When are Negative Online Reviews More Helpful than Positive Reviews? A Multi-method investigation into millions of online hotel reviews

Sanghyub, John, LEE

A thesis submitted to
Auckland University of Technology
in fulfilment of the requirements for the degree of
Master of Philosophy (MPhil)

2020

Faculty of Business, Economics and Law

Abstract

The role and volume of electronic word-of-mouth (e-WOM) online reviews worldwide are increasing rapidly so consumers, particularly in the tourism industry, may suffer from information overload. This, in turn, may impact on driving consumer behaviour. Thus, understanding factors that influence which reviews are perceived as helpful may be important for vendors in the tourism industry particularly in the hotel industry. This thesis suggests negativity bias and loss aversion as a theoretical anchor to illustrate the impact of star ratings on the perceived helpfulness of hotel reviews. Since prior research results appear to be diverse, there is a need to find which systemic moderators elicit different outcomes. This research seeks to provide a significant moderating role of reviews' differences such as consumer scepticism, and systematic information processing.

A quantitative approach via big data consisting of over two million online hotel reviews was adopted to address the inconsistent results. This research offers an enhanced predictive effect instead of small sample sized surveys used in prior studies. By deploying spatial regression discontinuity design between one-sided and two-sided reviews on Booking.com, as well as Agoda.com and Booking.com reviews, I proposed and validated the moderating role of consumer scepticism. This is expressed as 'too good to be true'. To make the results of the analysis more robust and addressing a small statistical effect size, the effect of the independent variables is not only measured in the statistical methods (regression and PROCESS macro) but also in traditional (bi-logistic regression) and new machine learning techniques (deep-learning).

The findings given in this work could offer pivotal implications for academics. They could additionally: (1) provide systemic moderators that elicit different outcomes; (2) illustrate a negative association between the review valence and the perceived helpfulness of the reviews; (3) document that when the level of consumer scepticism and heuristic information processing decreases, then negativity bias and loss aversion also weaken or are eliminated; (4) offer extensions for the broader research stream of e-WOM; and (5) extend the stream of research that utilizes big data with machine learning techniques. Limitations and directions for future research are discussed in the closing chapter.

Keywords

Big data, hotel online customer reviews, electronic word-of-mouth, negativity bias and loss aversion, machine learning model, deep learning

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person except where explicitly defined in the acknowledgements, nor material which to a substantial extent has been submitted for award of any other degree or diploma of a university or other institution of higher learning.

Name: Sanghyub, John, LEE

Signed: 이 상협.

March 2020

Acknowledgements

I see every day as an opportunity to challenge myself in order to achieve new chapters in my life, especially in the new technology area. Completing a thesis using new technology was not easy, but it was a meaningful journey. Without the support and encouragement of some people in my life, there would be no way for me to complete this journey.

I would like to express my sincere gratitude to my first and second supervisors, Dr Jungkeun Kim and Professor Roger Marshall. Amid many hardships, I almost gave up, but their guidance and support have been my biggest support to keep going. In the last MBA dissertation, I stated that I wanted to conduct big data research rather than using a survey with small sample size. Not only did they give me a lot of support to make the dream of big data research a reality, but also encouraged me to combine artificial intelligence and machine learning in order to take it one step further.

I would also like to express my sincere gratitude to Dr De Villiers, Rouxelle and Dr Kyuseop Kwak for demonstrating the core value of data analysis and thus broadening my horizons. I would also like to thank my colleague, Yuanyuan (Gina) Cui for the support we could give each other. Also, when presenting this research both in the 2019 ANZMAC Doctoral Colloquium and in the AUT Postgraduate Research Symposium, the invaluable and sharp feedbacks that mentors have provided has been a big step in solidifying the fundamentals of this study. I am grateful to John Bailie for providing a professional proofreading service.

Finally, I send my infinite love to my family, Helen, Tony, Renee and Claire who understood and encouraged me all the time even though I've been on my desk all day writing. They have helped me keep the hard work going.

Ethical Approval

In this study, ethical approval is not required because only secondary data is used.

Personal Motivation Statement

Good family bonding time and well-deserved rest to recharge a flat-out battery. Yes, it is a holiday season again. As a father of lovely three children, who have been eagerly waiting for this jolly time throughout the school term, I could not fail this annual mission of perfecting our holiday. I bravely dive myself into the countless information of destination and hotels to find that one place that would make us all happy.

I'd like to invite you to join me for this important journey of decision making through getting the right information with the help of large online review. So, after some hard-online research and thinking, I came to a point where I had to click and make the final payment. But wait, how can I know that I have made the right decision? So, I start to read some online reviews to back up my judgement. Then again, how do I know I can rely on certain reviews, let alone finding one in the swirling pool of reviews? I am already overloaded with tens of thousands of reviews for a single hotel.

First, I looked at other papers to find out which reviews were helpful, but I was confused because the results were not identical. To make my life easier and more efficient, I have developed my very own research assistant software which went into internet hotel Booking sites and collected over two million reviews from eight major cities in the United States. In other words, I collected big data. I believe that millions of actual online reviews can yield more reliable results than a few hundred questionnaires. Second, big data can give meaningful information by using artificial intelligence that mimics the human brain. With this, it can also predict which reviews will be more helpful.

My research came to an interesting conclusion that customers generally think negative reviews are more helpful. The more negative, the more helpful the review. On the other hand, in order for positive content to be helpful, it has to present both disadvantages and a lot of detail. My research also found that customers feel confident with 60.8% of accuracy with lengthier and two-sided reviews. Therefore, if you want to persuade others of positive information, don't just layout good things, but remember also to briefly mention the disadvantages as well.

Chapter 1: Introduction

Consumers often leave their evaluations and personal experience of staying in a hotel on the review boards of online travel websites. The peer-to-peer information such as star ratings and narrative reviews in online travel websites become vital sources of information that directly influences their purchase decisions. Researchers have found that online reviews help consumers have indirect experience in using or possessing the product (Simonson & Rosen, 2014), thus play a crucial role in generating sales (Chevalier & Mayzlin, 2006; Chintagunta et al., 2010; Duverger, 2013; de Langhe et al., 2015; Moe & Trusov, 2011; Sparks & Browning, 2011). Booking.com, one of the most popular travel online websites, has approximately 4.3 million hotel reviews posted on their travel site in February 2019, and approximately 40,000 online reviews are posted solely for the Pennsylvania hotel in New York. As vast amounts of information are available to use from the online reviews, consumers may become overloaded with information and have difficulty in evaluating the usefulness of the information created by the online users and consumers. This information overload has the potential to cause incorrect decision-making, inconvenience, confusion and stress (Frías et al., 2008). The information overload can be managed if consumers focus on useful information that other consumers have already rated as helpful (Liu & Park, 2015; Mudambi & Schuff 2010; Schindler & Bickart 2012; Schlosser, 2011; Willemsen et al. 2011).

This thesis hopes to answer questions regarding characteristics of an online review that may systematically affect a customer's judgment. Specifically, this study explores the impact of two key elements of online reviews, including star ratings (quantitative) and narrative reviews (qualitative) in order to predict the perceived helpfulness of online reviews. Several studies have examined the impact of review valence (i.e. star rating) on the perceived helpfulness of the reviews (Liu & Park, 2015; Mudambi & Schuff 2010; Schlosser, 2011; Schindler & Bickart 2012; Willemsen et al. 2011). However, the results of these studies are inconsistent (Mudambi & Schuff, 2010; Schlosser, 2011). Some researchers have confirmed the negativity effect by indicating that negative star ratings are important for the evaluation of the usefulness of the information (Willemsen et al., 2011). However, other researchers also have uncovered the positivity effect showing that the positive star ratings are highly valued in the evaluation of the usefulness of the information

(Liu & Park, 2015). Thus, this research attempts to provide the influence of moderators determining the direction of negativity or positivity bias. This thesis also asserts that the area of research conducted on which moderators influence the online review is limited (Mudambi & Schuff, 2010; Sen & Lerman, 2007; Willemsen et al., 2011; Zhang et al., 2010;).

Based on previous literature, this research suggests loss aversion (Barkley-Levenson et al., 2013; Tversky & Kahneman, 1992) as a theoretical anchor to explain the negativity bias effect of star ratings (quantitative review) on the perceived helpfulness of the reviews. This thesis also investigates the influence of narrative reviews (qualitative reviews) as moderators to predict the interaction effects with star ratings on the perceived helpfulness of the reviews. The narrative reviews include the type of information (one-sided information vs. two-sided information) and the length of comments. Specifically, two-sided information is more effective in reducing the doubt concerning the "too good to be true" effect. This thesis also proposes that the negativity bias will diminish as the length of narrative comments gets longer. In addition, this research proposes a three-way interaction of these moderators, by showing that a negativity bias will be diminished as the length of comments gets longer and two-sided information is provided, because consumers are more likely to engage in systematic information processing and the "too good to be true" effect would be weakened.

In order to address the research purpose, this research facilitates a higher predictive effect than the surveys and small sample sizes used in previous studies, by utilizing big data (Siegel, 2013). This thesis has used big data from two million reviews from both Agoda.com (60,266 reviews) and Booking.com (2,036,260 reviews), where consumers evaluate their experience with a particular hotel. Even though the age of big data is coming, traditional statistical data analysis is limited in analysing these large amounts of data. In order to analyse big data efficiently, the application of novel technology such as machine learning techniques is essential (Lei et al., 2016). Thus, to make the results of the analysis more robust, this research takes advantage of recent applications of machine learning in addition to statistical regression analysis.

This thesis contributes to the existing literature in several ways. Firstly, this research investigates the significant moderating role of narrative reviews, including the type of information and length of comments in the perceived helpfulness of the reviews.

Specifically, this research provides empirical evidence to show that negativity bias will diminish under systematic information processing only when the “too good to be true” effect is weakened. Second, this research incorporates secondary data from both Agoda.com (60,266 reviews) and Booking.com (2,036,260 reviews) to validate the findings from a series of studies. Two data sources that enable a spatial regression discontinuity design, as well as the size of the big data, improve the quality of the analysis results. Finally, this may be the first comprehensive study that explores the effect size of star rating on the perceived helpfulness of the reviews using statistical methods (regression and PROCESS macro), traditional (bi-logistic regression) and novel machine learning methods (deep-learning).

To this end, the contents of this thesis are organized as follows: Chapter 2 summarizes the theoretical grounding and Chapter 3 formulates and justifies the research hypotheses. Chapter 4 provides a description of the data collected, the methodology and analysis. This thesis also shows that the effect of interest cannot be explained by alternative mechanisms such as scarcity effects with a shortage of negative reviews, or cultural diversity in reviewers across both visited cities and reviewers’ nationalities. In Chapter 5, the results of traditional analysis (i.e., bi-logistic regression) and the more novel machine learning method are presented as an alternative approach for evaluating the 3-way interaction effect. This thesis concludes in Chapter 6 by discussing theoretical and practical implications and limitations.

Chapter 2: Review of Literature

2.1 Electronic word-of-mouth: Online Customer Reviews

2.1.1 Introduction

To properly commence this study, investigation of the comprehensive characteristics of online reviews described by previous studies is needed. A total of four main questions facilitate gaining insights regarding online reviews: (1) Why do customers post and view online reviews? (2) What is the impact of online reviews on customers? (3) What factors may influence customer star ratings as an indicator of customer satisfaction? (4) What are the characteristics of the online review text itself?

2.1.2 Motivation of online reviews

Why do customers post and view online reviews? Let's imagine this situation: when travellers visit a city for the first time and need to book a hotel, they may consider various strategies on how to go about booking a place to stay. Traditionally, they may ask friends and family or contact a travel agency. Word of mouth (WOM), which tends to be informal rather than formal, has a significant positive (Anderson, 1998) and negative impact (Richins, 1984) on customer satisfaction.

However, the way consumers communicate is evolving with the advent of web 2.0. In an information society where the Internet is universal, they can easily retrieve relevant information from the internet, whereas traditional WOM has limitations on the size of social networks (Hart & Blackshaw, 2006). Customers can post their opinions without difficulty on social media sites, blog sites, or sites that provide an online evaluation of specific products, while both the traditional and electronic WOMs have in common customers looking for advice and information from others in their decision-making process (Sparks & Browning, 2011). Thus, while WOM plays an essential role in customer purchase decisions, the role of electronic word-of-mouth (e-WOM) has become more critical in recent years due to the ease and availability of the Internet for both those posting opinions and reviews and those wanting to view the reviews.

As the supply and demand of e-WOM grow, Internet travel sites that offer online customer reviews of hotels in popular travel destinations are dramatically increasing, for example, sites such as www.Booking.com, www.Agoda.com and www.tripadvisor.com.

The online travel industry mainly consists of travel e-commerce sites and online customer review sites; revenue generated from online travel Bookings was 564.87 billion USD, and the growth rate of the worldwide online travel industry was 15.4% in 2016 (Statista, 2019). Therefore, the influence and economic impact of the online travel market that provides online customer reviews is large and increasing.

Specifically, the motivation for online reviews can be divided into two categories: posting motivations and viewing motivations. Firstly, to better understand the posting motivations of online reviews, Yoo and Gretzel (2008) found that traveller reviewers are motivated to express concerns and enjoyment for other consumers or travel service providers. Interestingly, writing negative online reviews to evoke negative emotions does not seem to be an important motivation. A systematic literature review by Litvin, Goldsmith and Pan (2008) described how posting motivations involve four factors: affect, altruism, self-interest and reciprocation. Moreover, the posting motivations for positive and negative reviews are induced differently. For positive reviews, a restaurant's food quality, service quality, and atmosphere have a positive effect on posting positive reviews to customers, whereas the restaurant's price equity does not have a positive effect on posting positive reviews for customers (Jeong & Jang, 2011). As for negative reviews, angered customers post negative reviews for revenge, while disappointed customers post negative reviews to warn others. Customers who experience regret post negative reviews to strengthen social ties or warn others (Wetzer, Zeelenberg & Pieters, 2007). Also, customers are motivated to write negative reviews by venting and altruism or revenge, posting detailed, convincing and reliable negative reviews (Sparks & Browning, 2010). In a study investigating travellers' personality traits, Yoo and Gretzel (2011) reported that personality dimensions such as neuroticism, extraversion, openness, agreeableness, and conscientiousness have a significant impact on posting barriers, as well as posting motivations. Finally, in an investigation into viewing motivations, Kim, Mattila & Baloglu (2011) found that women are likely to view online reviews for convenience, quality and risk reduction, while men are more likely to search and view online reviews based on their level of expertise. In the following section, extant literature concerning the impact of online reviews on online purchase intentions will be analysed to understand the various impacts of online reviews.

2.1.3 Impact on online purchase intentions

What is the impact of online reviews on customers? Both WOM and online customer reviews (eWOM) play an essential role in driving consumer behaviour. However, online consumer reviews have different characteristics. Firstly, unlike WOM, numerous positive and negative online consumer reviews are offered simultaneously, including written form and photos, on the same online site (Chatterjee, 2001). The second aspect of online consumer reviews is that they can be measured. Consumers can easily observe and measure the quantity and quality of positive and negative comments because online consumer reviews are posted in written form (Chevalier & Mayzlin, 2006). Star ratings also make it easy for consumers to measure their recommendations (Dellarocas & Narayan, 2006). This eliminates the need to rely on information provided by traditional product marketing that emphasises the positive side of their products. Customers can now put a higher value on electronic word-of-mouth, where they can evaluate reviews (Lee, Park & Han, 2008). The characteristics of these online reviews have a significant impact on customers' online purchase intentions. The findings of previous studies on online purchase intentions divided by reviews from travel and hospitality industries are as follows.

Firstly, a large and growing body of literature has investigated the relationship between online reviews of various industries and purchase intentions. In 2009, Cox et al. published a paper in which they described that while websites with online reviews are proliferating, they have not been considered as reliable as official source sites, such as government-sponsored tourism websites. As a result, travellers perceived that online reviews were not the only source of information but as an additional source of information as part of the information retrieval process. However, in 2010, Yoo & Gretzel demonstrated that credibility of online reviews depends on the type of website published and the perception of the reviewer. Also, consumers with high confidence in online reviews have a significant impact on their travel plans with the use of online reviews. For instance, as for restaurant reviews, the star rating of the online reviews on food quality, environment and services of the restaurant and the number of online reviews are positively related to the online popularity of the restaurants. Interestingly, official editor reviews have a negative relationship with the consumer's intention to visit a restaurant (Zhang et al., 2010). The valence and number of reviews, providing online coupons, and the number of search keywords all has a positive impact on weekly restaurant revenue (Lu et al., 2013). As for computer game reviews, online reviews have a more significant impact on the sales of games when they are less popular or the player has great internet experience (Zhu & Zhang, 2010).

Secondly, with regard to hotel agency, the existing literature is extensive and focuses particularly on the relationship between hotel reviews and hotel booking intentions. The presence of a reviewer's personally identifiable information has a positive effect on the credibility of online reviews which, in turn, positively affects the consumer's intention to book a hotel (Miao, Kuo & Lee, 2011). Customer reviews have a significant impact on online sales. Specifically, when the star ratings of customer reviews increase by 10%, online bookings increase by more than 5% (Ye et al., 2011). The quality and location ratings of customer reviews are essential factors in determining hotel room prices. However, the content of customer reviews, such as room, location, cleanliness and service, affect the room rates differently depending on the hotel ratings (Zhang, Ye & Law, 2011). Online travel sites such as Booking.com play an essential role in increasing hotel reputation. Specifically, efforts could be made to improve hotel staff service quality, and eventually consumers are likely to pay a price premium when a good score for the staff service rating in online reviews is presented (Öğüt & Onur, 2012). The ratings of a hotel room, location and staff in customer reviews is the most crucial attribute in positively affecting guest re-visit intentions and hotel recommendations (Yacouel & Fleischer, 2012). The higher the customer star rating, the greater the online sales of the hotel. Specifically, when the customer star rating increases by 1%, the sales per room increase by 2.68% in Paris and 2.62% in London. Also, the higher the guest star rating, the higher the hotel room price, and room prices in luxury hotels are more sensitive to customer star ratings (Zhang & Mao, 2012). Finally, there seems to a positive correlation between hotel purchase intentions and customer expectations and customer star ratings (Mauri & Minazzi, 2013). In the following section, literature concerning the link between the management of customer satisfaction and online review indications of customer satisfaction will be analysed, in order to better understand factors that influence high customer star ratings.

2.1.4 Management of customer satisfaction and online reviews

What factors might influence customer star ratings as an indicator of customer satisfaction? The company's management efforts concerning customer satisfaction can be divided into two factors; direct (response directly to consumer online reviews) and indirect (management of existing tangible and intangible assets). Firstly, a large and growing body of literature has investigated the relationship between direct management factors and customer satisfaction. Responses to online reviews are effective for customers with low star ratings, but do not appear to affect other customers. Specifically, dissatisfied customers who received a response to their online reviews increase their future satisfaction, however dissatisfied customers who do not receive a response decrease their future satisfaction (Gu & Ye, 2014). However, as sites offering online reviews proliferate, hotels are losing control of the content of online reviews, unlike advertising (Dwivedi, Shibu & Venkatesh, 2007). Rather, the hotel manager's response to customer reviews has a negative impact on hotel purchase intentions (Mauri & Minazzi, 2013). O'Connor (2010) found that very few hotels respond to, and manage, customer reviews directly on the hotel reservation site, TripAdvisor. For instance, in traveller online reviews, only one out of five negative reviews have been responded to by the hotels (Lee & Hu, 2004). Specifically, only highly-rated hotels typically respond to online complaints. The response usually consists of thanks, apologies, and explanations for mistakes, with few mentions of compensation adjustments (Levy, Duan & Boo, 2013). This suggests that direct management of soaring online reviews can be difficult.

Secondly, other studies have investigated the relationship between indirect management factors and customer satisfaction. Staff service, bathroom and room cleanliness and noise issues were found to be the most common complaints in customer reviews (Levy, Duan & Boo, 2013). Stringam and Gerdes (2010) found that the usage pattern of the words that are most used when the customer star rating is "high" or "low." Specifically, the lower the customer star rating, the more likely travellers were to post about hotel room beds and their components. In contrast, the word clean is used most often when the customer star rating is highest. The higher the rating, the more often the word clean is used. Thus, it suggests that the cleanliness management of the hotels has an essential effect on customer satisfaction. Moreover, high prices have a positive effect on quality among customer star rating factors, whereas it has a negative effect on value among customer star rating factors. Price also has a greater impact on the quality of a luxury hotel than low customer star ratings and hotel facilities (Ye et al., 2012). Interestingly, when the hotel provides free Wi-Fi to customers, it increases customer star ratings by up to 8%. Hotels that allow pets can improve customer star ratings up to 1% over hotels that do not (Bulchand-Gidumal, Melián-González & López-Valcárcel, 2011). In restaurants' online reviews, it was found to affect customer satisfaction in the following order: food, service, ambience, price, menu, and décor (Pantelidis, 2010). In the following section, the literature concerning opinion mining and sentiment analysis of online review texts will be analysed, to highlight the online review text itself.

2.1.5 Sentiment analysis and opinion mining of online review text

One of the major reasons to focus on the online review text itself is that customer star ratings, the most used quantitative variables, are biased towards positive ratings, while balanced and negative reviews are used relatively less (Racherla, Connolly & Christodoulidou, 2013). For instance, in online restaurant reviews, positive reviews make up approximately 78%, while negative reviews only make up approximately 22% (Pantelidis, 2010). Racherla, Connolly and Christodoulidou (2013) argue that customer star ratings might not be the ideal indicator of customer satisfaction. Thus, the analysis of the consumer review text itself could reveal nuanced opinions of customers that can be lost by customer star ratings. To this end, a variety of machine learning methods using artificial intelligence are used to analyse large amounts of review text. Customer review text analysis is divided into two categories: sentiment analysis and opinion mining.

Firstly, the existing literature mainly focuses on sentiment analysis of customer reviews by utilizing natural language processing techniques. When applying sentiment classification techniques to traveller reviews, all three machine learning algorithms, for example, N-gram, SVM, and Naive Bayes show more than 80% accuracy. Specifically, the N-gram and SVM algorithms showed better accuracy than the Naive Bayes algorithm (Ye, Zhang & Law, 2009). However, when using the improved Naive Bayes algorithm, the difference between positive and negative review accuracy is reduced to 3.6%. Similarly, compared to SVM, the gap in accuracy is reduced from 28.5% to 3.6% (Kang, Yoo & Han, 2012). When applying the Naive Bayes algorithm and the SVM algorithm to Cantonese written restaurant reviews, the Naive Bayes algorithm showed similar or better accuracy than the SVM algorithm (Zhang et al., 2011). Taken together, it suggests that various machine learning algorithms can categorise customer reviews with high accuracy, whether they are positive or negative, both in English and Cantonese.

Secondly, a large and growing body of literature has investigated opinion mining of customer reviews. Akehurst (2009) argued that despite a large amount of consumer review text, analysis requires steps to find, extract, and interpret reviews. However, this is not only a time-consuming task, but also has high labour costs. Accordingly, the value of the acquired consumer review text tends to be ignored. Fortunately, opinion mining, a technique of machine learning using artificial intelligence, can overcome obstacles and release the value of consumer review texts. For instance, high-frequency complaint words in customer review text are classified into eighteen problem categories, using Ward's clustering method. It is found that complaints about the hotel's services – for example, unfriendly staff, were most often mentioned (Lee & Hu, 2004). Pekar and Ou (2008) investigated a method to recognise the relationship between product features, for example, location, food, room, services, facilities, price and subjective expressions using sentiment lexicons such as GI, SWN, and Roget. Opinion mining of customer review text categorised into four quality classes (hotel service, hotel condition, room cleanliness, and room comfort), is strongly correlated with customer star ratings and recommendations. Specifically, hotel service and room comfort were most strongly correlated followed by hotel condition and room cleanliness (Stringam, Gerdes & Vanleeuwen, 2010). As mentioned, sentiment analysis and opinion mining of online review text is a technique for identifying nuanced opinions of customers. However, analysing the online review text itself could provide further insight into this study. Sentiment analysis using natural language processing techniques will be mentioned later in the validation processes of spatial regression discontinuity design.

2.1.6 Conclusion

As shown above, this study examines the comprehensive characteristics of online reviews as depicted through previous studies. A total of four main questions facilitate gaining insights regarding online reviews; 1) Why do customers post and view online reviews? 2) What is the impact of online reviews on customers? 3) What factors may influence customer star ratings as an indicator of customer satisfaction? 4) What are the characteristics of the online review text itself?). Next, this study builds on these insights to further explore the key variables for further enquiry.

2.2 Online customer reviews' characteristics

2.2.1 Introduction

Having reviewed some current literature, the characteristics of online reviews have been identified. The research reported in this thesis is based on big data, which has the advantage of increasing the predictive effect, but also has the disadvantage of limiting the selection of variables. Traditional surveys can generally collect different questions related to topics and demographic data; however, big data cannot collect personally identifiable information and only includes variables collected for the original purpose of the data set (Whitaker, 2018). Thus, this study will investigate the important variables that can be collected using big data (i.e. perceived helpfulness of online reviews, star ratings and narrative reviews).

2.2.2 Perceived helpfulness of online reviews

At present, most retailers offer online star ratings and narrative reviews that users generate online. Star ratings play a crucial role in product sales (de Langhe et al., 2015). The expansion of peer-to-peer information, including online star ratings, allows consumers to quickly and easily predict rich and specific information, such as the possession or use of a product (Simonson & Rosen, 2014). However, due to rapidly increasing electronic word-of-mouth, consumers may be subject to information overload.

As more people access and supply information on the Internet, the information on the Internet is exploding. According to Internet Live Stats (2019), about 6,000 tweets are tweeted, more than 40,000 Google searches are conducted, and over 2 million emails are sent per second. In terms of online hotel reviews, according to Booking.com (2019), approximately 4.3 million hotel reviews were posted on their travel site from February 2019 to November 2019. As an example of the extent of this phenomenon, approximately forty thousand online reviews are posted solely for the Pennsylvania hotel in New York!

Information overload has the potential to adversely affect decision-making and add inconvenience, confusion and stress to customers (Frías et al., 2008). In this regard, only a small amount of information accessed by customers is voted as helpful. Moreover, the helpfulness and usefulness of information play a mediating role between influence processes and information adoption (Sussman & Siegal, 2003). Compared to the importance of helpful information, most reviews are not perceived as such, thus, I believe it vital to identify factors that determine which reviews are perceived as useful from those that are not.

2.2.3 Star ratings and Narrative reviews

As asserted earlier, online reviews can primarily be divided into quantitative and qualitative elements. Reviewers provide an overall quantitative assessment of product experience (i.e. online star ratings) followed by detailed qualitative assessments (i.e. online narrative reviews) (Sridhar & Srinivasan, 2012). Review valence (star ratings) represents the number of stars rated by the reviewer and represents an assessment of a hotel's facilities and services. The overall rating is considered a useful clue to reflect consumer attitudes and helps consumers to assess the quality of the hotel (Liu & Park, 2015).

Some researchers have shown a positive relationship between online review ratings and customer behaviour (Gauri et al., 2008; Ho-Dac et al., 2013; Lee et al., 2008; Liu & Park, 2015; Ye et al., 2009). Consumers take a more favourable attitude toward the product when more positive reviews are received (Lee et al., 2008). Positive online reviews have the most significant impact on customers' purchase intentions (Gauri et al., 2008). Ye et al., (2009) also claim that positive hotel online reviews can significantly increase the number of bookings at a hotel. Also, cumulative positive reviews appear to have a more significant impact on product sales than cumulative negative reviews (Ho-Dac et al., 2013). In a previous study of online reviews of experience goods (hotels and restaurants), it was suggested that a star rating had a positive relationship with the usefulness of reviews (Liu & Park, 2015).

On the other hand, some researchers have shown a negative relationship between online review ratings and customer behaviour (Ba and Pavlou, 2002; Chevalier & Mayzlin, 2006; Lee et al., 2008; Papathanassis & Knolle, 2011). Negative reviews tend to have more impact on consumers than positive reviews in processing and adoption of, for instance, online holiday reviews (Papathanassis & Knolle, 2011). Lee et al., (2008) add that high-quality negative online consumer reviews have a more significant impact on consumer attitudes than low-quality negative online consumer reviews. Even if the number of positive reviews overwhelms the number of negative reviews, negative reviews still affect consumers, in the sense that negative reviews are more effective than positive reviews

(Chevalier & Mayzlin, 2006). This is because negative review ratings have a more significant negative weight than a positive review rating in terms of the buyer's confidence in the seller (Ba & Pavlou, 2002).

As star ratings (quantitative) have a significant impact on consumer attitudes, narrative reviews (qualitative) also have a significant impact on consumer attitudes. Prior studies have noted the importance of narrative reviews. In addition to star rating in online reviews, the quality of online reviews affects consumers' purchase intent positively (Park, Lee & Han, 2007). After analysing the review length data, Chevalier & Mayzlin (2006) found that the customers are actually reading the review text. Community members rely less on reviewer information disclosure in helpful review ratings if the review text is clear and informative (Forman, Ghose & Wiesenfeld, 2008). Taken together, it transpires that consumers use not only star ratings (quantitative) but also narrative reviews (qualitative) as important indicators for their choices. Table 1 shows the key articles for star ratings and narrative reviews.

Table 1: Key reading table for Star Ratings and Narrative reviews

Source	Method	Context	Findings
		As for dual valance effect	
Vermeulen and Seegers, 2009	168 respondents completed the entire experiment	Online hotel reviews	Positive or negative reviews increase the consumer's perception of the hotel, while positive reviews improve the attitude toward the hotel. This effect is more powerful in less-known hotels.
Sparks and Browning, 2011	554 community members who had been randomly assigned to one of 16 conditions	Online hotel reviews	Consumers are more affected by initial negative information, especially when the overall set of reviews is negative. However, positively organized information along with numerical star ratings information increases both Booking intent and consumer confidence.
Zhang, Craciun and Shin, 2010	A sample size of 150 undergraduate students	Photo-editing program, anti-virus program	Consumers who evaluate products associated with promotion consumption goals perceive positive reviews to be more persuasive than negative reviews. In contrast, consumers who evaluate products associated with prevention consumption goals perceive negative reviews to be more persuasive than positive reviews.
		As for positive valance effect	
Park, Lee and Han, 2007	352 students with a manipulated experiment	Online shopping malls.	The quality of online reviews has a positive effect on the consumer's intention to buy. As the number of reviews increases, the purchase intent increases.

Gupta and Harris, 2010.	Respondents were directed to the AtoZTronics website for questionnaire.	Online shopping mall. Laptops	e-WOM allows consumers to spend more time considering the recommended products.
Gauri, Bhatnagar and Rao, 2008	BizRate.com and Alexa.com. After a purchase, requested to complete a feedback form.	An online price comparison Web site	It is not the total number of reviews that affect customer repurchase intentions, but the percentage of positive reviews.
Ye, Law and Gu, 2009	The data, 3625 reviews for 248 hotels, were retrieved from Ctrip.	Online hotel reviews	Online consumer-generated reviews have a positive impact on hotel room sales.
Ho-Dac, Carson and Moore, 2013	Data, a total of 3,341 OCRs, from Amazon.com, except for advertising data	Amazon.com DVD players	Positive OCR increases sales of models of weak brands. Higher sales generate more OCR. This creates a positive feedback loop between sales and positive OCR for weak brand models, which not only helps sales, but also increases overall brand equity.
As for negative valance effect			
Lee, Park and Han, 2008	354 students with a manipulated experiment	Online shopping mall. MP3 player	A high percentage of negative online consumer reviews will have a conformity effect. As the proportion of negative online consumer reviews increases, product attitudes become more adverse, depending on the quality of negative online consumer reviews.
Papathanassis and Knolle, 2011	22 sessions qualitative approach was adopted	Online holiday reviews Hotels, and resorts	Negative reviews appear to have a greater impact than positive reviews. Consumers are only 40% leveraging and combining various content sources directly related to online vacation reviews.
Chevalier and Mayzlin, 2006	Recent 500 reviews of the book, collected from Amazon.com and bn.com	Amazon.com Barnesandnoble.com Books	Reviews on both target sites are overwhelmingly positive, but the impact of 1-star reviews is higher than that of 5-star reviews. Evidence of the review length data indicates that the customer read the review text.
Ba and Pavlou, 2002	393 users who received an email telling them how to access the web site	An online auction market, eBay.com	The negative rating itself did not have a high impact on the price premium of the eBay data. In fact, the only time negative ratings were important was when expensive products were involved in the deal.
Park and Kim, 2008	Two hundred and twenty-two undergraduate and graduate students participated in the experiment.	The portable multimedia player (PMP)	Novices will be more sensitive to benefit-based negative reviews, and experts will be more sensitive to attribute-based negative reviews in terms of cognitive fit with reviews.

Chatterjee, 2001	Undergraduate marketing (314) and physics (105) students in two north-eastern universities	The recommended course textbook	The detrimental effect of negative consumer reviews on retailers' perceptions of their trust and intention to buy is mitigated by the consumer's familiarity with retailers. Also, for companies that are not familiar to consumers, the effect may be more negative.
Bambauer-Sachse and Mangold, 2011	216 people participated in the study	Computer notebook, digital camera	Negative online product reviews have a significant detrimental effect on consumer-based brand assets, which leads to significant brand asset dilution even when the brand is familiar to customers.

2.2.4 Conclusion

This Chapter has examined essential and collectable variables through collected big data. The characteristics of perceived helpfulness of online reviews, star ratings and narrative reviews and findings from previous studies were also reviewed. The next section builds on this insight to further explore the relationship of variables and related theories pertinent to this study.

2.3 Mixed findings of star ratings on perceived helpfulness of the review

Several studies have examined the perceived helpfulness of the review. (Liu & Park, 2015; Mudambi & Schuff 2010; Schindler & Bickart 2012; Schlosser, 2011; Willemsen et al. 2011). Previous research on the moderators of existing online review studies was generally focused on product characteristics. For example, for products related to promotional consumption goals (e.g., photo-editing software), consumers express a positive bias by evaluating positive reviews as more persuasive than negative reviews. Conversely, consumers have a negative bias about products related to the goal of preventing consumption, for example, antivirus software (Zhang et al., 2010). However, previous studies also evaluate the perceived helpfulness of the review observed inconsistent results on experience products such as hotels.

As for experience products, economists and marketers categorize products or services into a search for experience products according to the degree to which consumers can evaluate or obtain information. Search products are mostly products that can be evaluated before purchase (e.g., clothing, office supplies, furniture). In contrast, experiential products are mostly services (e.g., hotels, restaurants, travel, vacations) that can be accurately evaluated only after purchasing and experiencing the product (Bei et al., 2004).

First, analysing online reviews of MP3 players, music CDs, and PC video games on Amazon.com reveals that moderate reviews receive higher perceived helpfulness scores than extreme reviews (whether strong positive or negative) of experience products (Mudambi & Schuff, 2010). In other words, experience products display an inverted U-shape relationship between the perceived helpfulness of the review and star ratings.

Second, analysing online reviews of books on Amazon.com reveals that extreme reviews (1, 2, and4, 5 points) receive higher perceived helpfulness ratings than intermediate reviews (3 points) (Forman, Ghose & Wiesenfeld, 2008). In other words, there is a U-shape relationship between the perceived helpfulness of the review and star ratings for experience products. Apart from amazon.com data, analysing 48 different scenarios of hotels reveals that unbalanced (positive or negative) review sets are perceived more useful than balanced (neutral) review sets (Purnawirawan, Pelsmacker & Dens, 2012). Again, experience products exhibit a U-shape relationship between the perceived usefulness of the review and star ratings.

Third, analysing online reviews of experience products (sunscreen, an espresso machine, running shoes, shaving equipment and diet pills) on Amazon.com reveals that negatively valanced reviews receive higher perceived helpfulness scores than positively valanced reviews (Willemsen et al., 2011). In other words, experience products have a negative relationship between the perceived helpfulness of the review and star ratings.

Finally, analysing online reviews of E-books on Amazon’s KINDLE reveals that positively valanced reviews receive higher perceived helpfulness than negatively valanced reviews (Li & Zhan, 2011). In other words, experience products appear to generate a positive relationship between the perceived helpfulness of the review and star ratings. Apart from Amazon.com data, analysing online reviews of restaurants on Yelp.com reveals that positively valanced reviews be seen as more helpful than negatively valanced reviews (Liu & Park, 2015). That is, for experience products there seems a positive relationship between the perceived helpfulness of the review and star ratings. One interesting finding is that hedonic products such as music CDs, fiction books, general magazines, movie videos, and DVDs also show a positive relationship between the perceived helpfulness of the review and star ratings (Sen & Lerman, 2007).

Since the results of previous studies on experience products are different, research is needed to address this inconsistency. As the area of research conducted concerning moderators' influence the online review is limited (Mudambi & Schuff, 2010; Sen & Lerman, 2007; Willemsen et al., 2011; Zhang et al., 2010), there is a need to find systemic moderators that elicit different outcomes. Table 2 shows the key articles for helpfulness of the reviews.

Table 2: Key reading table for helpfulness of the reviews

Source	Method	Context	Findings
Forman, Ghose and Wiesenfeld, 2008	Data using automated Java scripts. 175,714 reviews of Amazon's 786 unique books	Amazon.com Books	Identity information about reviewers determines community members' product and review judgment. For books, moderate reviews (3 points) are less helpful than extreme reviews (1,2,4,5 points).
		Books	U Shape
Mudambi and Schuff, 2010	The online reviews available through Amazon.com, resulting in a data set of 1,587 reviews of the 6 products	Amazon.com MP3 player, Music CD, PC video game, Cell phone, Digital camera, Laser printer	Moderate product reviews are more useful than extreme reviews (strong positive or negative) on experience products, but not useful for search products. Also, longer reviews generally increase the usefulness of the review, but this effect is greater for search products than experience products.
		Experience goods	Inverted U Shape
Sen and Lerman, 2007	100 reviews, One hundred thirty-seven MBA students	The utilitarian products were cell phones, digital cameras, PDAs, computer monitors and printers, and hedonic were music CDs, fiction books, general magazines, movie videos, and DVDs.	Readers of negative reviews of utilitarian products showed negativity bias. 61% of all those who rated the help of negative reviews on utilitarian products pointed out that this is helpful. In the case of hedonic, a large percentage (72%) of readers found that negative reviews were "not useful".
		Utilitarian products Hedonic products	Negative relationship Positive relationship
Purnawirawan, Pelsmacker and Dens, 2012	413 respondents were randomly assigned to the 48 different scenarios	Hotels	The perceived usefulness of an online review set is affected by its balance and order. An unbalanced (positive or negative) review set is considered more useful than a balanced (neutral) evaluation set. The review order affects the perceived usefulness of the review set only for unbalanced review sets.
		Hotels	U Shape

Willemsen, Neijens, Bronner and De Ridder, 2011	42,700 reviews covering 38.745 reviews of cameras, 2,497 of DVD players, 1,032 of running shoes and 426 of sunscreen from Amazon.com	Search products (i.e., a digital camera, a laser printer, a DVD player and a food processor) experience products (i.e., sunscreen, an espresso machine, running shoes, shaving equipment and diet pills).	Negatively rated reviews lead to higher perceived usability than positively rated reviews. This effect is more noticeable on experience products than on search products.
		Search products Experience products	Weak negative relationship Strong negative relationship
Li and Zhan, 2011	1,793 individual reviews. Participants were 104 university students	Amazon's KINDLE	Negative reviews did not have a higher level of usability than balanced reviews. Review readers with a positive prior attitude to the product tended to prefer positive reviews over negative reviews, and readers with negative attitudes rated the positives more favourably, but only when they were more engaged.
		Online E-books	Positive relationship
Liu and Park, 2015	5090 online restaurant reviews from Yelp.com	Restaurants	The star rating of an online review has a positive relationship with the perceived usefulness of the review.
		Restaurants	Positive relationship

2.4 Negativity bias and loss aversion and Positivity offset

As shown in the findings above, consumers tend to weigh one side more prominently, rather than according positive and negative online consumer reviews the same weight. The main difference is that when a consumer puts more weight on negative information, it is called a negativity bias, whereas when they put more weight on positive information, it is called a positivity offset (Ito & Cacioppo, 2005).

On the one hand, according to negativity bias, despite rational thinking requires that information of a negative and positive nature should be evaluated the same when they have the same intensity, negatively valanced information, such as unpleasant thoughts, feelings or social interactions, have a greater impact on psychological state and processing than of neutral or positively valanced information (Baumeister et al., 2001). Rozin and Royzman (2001) argue that humans place greater weight on negative entities than positive entities (e.g. events, things and personal characteristics). Also, negative entities are more emotionally contagious than positive entities.

Another important finding relating to this behaviour comes from research that shows that the human brain tends to put more weight on negative information. This was discovered by showing positive, negative and neutral pictures and recording event-related brain potentials (Ito et al., 1998). Even infants of three months old evaluated people based on others' social behaviours, and a negativity bias was observed in their social evaluation (Kiley-Hamlin, Wynn & Bloom, 2010). Therefore, customers will weigh negative reviews with low star ratings more than positive reviews with high star ratings.

Apart from a negativity bias, Tversky and Kahneman (1992) assert demonstrate that the loss aversion motivation is strong in decision-making, where losses (or potential losses) are more powerful than (potential) gains. The value function graph of their Prospect Theory has an asymmetric “S” shape. The loss side of their graph shows a steeper slope than the profit side. If they were the same, the value of the loss would be twice the value of the profit. Investigating the effects of potential losses and benefits in a situation of risk, the potential loss affects choice more powerfully than potential benefits, not only with regard to adults but also with regard to adolescents. The stronger the loss aversion, the stronger the risk avoidance (Barkley-Levenson et al., 2013). This theory also predicts that consumers will weigh negative reviews, that avoid risks, more than positive reviews that recommend hotels.

Together, these studies indicate that negative reviews could have a greater effect on the perceived helpfulness of the reviews than positive reviews. On the other hand, according to the positivity offset, people interpret a neutral situation as slightly positive (Cacioppo, Gardner & Berntson, 1999) and evaluate most of their life as “a good time” (Diener & Diener, 1996). The positivity offset is asymmetric to the negativity bias. One of the reasons people show positivity offset maybe because people positively evaluate their surroundings when there are no perceived threats, a neutral situation. Therefore, in a neutral situation, people are motivated to explore and participate in the environment while they choose to avoid risk or explore. Without such a motivation for exploration, people may not be able to reap the potential reward value they can get from a neutral situation. Because of positivity offsets, people who tend to explore in neutral situations not only increase survival value but also social cohesion (Cacioppo, Gardner & Berntson, 1999). Furthermore, individuals with stronger positivity offsets evaluated only neutral information more positively, whereas individuals with stronger negativity bias evaluated negative information more negatively (Ito & Cacioppo, 2005).

Further to this point, surveys were conducted involving respondents from 43 countries in order to ascertain those who were happy with their lives and those who were not. Most people, that is up to 86% of the 43 countries, reported a positive level of subjective well-being and said they were satisfied with their lives. A positive assessment of the past can help people to be positively motivated in their upcoming exploration. This plays an important role not only in human survival but also in social organization (Diener & Diener, 1996). However, in online review ratings that require an immediate and intense response, as customers interpret the neutral reviews and take time to re-evaluate reviews, the positivity offset does not explain how neutral and positive reviews could have a more significant effect on the perceived helpfulness of the reviews than negative reviews.

In summary, this research suggests negativity bias and loss aversion as theoretical anchors to illustrate the impact of star ratings on the perceived helpfulness of the reviews. However, as mixed findings were found in previous research, this research will address this inconsistency, within the research question presented below:

When are negative reviews more helpful than to customers than positive reviews?

2.5 Consumer scepticism; "too good to be true"

In addition to the quantitative assessment of product experience (online star ratings), qualitative assessments (online narrative reviews) influence consumers' decisions (Sridhar & Srinivasan, 2012). A significant element of argument quality in online narrative reviews is whether the information contains one-sided or two-sided information. One-sided information provides biased information, either positive or negative. Two-sided information contains both negative and positive sides. For example, two-sided persuasion is information that provides the nature of a product or service, together with both positive and negative information. Two-sided information is more effective in increasing source credibility than one-sided information (Li & Zhan, 2011).

On the one hand, source credibility is important because of consumer scepticism, "too good to be true," which can be triggered by fake reviews. CNBC (2019), by using artificial intelligence to judge the authenticity of online reviews, found that suspicious fake online reviews of Amazon.com have increased from 16.34% in 2018 to over 34% in all product categories in 2019. Indeed, tens of thousands of suspicious, probably fake, five-star ratings online reviews are found at Amazon.com. Generally, two-sided information is believed to be more credible and persuasive for consumers because each product has positive and negative characteristics (Cheung et al., 2009; Crowley & Hoyer, 1994). If negative reviews are included in online reviews, consumers can be given increased believability (Schindler & Bickart, 2005). For example, providing both positive and negative information in an online review of a hotel can help customers better assess the quality of the hotel (Filieri & McLeay, 2014). Negative reviews, which account for a small percentage of the overall review, could be beneficial to consumers (Mulpuru, 2007). Moreover, if all online review messages are positive, the credibility of the website and online review messages may decline over the long term (Doh & Hwang, 2009). Furthermore, existing studies argue that two-sided information is more effective at persuading consumers to use advertised products than one-sided information. (Belch, 1981; Etgar & Goodwin, 1982; Golden & Alpert, 1987; Hastak & Park, 1990; Kamins & Assael, 1987; Kamins & Marks, 1987; Kamins, 1989; Pechmann, 1992; Sawyer, 1973; Settle & Golden, 1974; Smith & Hunt, 1978; Stayman et al., 1987). For instance, researchers measured the effects of advertisements including negative information for products such as a pen (Hastak & Park, 1990; Kamins & Assael, 1987; Kamins & Marks, 1987; Sawyer, 1973; Settle & Golden, 1974), a grocery store (Belch, 1981), remedies (Etgar & Goodwin, 1982), deodorants (Golden & Alpert, 1987), home computers (Kamins, 1989), ice cream (Pechmann, 1992), televisions (Smith & Hunt, 1978), and clocks (Stayman et al., 1987).

On the other hand, O'Connor (2010) argued that the belief that fake reviews compromised the credibility of user review sites does not hold enough evidence. There is little solid evidence for reviews with characteristics that represent fake reviews. Moreover, each year, hundreds of millions of potential hotel visitors refer to these travel sites, and over eighty per cent of these visitors are influenced by the choice of hotel online customer reviews reviewed. (Vermeulen & Seegers, 2009). Furthermore, half of the consumers planning to travel examine other consumer posts with regard to their travel plans, and eight out of ten examine online travel reviews. Also, eight out of ten Internet users at least somewhat trust the information posted by other travellers (Gretzel et al., 2010). Nevertheless, consumer scepticism about fake reviews has not yet disappeared as news of fake reviews continues to be reported. For instance, according to BBC News (6 September

2019), Trip Advisor could not prevent attempts by numerous fake reviews to increase hotel ratings through artificial manipulation.

Collectively, these studies indicate that because of consumer scepticism, two-sided reviews could have a greater effect on the perceived helpfulness of the reviews than one-sided reviews. Table 3 shows the key articles for two-sided information.

Table 3: Key reading table for two-sided information

Source	Method	Context	Findings
Filieri and McLeay, 2014	An online questionnaire with a convenience sample of 55.	Hotel accommodations while planning their holidays	Information quality related to the adoption of information in online reviews leading value-added information. By providing both positive and negative information, travellers can provide an important description of their accommodation and rate better quality.
Crowley and Hoyer, 1994	In prior research		Two-sided arguments are more persuasive than one-sided arguments.
Eisend, 2007	With 190 participants, a fictitious brand (a pizza restaurant); with 186 participants,	A real brand (a Sony notebook).	Two-sided messages enhance perceived novelty, attention and motivation reinforce attitudes toward advertising, and attitudes toward advertising reinforce attitudes toward brands. However, if too much novelty is induced, the perceived novelty can have a negative effect.
Golden and Alpert, 1987	568 respondents reached by telephone.	Economy for bus, protection from odour for deodorant	The number of negatively mentioned attributes had the greatest impact when it was two of five dependent variables. The perceived quality of information peaked at three positive claims and two negative claims and fell for more and less consisted.
Cheung, Luo, Sia and Chen, 2009	A questionnaire with 1,195 respondents.	On-line consumer discussion forum, www.myetone.com	Two-sided information is generally considered more reliable for consumers because each product has both positive and negative features. Two-sided descriptions are perceived as more detailed information that affects the strength of the argument.
Schindler and Bickart, 2005	19 consumers who claimed they “frequently shop online.”	Online shops	Reliability can be improved if the content of the Internet WOM message contains negative information.
Doh and Hwang, 2009	143 samples in South Korea with self-administered questionnaires	Movies and digital cameras	If all eWOM messages are positive, then the credibility of the website and eWOM messages can decline in the long run.

2.6 Systematic versus heuristic information processing

According to Kahneman (2011), human psychology is characterised by various heuristics and biases, essentially because people have two systems of thinking.

System 1: Heuristic information processing. System 1 is an automatic, fast and unconscious mindset. It is autonomous, efficient, and requires little energy or attention. However, it often leads to prejudice and systematic errors.

System 2: Systematic information processing. System 2 is a difficult, slow and controlled mindset, and need energy and attention. Also, it can filter the biases and systematic errors in System 1.

In the context of online reviews, the elaboration likelihood model (ELM) distinguishes two systems of thinking: System 1: Peripheral route' and System 2: Central route. The distinction between the level of involvement and motivation needed to handle the quantitative and qualitative aspects of online product reviews drive persuasive dual-process models. Consumer attitudes are determined by the degree of elaboration effort the consumer tends to take, depending on the level of involvement at the time of processing the information (Petty, Cacioppo & Goldman, 1981). The information processing result forms the attitude of the consumer through one, the other, or some combination of the following two paths (Petty & Cacioppo, 1986). The peripheral heuristic route, when deep thought is unlikely, the attitude of the consumer is mainly influenced by cues and associated clues that are not directly related, such as models, colours or background music. These are called peripheral cues and the attitude formation path at this time is called the peripheral path. In this sense, those who have lacked the motivation or ability to read online reviews will be influenced, perhaps even unconsciously, by other clues and cues. For example, review valence and star ratings that are not directly related to argument quality (Cheung et al., 2009; Filieri & McLeay, 2014; Kim et al., 2018; Li & Zhan, 2011).

The central, systematic, route is used when there is elaboration of the argument – argument and counterargument. In this route, the attitude of the consumer is mainly influenced by the strength of specific items information. According to Petty and Cacioppo (1986), product information provides the central clues, and this attitude formation path is thus called the central route. A central cue refers to supporting material and ideas directly related to the quality of the message claim. Thus, those who are motivated, can read reviews and pay attention to argument quality cues such as the length of verbal content and the presence or not of two-sided information in other customers' reviews of products that they consider purchasing (Cheung et al, 2009; Filieri & McLeay, 2014; Kim et al., 2018; Li & Zhan, 2011) will follow this route to persuasion.

Moreover, argument quality has a positive impact on information adoption and purchase intention when the central route is used. One of the more significant elements of argument quality is the length of a comment (Kim et al., 2018). Another study has pointed out that the elaboration of the message can play an influential role in the message-generated persuasion process. In other words, an online review with engaging or even challenging information mitigates customer uncertainty about product quality and confidence in the decision-making process. Thus, the length of online reviews positively affects purchase intention while enhancing the elaborateness of the reviews (Liu & Park, 2015). The length of online reviews can be correlated to the reviewer's level of enthusiasm, which can affect customer judgment (Purnawirawan et al., 2012).

Another study has shown that consumers do not only refer to the average star ratings that websites offer, but also actually read and respond to reviews (Li & Zhan, 2011). Additionally, according to the ELM, high involvement consumers (or consumers in a high-involvement mode) are motivated to process information when specific information of interest is presented. Thus, when reviews become longer and more specific, the motivation of consumers to process information centrally will increase. Contrarily, when the length of online reviews is shorter, the peripheral route will be preferred, Together, these studies indicate the possibility that in terms of information processing, longer reviews could have a more significant effect on the perceived helpfulness of the reviews than shorter reviews.

Chapter 3: Hypotheses Development

3.1 Introduction

Having reviewed the relevant current literature, a research gap has been identified; only limited attention has been given to the mixed findings found in prior research regarding the effects of review valence (i.e., star ratings of online reviews) on the perceived helpfulness of the reviews. Thus, this research aims to develop and empirically address a hypothesis regarding the effects of variables of review valence, including the moderating role of consumer scepticism, “too good to be true,” and systematic information processing on the perceived helpfulness of the reviews.

3.2 Hypothesis 1: Negative effect of review valence

Previous studies have suggested that review valence is a key factor that affects the perceived helpfulness of the reviews (Forman, Ghose & Wiesenfeld, 2008; Li & Zhan, 2011; Liu & Park, 2015; Mudambi & Schuff, 2010; Purnawirawan, Pelsmacker & Dens, 2012; Sen & Lerman, 2007; Willemsen et al., 2011). As shown in the literature review, because of negativity bias and loss aversion, negative reviews could have a more significant main effect on the perceived helpfulness of the reviews than positive reviews.

Specifically, as an underlying mechanism of negativity bias, heuristic information processing plays a key role in hypothesis development. Prior studies that have noted the importance of adaptive human heuristics could be reflected by a negativity bias, and loss aversion. For example, Haselton and Nettle (2006) argue that the close relationship between heuristics and negativity bias and loss aversion is reinforced from an evolutionary psychological process because historically humans have experienced a greater cost when they misrepresent negative information. Krueger and Funder (2004) argue that heuristics can generate errors and reinforce both negative bias and loss aversion. Kahneman (2011) also argues there are unique patterns in people-generated errors that occur repeatedly in certain circumstances and at predictable levels. Therefore, various heuristics are the occurrence of System 1 according to the situation, and the systematic errors created by System 1’s mechanism often lead to biases. In summary, loss aversion is a typical example of heuristic-based judgment. (Kahneman, Knetsch, & Thaler, 1991).

Furthermore, as shown earlier in this research, the overwhelming amount of information on the Internet causes consumers to find it difficult in understanding an issue and effectively making decisions (Liu & Park, 2015). Therefore, consumers will tend to apply a heuristic information process that selects and examines online reviews that are considered to be important, rather than a systematic information processing that examines all online reviews one by one. Taken together, the helpfulness of negative hotel reviews will likely be more influenced by strengthened loss aversion through the heuristic information processing of consumers. Hence, Hypothesis 1 is proposed as follows:

H1: Review valence has a negative relationship with the perceived helpfulness of the reviews.

3.3 Hypothesis 2: The moderating role of consumer scepticism, "too good to be true."

Previous studies have suggested that the type of information is a key factor that affects the source credibility of the reviews (Doh & Hwang, 2009; Filieri & McLeay, 2014; Schindler & Bickart, 2005). As shown in the literature review, because of consumer scepticism, "too good to be true," two-sided reviews could have a greater main effect on the perceived helpfulness of the reviews than one-sided reviews.

Specifically, as the moderating role of consumer scepticism, 'too bad to be true,' plays a key role in this hypothesis development. Since fake reviews are mostly found to be positive, it is expected that the effect of "too bad to be true" is smaller than the effect of "too good to be true." Mukherjee, Liu and Glance (2012) argue that positive online reviews can bring significant financial benefits to organizations and individuals, so there is a temptation to make fake online reviews. Although it is difficult to detect fake online reviews, analyses using artificial intelligence revealed that many fake reviews were found among the five-star ratings online reviews. Thus, consumers have some scepticism about fake reviews, which are mostly found in positive reviews, and they consequently doubt the source credibility of positive reviews. In other words, consumer scepticism is greater in positive, 'too good to be true' reviews than in negative 'too bad to be true' reviews.

It seems that two-sided information not only improves source credibility and believability more than one-sided information, but also positively affects the formation of attitudes of customers by reducing consumer scepticism of, ‘too good to be true’ positive reviews more than negative ‘too bad to be true’ reviews. Hence, Hypothesis 2 is proposed:

H2: The type of information moderates the impact of review valence on the perceived helpfulness of the reviews. Specifically, when reviews are one-sided information, review valence has a strong negative impact on the perceived helpfulness of the reviews. However, when reviews are two-sided information, the negative impact of review valence on the perceived helpfulness of the reviews is relatively weak.

3.4 Hypothesis 3: The moderating role of systematic information processing

Previous studies have suggested that the length of review comment is a key factor that affects the perceived helpfulness of the reviews (Liu & Park, 2015; Mudambi & Schuff, 2010). As shown in the literature review section of this research, because of systematic information processing, longer reviews could have a more significant main effect on the perceived helpfulness of the reviews than shorter reviews.

Specifically, as loss aversion is a typical example of heuristic-based judgment (Kahneman, Knetsch, & Thaler, 1991), heuristic information processing plays a crucial role in this hypothesis’ development. Because of the close relationship between heuristic information processing, negativity bias and loss aversion, Hypothesis 1 was developed: Review valence, star ratings of online reviews, has a negative relationship with the perceived helpfulness of the reviews. In other words, the lower level of heuristic information processing, the lower level of the negativity bias and loss aversion. Thus, when the length of reviews is longer, not only will he heuristic information processing be minimized, but so too will negativity bias and loss aversion.

The helpfulness of negative hotel reviews will, then, be more influenced by strengthened loss aversion through the peripheral, heuristic information processing route, typically associated with shorter length reviews. In contrast, the helpfulness of negative hotel reviews will be less influenced by weakened loss aversion through the central route, where systematic information processing of longer length reviews may take place. Hence, the following hypothesis is proposed:

H3: Systematic information processing moderates the impact of review valence on the perceived helpfulness of the reviews. Specifically, when the length of reviews is shorter, review valence has a strong negative impact on the perceived helpfulness of the reviews. However, when the length of reviews is longer, review valence has a positive impact on the perceived helpfulness of the reviews.

3.5 Hypothesis 4: The moderating role of both consumer scepticism, “too good to be true” and systematic information processing

As shown above, in the development of Hypotheses 2 and 3, the presence of two-sided information, consumer scepticism, “too good to be true,” and the length of reviews could have an impact on perceived helpfulness. I expect both consumer scepticism and heuristic information processing to strengthen negativity bias and loss aversion. This study identifies consumer scepticism and systematic information processing as the two most likely important moderators. If two moderators fully describe negativity bias, loss aversion, then negativity bias and loss aversion will be eliminated as both consumer scepticism and heuristic information processing is eliminated. When consumer scepticism is sufficiently weakened by two-sided information, and simultaneously systematic information processing is activated by the longer length of reviews, negativity bias and loss aversion should be largely eliminated and can be turned into a positivity bias. In contrast, even with high-level systematic information processing enabled, if consumer scepticism is still present, then a negativity bias and loss aversion will still be observed.

Thus, only when reviews are two-sided with systematic information processing might negativity bias and loss aversion be eliminated, resulting in the helpfulness of positive hotel reviews becoming enhanced. Hence, Hypothesis 4 is proposed:

H4: Both length of comments and type of information moderate the impact of review valence on the perceived helpfulness of the reviews. Specifically, when reviews are two-sided, review valence has a positive relationship with the perceived helpfulness of the reviews with longer comments, whereas review valence has a negative impact on the perceived helpfulness of the reviews with shorter comments. Alternatively, when reviews are one-sided, review valence has a negative impact on the perceived helpfulness of the reviews in both shorter and longer comments.

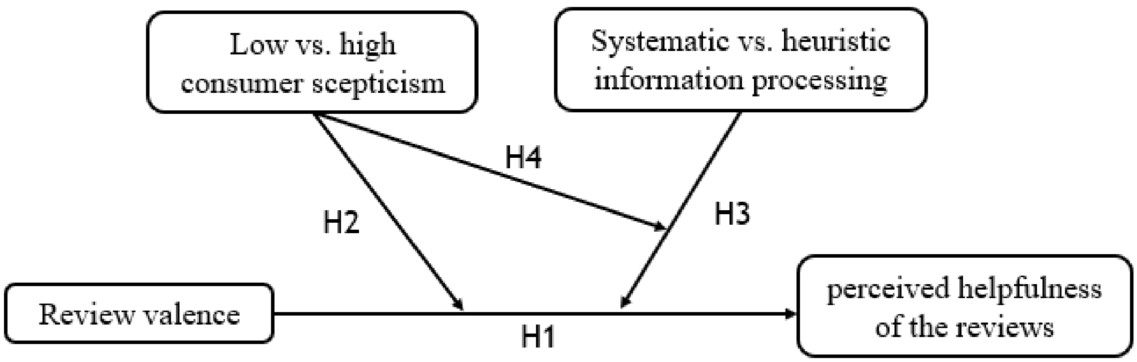

Figure 1: Conceptual framework

Chapter 4: Research Design and Data Analysis

4.1 Introduction

Having developed the hypotheses regarding the relationships of the important variables in this context, this section will illustrate how these relationships will be examined and tested empirically by using big data. In order to address the research purpose, this research offers an enhanced predictive effect than surveys or small sample size survey used in previous studies by utilizing big data (Siegel, 2013). The remainder of this chapter will elaborate on the methodology adopted in this research, the procedures of spatial regression discontinuity design, and trustworthiness of big data collection as well as how alternative explanations will be dealt with.

4.2 Big data

Laney (2001) states that the characteristics of big data can be defined with three aspects: the 3Vs of big data. First, the volume of data is large. Nowadays, the amount of data available is soaring, due to the digital revolution and the emergence of social media. For instance, in 2012, not only were 2.5 trillion bytes of data generated daily, but 90 per cent of the world's data had been produced between 2012 and 2014 (Wu et al., 2014). In 2017, the data generated every day not only increased to 2.5 quintillion bytes, but again, 90 per cent of the world's data had been produced between 2015 and 2017 (Marr, 2018). On top of that, 510,000 comments were posted, and 136,000 photos uploaded every minute on Facebook (The Social Skinny, 2019). Second, the velocity that is generated in real-time streamed data is very high. Finally, in addition to the refined structured data such as numbers, there is a variety of unstructured data such as text, images and video (Zikopoulos & Eaton, 2011).

As the role and importance of big data grow, new aspects are being added in addition to the traditional 3Vs (i.e., volume, velocity, and variety). Exhaustivity suggests that sample data can be collected across a whole population rather than a small subset of it (Mayer-Schonberger & Cukier, 2013). Thus, data pattern analysis using machine learning is becoming more important than population estimation in statistics. Also, value suggests that many insights can be derived from a single big data source (Marr, 2015). However, as the data surges, the most fundamental challenge in big data analysis is to explore large amounts of data and extract useful information or knowledge (Al-Jarrah et al., 2015). In the past, statistical techniques focused on identifying populations using a small amount of data, often used due to the limitations of data collection. However, with the explosion of data due to the development of the Internet, artificial intelligence techniques are emerging that capture meaningful information that the data itself has, rather than estimating it from the population. Therefore, big data that gather most of the population needs to be applied not only to statistical analysis but also a new analysis method using artificial intelligence. In other words, a real-world experiment with real-world data enables a variety of insights to be gained by utilising novel analytical methods, such as machine learning. Thus, in this study, traditional statistical analysis techniques of regression and Hayes PROCESS moderator method are used, but also machine learning, deep learning, a new frontier in artificial intelligence is used to extract meaningful information from the large database.

Paas and Morren (2018) point out problems with surveys using online panels (e.g. M-Turk, SurveyMonkey). Approximately one-third of the total respondents answered a question that should be skipped (i.e. 'Please do not answer if you are reading this'). Systematic biases may also occur in survey responses. Again, small and biased samples may affect the generalized meaning of findings (Schuckert, Liu & Law, 2015). Thus, the big data collected by this study to address the inconsistent results of existing studies are as follows: as for Booking.com, this study collected 2,036,260 reviews of 2,238 hotels in eight US cities via the web scraping software which this research developed. As for Agoda.com, this study has collected 60,266 reviews of 514 hotels throughout the US as of November 2018 via the web scraping software which the author developed.

One concern is that the difference between the total amount of data on Booking.com and Agoda.com is more a distraction than a plus. Despite scraping all reviews published with web scraping software, Booking.com has approximately 33 times more reviews than Agoda.com. For a fair comparison, choosing another famous hotel booking site like TripAdvisor.com, which has a higher total number of reviews than Agoda.com, could be an alternative. However, unlike Booking.com and Agoda.com, which are offered by the same company and offer the same 10-point evaluation form, TripAdvisor.com, which is offered by another company offers a 5-point evaluation form. Not only does the difference in ratings appear to produce a greater bias than the difference in the total number of reviews, but Agoda.com's approximately 60,000 reviews could be considered sufficient as they provide more data than the datasets used in many previous studies. Thus, despite the difference in the total amount of data, it was decided to scrape online reviews from Booking.com and Agoda.com.

4.3 Spatial regression discontinuity design for Big data

Apart from big data, this study applied a spatial regression discontinuity design to two online travel sites with reviews based on the type of information which is critical to this study. Even though Agoda.com and Booking.com, two of the largest hotel reservation sites in the United States as well as worldwide, are operated by the same company (Booking Holdings, 2019), their online review input systems for consumers are different. The main difference is that Agoda.com enables consumers to enter one-sided information into their reviews, whereas Booking.com enables consumers to enter both one-sided and two-sided information into their reviews. Specifically, Agoda.com provides only one input box (one-sided information) for general comments on the evaluation form, whereas Booking.com provides two input boxes (two-sided information) for dividing positive and negative comments on the evaluation form. Also, as it is not compulsory to enter both positive and negative comments (two-sided information) on Booking.com, reviewers may leave: 1) both positive and negative comments (two-sided information) or 2) only one side of positive or negative comments (one-sided information) or 3) no comments.

Also, the online review text was analyzed by sentiment analysis using natural language processing techniques to uncover differences in the characteristics of one-sided and two-sided information. The IBM SPSS Modeler Text Analytics, which is one of the widely used text mining tools in the hotel industry (Lau et al., 2005), was used in this sentiment analysis. Through sentiment analysis nodes coding, the total number of positive words and negative words in the review text written in English were captured as new variables. Specifically, first, for one-sided reviews of Agoda.com, it showed that the mean of the total number of positive words was 4.74 and the mean of the total number of negative words was 2.20. Second, for one-sided reviews of Booking.com, it showed that the mean of the total number of positive words was 2.46 and the mean of the total number of negative words was 1.56. Third, for two-sided reviews of Booking.com, it showed that the mean of the total number of positive words was 2.68 and the mean of the total number of negative words was 2.18. Finally, by comparing the ratio of negative words to positive words, it verified that one-sided information is relatively weighted information (Ratio_one-sided_Agoda.com = 46.4%, Ratio_one-sided_Booking.com = 63.4%), whereas two-sided information is relatively balanced information (Ratio_two-sided_Booking.com = 81.3%).

Generally, a randomized clinical trial is one of the best data analysis methods to reveal causal relationships and is an artificial way in which researchers design experiments and collect and analyse data through intervention. However, it is difficult to carry out such experiments on a large scale in terms of budget and manpower. Therefore, analysts have turned to a spatial regression discontinuity design, one of the natural experiments and quasi-experimental methods, to collect data using an artificially created boundary line (Keele & Titiunik, 2015). Therefore, when analysing big data without experiments in this research, it is vital to collect data from natural experimental comparison groups using the spatial regression discontinuity design. For example, if different electric utilities supply different prices based on a boundary line at the same time, it is possible to analyse how consumers reacted to changes in electricity prices (Ito, 2014). Applying geographic boundaries as online boundaries, the two hotel reservation sites provide unique and relevant research data for this research by providing different online hotel evaluation forms. Thus, as shown in Table 4, this research investigates the spatial regression discontinuity between one-sided and two-sided reviews on Booking.com as well as Agoda.com and Booking.com reviews.

Table 4: Spatial regression discontinuity design

Type of information	**Agoda.com**	**Booking.com**
One-sided reviews	60,258	338,490
Two-sided reviews	NA	872,685
Zero length reviews	8	825,085
Total Sample size	60,266	2,036,260

4.4 Big Data Sample

4.4.1 Agoda.com

The author collected reviews from Agoda.com, one of the most popular global online travel agencies for hotels, vacation rentals, flights and airport transfer. Agoda.com allows consumers who have booked hotels through Agoda.com to post online reviews, which are used as an indicator of the quality at hotels worldwide.

While Agoda.com's primary offering to consumers is the online Booking service of hotels and the rating of them through online reviews, for this research I gathered all publicly available online reviews in the US until Nov 2018, collecting a total of 60,275 reviews. In addition to the review text, metadata containing information about star ratings of a review, in an interval scale from 2 to 10, helpfulness of a review, name of a writer, nationality of writer, nights and date of stay, and type of travel, were also collected (see Figure 2). Specifically, as shown in Figure 3, Agoda.com aggregates the ratings given to individual aspects to form the overall score. The 60,266 reviews in this data set are written by customers originating from over two hundred countries, providing ratings for 513 hotels located in the US.

Figure 2: Publicly available online reviews of Agoda.com

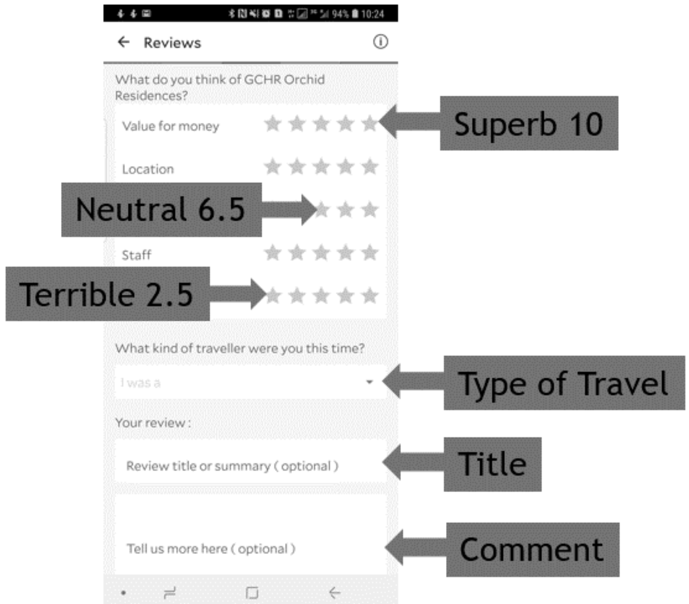

Figure 3: Individual aspects rating system of Agoda.com

Table 5 provides the descriptive statistics of the Agoda.com sample. First, an average perceived helpfulness of online reviews is approximately 2% (SD = .001, Min = 0, Max = 8). Second, the average overall rating for all reviews in this sample was relatively positive and left-skewed (M = 7.59, SD =. 008, Min = 2, Max = 10). Specifically, both Table 6 and Figure 4 provide a distribution of Agoda.com hotel reviews' star ratings in this sample. Lastly, an average length of online reviews is 215 characters (SD = .942, Min = 0, Max = 2139). Although the data did not follow the normal distribution, big data used in other studies also found that their data did not follow the normal distribution. For instance, this research found that the distribution of review star ratings is left-skewed. Another big data usage, which analysed over 1.2 million Booking.com online hotel reviews, it was also found that the distribution of review star ratings is left-skewed (Mariani & Borghi, 2018).

Table 5: Descriptive statistics of Agoda.com

IV	N	Mean	Std. Error of Mean	Minimum	Maximum	Kurtosis	Skewness
Helpfulness of Reviews	60266	.02	.001	0	8	237.171	11.552
Star Ratings	60266	7.587	.0076	2	10	.191	-.812
Length of Comments	60266	215.07	.942	0	2139	9.692	2.53

Sqrt Length of Comments	60266	12.9838	.02778	0	46.25	1.065	.941

Table 6: Frequency of star ratings category of Agoda.com

Star Ratings Category	**Frequency**	**Valid Percent**	**Cumulative Percent**
2, 2.5	806	1.3	1.3
2.5, 3	476	.8	2.1
3, 3.5	836	1.4	3.5
3.5, 4	1687	2.8	6.3
4, 4.5	998	1.7	8.0
4.5, 5	1154	1.9	9.9
5, 5.5	2276	3.8	13.7
5.5, 6	4447	7.4	21.0
6, 6.5	2624	4.4	25.4
6.5, 7	3010	5.0	30.4
7, 7.5	6024	10.0	40.4
7.5, 8	4434	7.4	47.7
8, 8.5	10235	17.0	64.7
8.5, 9	4397	7.3	72.0
9, 9.5	6978	11.6	83.6
9.5, 10	9884	16.4	100.0
Total	60266	100.0	

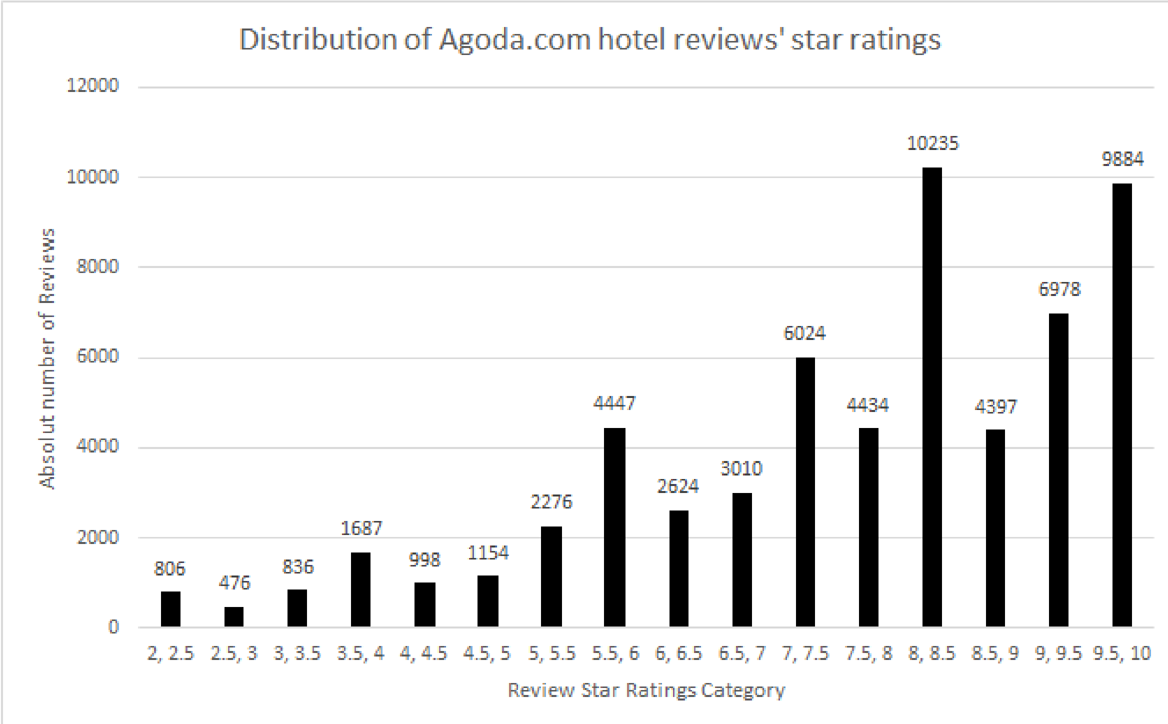

Figure 4: Distribution of Agoda.com hotel reviews' star ratings, US, Sep 2007 - Nov 2018

4.4.2 Booking.com

Apart from Agoda.com, I collected reviews from Booking.com, one of the most popular global travel fare aggregator website and travel metasearch engine for lodging reservations. Because of the unexacting availability of data content to researchers, Booking.com has been heavily used in the literature of online hotel reviews (Mariani & Borghi, 2018). Booking.com allows consumers who have booked hotels through Booking.com to post online reviews, which are used as an indicator of the quality of hotels worldwide.

While Booking.com's primary offering to consumers is the online Booking service of hotels and the rating them through online reviews, for this research I gathered all publicly available online reviews in the major eight US cities until Nov 2018, collecting a total of 2,036,260 reviews. In addition to the review text divided into positive and negative, metadata containing information about star ratings of a review (on an interval scale from 2.5 to 10), helpfulness of a review, name of the writer, nationality of writer were also collected (see Figure 5). Specifically, as shown in Figure 6, Booking.com aggregates the ratings given to individual aspects to form the overall score. The 2,036,260 reviews in this data set are written by customers originating from over two hundred countries, providing ratings for 2,238 hotels located in the major eight US cities.

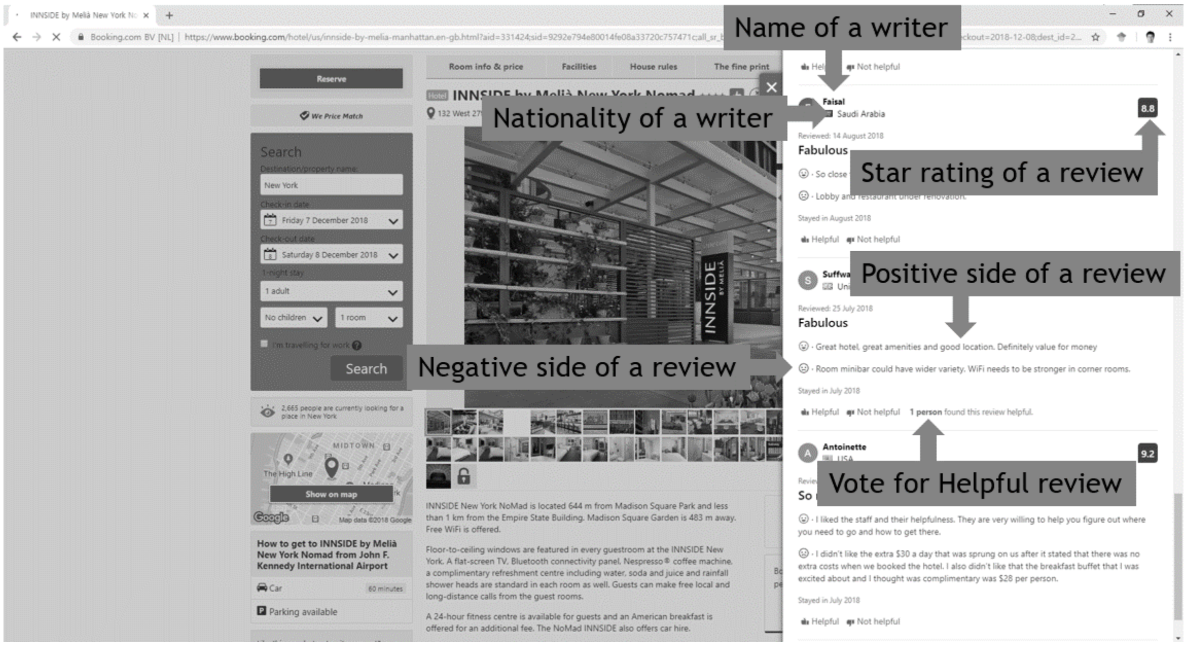

Figure 5: Publicly available online reviews of Booking.com

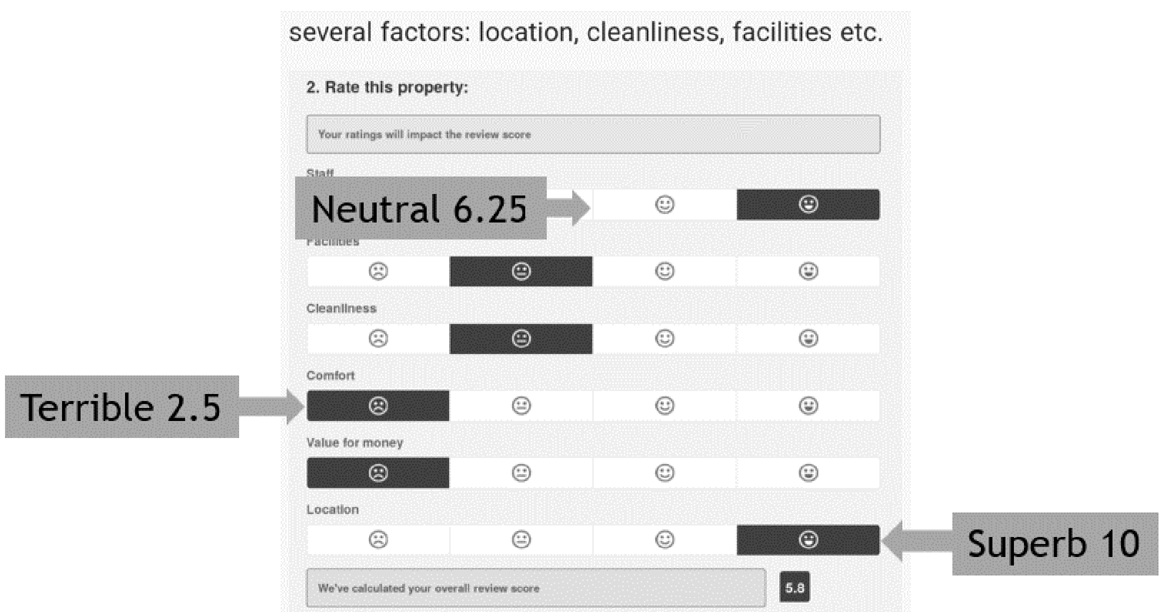

Figure 6: Individual aspects rating system of Booking.com

Table 7 provides the descriptive statistics of the Booking.com sample: First, an average perceived helpfulness of online reviews is approximately 4% (SD = .000, Min = 0, Max = 63). Second, the average overall rating for all reviews in this sample was relatively positive and left-skewed (M = 7.911, SD =. 001, Min = 2.5, Max = 10). Specifically, Figure 7 in trustworthy of data collection provides a distribution of Booking.com hotel reviews' star ratings in this sample. Lastly, an average length of online reviews is 123 characters (SD = .144, Min = 0, Max = 5869). Specifically, an average length of positive side of online reviews is 63 characters (SD = .083, Min = 0, Max = 3365), whereas an average length of negative side of online reviews is 60 characters (SD = .097, Min = 0, Max = 2717). However, However, 825,085 reviews (40.5%) of this data showed no comments in the online reviews (length of comments = 0) – these reviews, without comments, were excluded from further analysis.

Table 7: Descriptive statistics of Booking.com

IV	**N**	**Mean**	**Std. Error of Mean**	**Minimum**	**Maximum**	**Kurtosis**	**Skewness**
Helpfulness of Reviews	2036260	.04	0	0	63	1882.517	19.558
Star Ratings	2036260	7.911	.0013	2.5	10	.164	-.897
Length of Comments	2036260	122.67	.144	0	5869	24.469	3.785
Length of Positive Side	1211175	106.18	.124	1	3365	34.265	4.441
Length of Negative Side	872685	138.86	.196	1	2717	23.232	3.899
Sqrt Length of Comments	2036260	7.6280	.00563	0	76.61	.841	.961

4.4.3 Trustworthy of big data collection

To determine whether the web-scraping software the author developed is trustworthy, this research tests its source data against already published source data. According to the big data usage, which analysed over 1.2 million Booking.com online hotel reviews, it was found that the distribution of review star ratings is left-skewed (Mariani & Borghi, 2018). When comparing the distribution of their over 1.2 million review star ratings with the distribution of the over two million review star ratings collected here, the trustworthiness of the data is supported due to exhibiting the same patterns of distribution in the graphs observed. Specifically, comparing Figure 7 and Figure 8, the same patterns were observed in that both distributions of review star ratings is left-skewed, and the same relatively high frequency was observed in both category (5, 5.5), category (7.5, 8) and category (9.5, 10). This is presumably because Booking.com uses a four-level item rating system, with many users consistently giving the same score in each rating category.

Table 8: Frequency of star ratings category of Booking.com

Star Ratings Category	**Frequency**	**Valid Percent**	**Cumulative Percent**
2.5, 3	46141	2.3	2.3
3, 3.5	22766	1.1	3.4
3.5, 4	28142	1.4	4.8
4, 4.5	30736	1.5	6.3
4.5, 5	35182	1.7	8
5, 5.5	110271	5.4	13.4
5.5, 6	61181	3	16.4
6, 6.5	72537	3.6	20
6.5, 7	89699	4.4	24.4
7, 7.5	119175	5.9	30.2
7.5, 8	365923	18	48.2
8, 8.5	154403	7.6	55.8
8.5, 9	150320	7.4	63.2
9, 9.5	168284	8.3	71.4
9.5, 10	581500	28.6	100
Total	2036260	100	

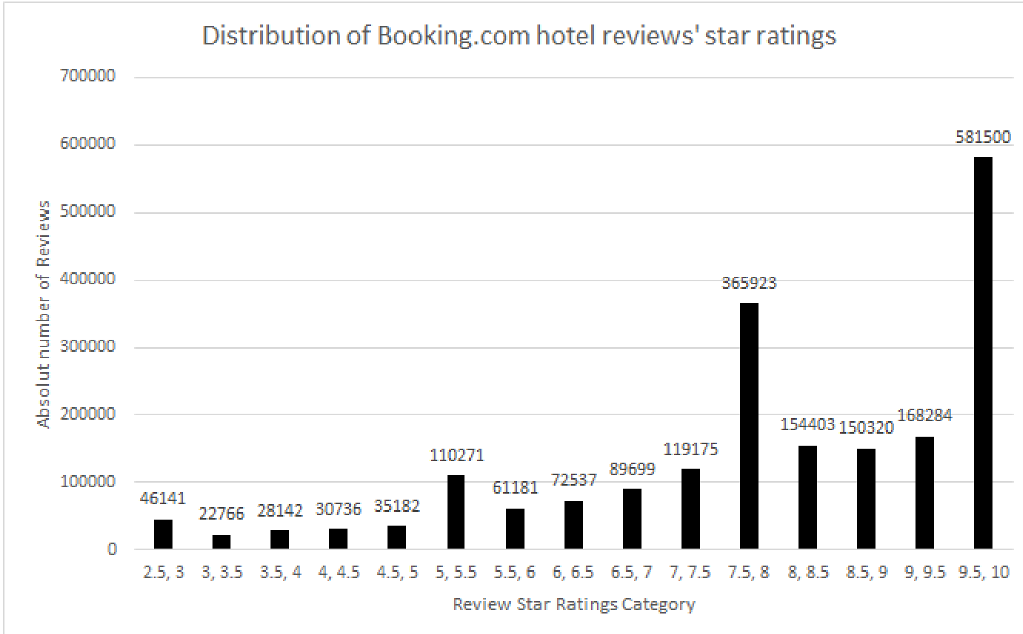

Figure 7: Distribution of Booking.com hotel reviews' star ratings, US, Nov 2016 - Nov 2018

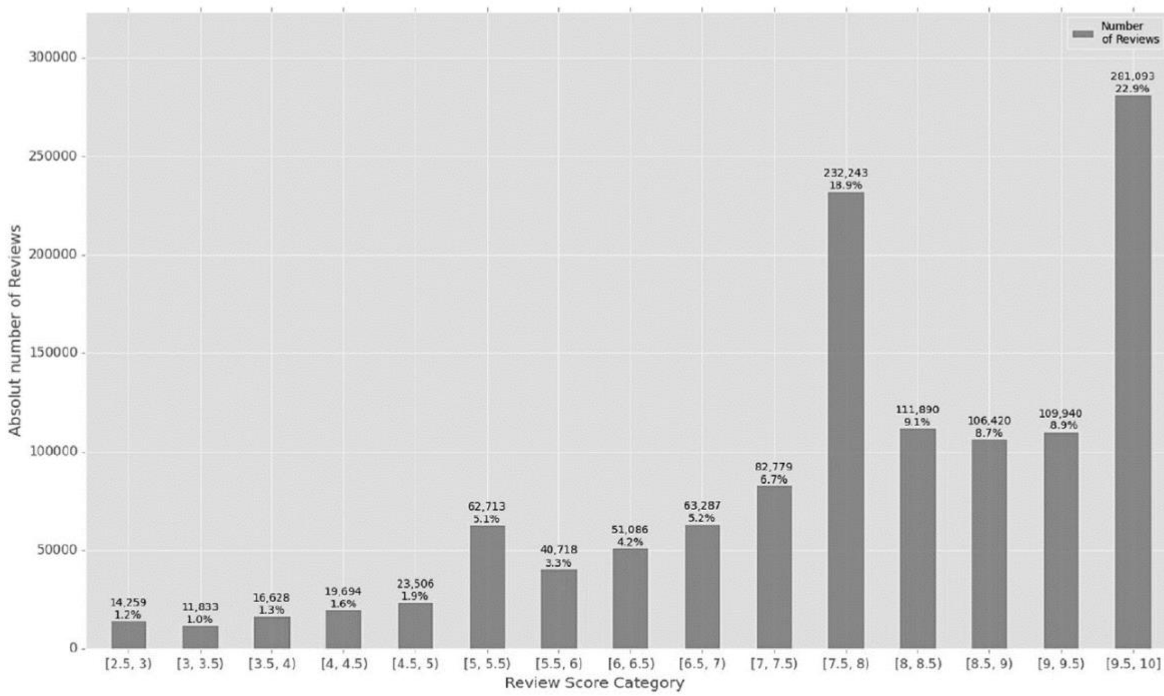

Figure 8: Distribution of Booking.com hotel reviews' scores, London, Jan 2015 - Jan 2017 (Mariani & Borghi, 2018)

4.4.4 Assumption of normality

Among the independent variables used in this study, the 'length of comments' variable did not meet the normality assumption criteria. Thus, it was square-root transformed if necessary. The dependent variable, ‘helpfulness of reviews’, did not meet the normality assumption criteria. Nevertheless, violation of normality would not be a big issue because of Li et al. (2012). Details are as follows:

In general, before analysing linear regression, the normality assumption should be confirmed by checking the skewness and kurtosis to see if the variables have a normal (Gaussian) distribution. Skewness indicates whether the distribution of the data is skewed left or right around the mean, and kurtosis indicates whether the distribution of the data has a sharp or gentle slope around the mean (Hair et al., 2013). To meet the normal distribution criteria, skewness values should be ± 2 intervals and kurtosis value should be ± 7 intervals (West, Finch & Curran, 1995). The normal distribution analysis for the three variables used in this study is as follows.

First, as shown in Table 5 and Table 7, star ratings as the independent variable met the normal distribution criteria of skewness values (Agoda.com = -.812 and Booking.com = -.897) because they were ± 2 intervals and kurtosis values (Agoda.com = .191 and Booking.com = .164) were ± 7 intervals (West, Finch & Curran, 1995).

Second, length of comments as an independent variable did not meet the normal distribution criteria in that the skewness values (Agoda.com = 2.53 and Booking.com = 3.785) were not ± 2 intervals and kurtosis values (Agoda.com = 9.692 and Booking.com = 24.469) were not ± 7 intervals.

When the normality assumption of independent variables was required, the ‘length of comments’ variable was square-root (sqrt) transformed to ‘sqrt length of comments’ variable, and it met the normal distribution criteria as the skewness values (Agoda.com = .941 and Booking.com = .961) were ± 2 intervals and kurtosis values (Agoda.com = 1.065 and Booking.com = .841) were ± 7 intervals (West, Finch & Curran, 1995). Thus, ‘sqrt length of comments’ variable was used in the linear regression and negative binomial regression analysis, which required the normality assumption of independent variables (Hair et al., 2013).

When the normality assumption of independent variables was not required, ‘length of comments’ variable was used without sqrt transformation. As for Hayes PROCESS macro, the generation of confidence intervals for the significance test was analyzed using the bootstrap procedure, where non-normality is assumed as the entire model or a single path is tested with PROCESS macro (Hayes, 2018). Bi-logistic regression and deep learning analysis also do not require the normality assumption of independent variables based on ordinary least squares algorithm, which do not require a linear relationship between the independent and dependent variables (Guijarro-Berdiñas et al., 2007; Statistics Solutions, 2019).

Finally, 'helpfulness of reviews' as a dependent variable also did not meet the normal distribution criteria in that skewness values (Agoda.com = 11.552 and Booking.com = 19.558) were not ± 2 intervals and kurtosis values (Agoda.com = 237.171 and Booking.com = 1882.517) were not ± 7 intervals. In addition, the square root transformed 'sqrt helpfulness of reviews' also fails to meet the normal distribution criteria of that skewness values (Agoda.com = 7.272 and Booking.com = 5.594) were not ± 2 intervals and kurtosis values (Agoda.com = 55.045 and Booking.com = 35.119) were not ± 7 intervals (West, Finch & Curran, 1995).

Nevertheless, in large samples, by the central limit theorem and the law of large numbers, even if the dependent variable does not meet the normal distribution criteria, it is considered valid to use linear regression techniques as estimated parameters remain robust (Li et al., 2012). Since this study analyses big data, 'helpfulness of reviews' will be used as a dependent variable for linear regression and PROCESS macro analysis without transformation. Also, because of negative binomial regression analysis using count dependent variable, and bi-logistic regression and deep learning analysis using a categorical dependent variable, they do not require the normality assumption of dependent variables (Gardner, Mulvey & Shaw, 1995; Guijarro-Berdiñas et al., 2007; Statistics Solutions, 2019).

4.5 Analysis of Hypothesis 1:

The purpose of the first analysis is to test the hypothesis that the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews. With the setting to test out the prediction, this research analyses online hotel reviews generated by consumers on Booking.com and Agoda.com.

4.5.1 Data Set

This thesis analysed two full data sets: one composed of 2,036,260 online reviews for 2,238 hotels in eight US cities and a second composed of 60,266 online reviews for 514 hotels throughout the US. The reviews were posted from 2007 through 2018 and referenced a total of 2,752 hotels listed on both Booking.com and agada.com website. Each post contains the descriptive text of the review, the name of the hotel reviewed, the star rating and the helpfulness vote.

4.5.2 Helpfulness Star ratings Data Collection

As for the descriptive statistics of star ratings and helpfulness vote, the star rating is overwhelmingly biased toward positive reviews on both sites ($M\_{\text{Booking.com}}$ = 7.91, Min = 2.5, Max = 10.0, SD = .92, and $M\_{\text{Agoda.com}}$ = 7.59, Min = 2.0, Max = 10.0, SD = 1.86). Moreover, as for the hotel rating scale of both sites, many previous studies have suggested that the rating scale of Booking.com was 0-10 or 1-10. However, Booking.com rates the hotels based on a four-level item, a 2.5 to 10 rating scale is used, the exact average is 6.25 points, not 5 points (Mellinas et al., 2015). Agoda.com rates the hotels based on a five-level scale (from 2.0 to 10), where the exact average is 6.0 points. Therefore, this study distinguishes between positive and negative reviews based on 6.25 points for Booking.com and 6.0 points for Agoda.com. Specifically, to unify different centre points, the formula (i.e., (star rating of Agoda.com -6)* .9375 + 6.25 = star rating of Booking.com) can be used that transforms the average rating of Agoda.com from a five-level item to a four-level item (i.e., 7.59 star rating of Agoda.com = 7.74 star rating of Booking.com). As for the helpfulness vote, both sites put helpful buttons at the bottom of each review so that consumers can evaluate whether the review was helpful.

Specifically, only a small number of reviews, less than 5% on both sites, are rated as helpful by consumers ($M\_{Booking.com}$ = 4.46%, Min = 0, Max = 63, SD = .28, and $M\_{Agoda.com}$ = 2.29%, Min = 0, Max = 8, SD = .18).

4.5.3 Hierarchical Regression Analysis Results

To formally test the negativity bias, hierarchical regression analysis was conducted (Hair et al., 2013), with 'star ratings' and 'sqrt length of comments' as independent variables, and 'perceived helpfulness of the reviews' as the dependent variable. For Booking.com there is another independent variable that will be used for further analysis, but only includes the same variables used on agoda.com for direct comparison of the two sites. Hierarchical regression analysis is the most commonly used technique to validate mediating and moderating effects. This study can separate the stages of independent variables and analyse them in two or more stages. By adding variables in several stages, this study can identify how the explanatory power changes in stages (Gelman & Hill, 2006).

As for Booking.com, the regression model is significant in both model 1 (F (1, 2036258) = 3148.20, p < .001) and model 2 (F (2, 2036257) = 39606.894, p < .001). The explanatory power of the regression model is .2% in model 1 (R^2 = .002, Adjusted R^2 = .002) and 3.7% in model 2 (R^2 = .037, Adjusted R^2 = .037). The Durbin-Watson statistic is .496, showing a value of less than 1, which is not considered suitable for the assumption of residual independence (Field, 2013). This study will verify the residual independence assumption problem through additional negative binomial regression analysis. As shown in the coefficients output in collinearity statistics, all VIF values are 1.022, meaning that the VIF values obtained are between 1 to 10. Therefore, multicollinearity symptoms are not observed.

Specifically, in model 1, 'star ratings' (β = -.039, t = -56.11, p <. 001) significantly influence perceived helpfulness of the reviews, suggesting the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews. In model 2, the control variables are also significant; 'star ratings' (β = -.011, t = -16.04, p < .001) and 'sqrt length of comments' (β = .192, t = 275.59, p < .001) significantly influence perceived helpfulness of the reviews, suggesting the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews.

As for Agoda.com, the regression model is significant in both model 1 (F (1, 60264) = 219.70, $p < .001$) and model 2 (F (2, 60263) = 215.51, $p < .001$). The explanatory power of the regression model is .4% in model 1 ($R^2 = .004$, Adjusted $R^2 = .004$) and .7% in model 2 ($R^2 = .007$, Adjusted $R^2 = .007$). The Durbin-Watson statistic is 1.878, showing a value close to 2, which is considered suitable for the assumption of residual independence (Field, 2013). As shown in the coefficients output in collinearity statistics, all VIF values are 1.013, meaning that the VIF values obtained are between 1 to 10, so multicollinearity is not an issue. Specifically, in model 1, 'star ratings' ($\beta = -.06$, $t = -14.82$, $p < .001$) significantly influences perceived helpfulness of the reviews suggesting the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews. In model 2, control variables were significant, 'star ratings' ($\beta = -.054$, $t = -13.1$, $p < .001$) and 'sqrt length of comments' ($\beta = .059$, $t = 14.51$, $p < .001$) significantly influence perceived helpfulness of the reviews suggesting the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews.

Overall, hierarchical regression analysis of both sites shows a negativity bias effect. However, the β value of Agoda.com is greater than the β value of Booking.com. Therefore, it is expected that the two-sided information provided only on Booking.com was weakening the negativity bias effect. Further detailed analysis of the type of information will be conducted in the second analysis. Table 9 and Figure 9 show a summary of the results for Booking.com. Also, Table 10 and Figure 10 show a summary of the results for Agoda.com.

Table 9: Results of first analysis: Hierarchical regression for Booking.com

Model	IV	Unstandardized Coefficients		Standardized Coefficients	t	P value
		B	Std. Error	Beta		
1	(Constant)	.090	.001		108.001***	.000
	Star Ratings	-.006	.000	-.039	-56.109***	.000
F (1,2036258) =3148.20 (p<.001), R^2=.002, Adjusted R^2=.002						
2	(Constant)	.007	.001		8.069***	.000
	Star Ratings	-.002	.000	-.011	-16.037***	.000
	Sqrt Length of comments	.007	.000	.192	275.587***	.000
F (2,2036257) =39606.894 (p<.001), R^2=.037, Adjusted R^2=.037						

Figure 9: Main effect of star ratings for Booking.com

Table 10: Results of first analysis: Hierarchical regression for Agoda.com

Model	IV	Unstandardized Coefficients		Standardized Coefficients	t	P value
		B	Std. Error	Beta		
1	(Constant)	.066	.003		22.024***	.000
	Star Ratings	-.006	.000	-.060	-14.822***	.000
F (1,60264) =219.7 (p<.001), R^2=.004, Adjusted R^2=.004						
2	(Constant)	.041	.003		12.030***	.000
	Star Ratings	-.005	.000	-.054	-13.101***	.000
	Sqrt Length of comments	.002	.000	.059	14.511***	.000
F (2,60263) =215.51 (p<.001), R^2=.007, Adjusted R^2=.007						

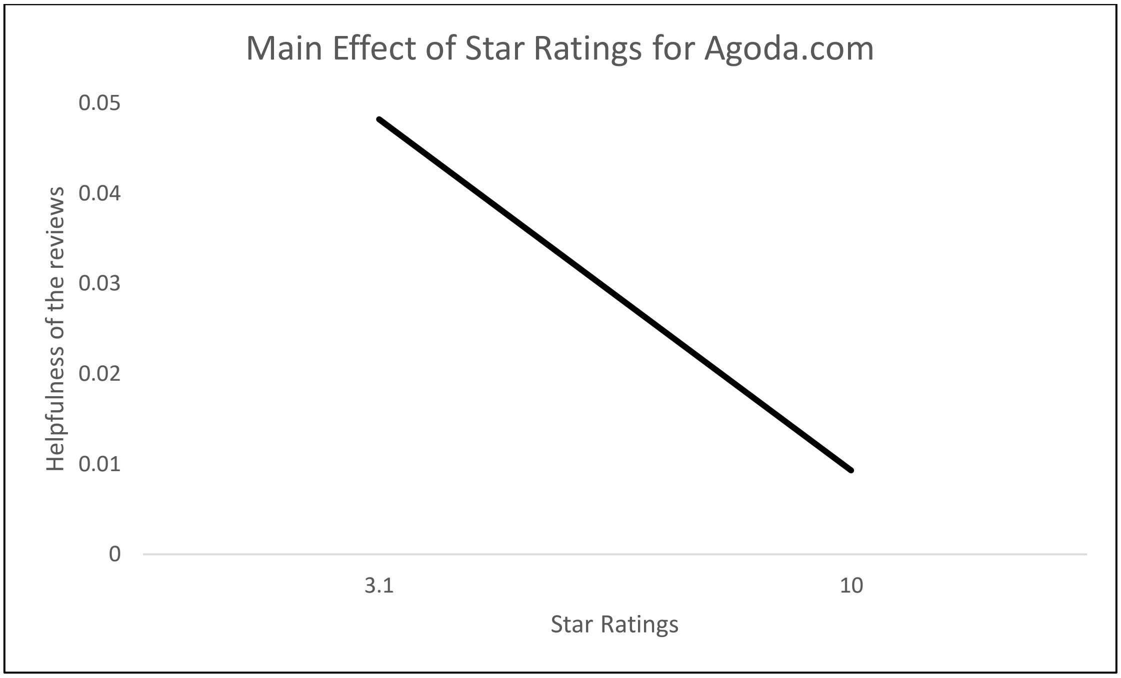

Figure 10: Main effect of star ratings for Agoda.com

4.5.4 Additional Negative Binomial Regression Analysis Results

As shown above, when the continuous variable is analysed as a linear regression model as a dependent variable, the model is observed might be inadequate due to the residual independence assumption problem. Helpfulness votes could be treated as continuous as well as frequency data. Therefore, further analysis can use Poisson regression, which is often used when the dependent variable is count data. However, Poisson regression analysis is a statistical analysis module that performs regression analysis of the generalized linear model assuming that dependent variables follow the Poisson distribution (Cameron & Trivedi, 1990). Dependent variables along the Poisson distribution should have the same mean and variance, and the dependent variable in this study observed an overdispersion problem where the variance is above the mean ($M\_{Booking.com}$ = .04, $Variance\_{Booking.com}$ = .077, $M\_{Agoda.com}$ = .02, $Variance\_{Agoda.com}$ = .031). When there is an overdispersion problem, the analysis can be performed with negative binomial regression (Gardner, Mulvey & Shaw, 1995). Thus, this research conducted additional negative binomial regression analysis with the dependent variable as perceived helpfulness of the reviews and the independents as star ratings and sqrt length of comments.

As for Booking.com, the results of the Omnibus test shows that goodness-of-fit for the negative binomial regression model is significant ($\chi 2 = 94433.777$, $p < .001$), and the Pearson Chi-Square is 1.024, showing a value close to 1, which is considered a suitable model that does the fit the data. Second, as shown in results of the parameter estimates, this research finds that the main effects of star ratings' ($B = -.031$, SE = .0018, Wald = 315.402, $p < .001$) and sqrt length of comments ($B = .105$, SE = .0004, Wald = 87562.874, $p < .001$) are significant. In other words, the higher the star ratings, the lower the helpfulness votes, and the longer the review, the higher the helpfulness votes. Third, analysis finds that the number of helpfulness votes decreases by 3.1% with a one-point increase in star ratings ($OR = .969$) suggesting the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews.

As for Agoda.com, the results of the Omnibus test show that goodness-of-fit for the negative binomial regression model is significant ($\chi 2 = 484.842$, $p < .001$), and the Pearson Chi-Square is 1.247, showing a value close to 1, which is considered a suitable model that does fit the data. Second, as shown in the results of parameter estimates, analysis finds that the main effects of star ratings ($B = -.187$, SE = .0131, Wald = 203.251, $p < .001$) and sqrt length of comments ($B = .053$, SE = .0034, Wald = 249.097, $p < .001$) are significant. This means that the higher the star ratings, the lower the helpfulness votes, and the longer the review, the higher the helpfulness votes. Third, analysis finds that the number of helpfulness votes decreases by 17% with a one-point increase in star ratings ($OR = .83$); suggesting the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews.

Overall, negative binomial regression analysis of both sites also shows the negativity bias effect. However, the *OR* value of Agoda.com is greater than the *OR* value of Booking.com. Therefore, it is reasonable to assume that the two-sided information provided only on Booking.com has weakened the negativity bias effect. Further detailed analysis of the type of information will be conducted in the second analysis. Tables 11 and 12 show a summary of additional results for both Booking.com and Agoda.com.

Table 11: Results of additional first analysis: Negative binomial regression analysis for Booking.com

	B	**S. E.**	**Wald**	**P value**	***OR***	**95% CI**
Star Ratings	-.031	.0018	315.402***	.000	.969	(.966 ~.973)
Sqrt Length of comments	.105	.0004	87562.874***	.000	1.111	(1.11~1.112)
(Intercept)	-4.125	.0156	69907.985***	.000	.016	(.016~.017)
Pearson $\chi2$=1.024, Omnibus Test: Likelihood Ratio $\chi2$=94433.777 (p<.001)						

Table 12: Results of additional first analysis: Negative binomial regression analysis for Agoda.com

	B	**S. E.**	**Wald**	**P value**	***OR***	**95% CI**
Star Ratings	-.187	.0131	203.251***	.000	.83	(.809~.851)
Sqrt Length of comments	.053	.0034	249.097***	.000	1.054	(1.047~1.061)
(Intercept)	-3.21	.114	793.362***	.000	.04	(.032~.05)
Pearson $\chi2$=1.247, Omnibus Test: Likelihood Ratio $\chi2$=484.842 (p<.001)						

4.6 Alternative Explanations for Hypothesis 1

The results of the analyses support the predictions that (i) the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews, and (ii) this effect is driven by a negativity bias, loss aversion. However, the observed negative relationship may have been caused by other factors associated with the bias towards positive reviews, such as the scarcity effects (Lynn, 1989) or cultural diversity in reviewers across cities (Ferris, Frink & Galang, 1993). I next test and report for the significance of these possible alternative explanations.

4.6.1 Scarcity Effects

Firstly, one possible alternative explanation is the scarcity effect. Several reports have shown that online reviews consist of a majority of positive reviews and a minority of negative reviews. Chevalier and Mayzlin (2006) found that the majority of online reviews of books are positive at both Amazon.com and Barnesandnoble.com. Liu and Park (2015) found that the average star rating was 4.28 out of 5 based on 5,090 online restaurant reviews from Yelp.com. As shown in both Figures 7 and 8, this study also found that online hotel reviews are biased towards positive reviews at both Booking.com and Agoda.com. Furthermore, scarcity effects are seen in goods with more demand than supply. According to economic law, scarce goods are more costly than goods with abundance (Lynn, 1989). Thus, as consumers might want positive and negative reviews at an equal ratio, a relatively small number of negative reviews will be more valuable. Consumers can, therefore, assess scarce negative reviews as more helpful.

As for Booking.com (see Table 13), to an additional test of the scarcity effects explanation, hierarchical regression analysis was conducted (Hair et al., 2013), with star ratings as the independent variable, perceived helpfulness of the reviews as the dependent variable, total number of reviews by hotels and the overall star ratings by hotels as the control variables, controlling for the following scarcity effects markers: relatively high number of positive reviews and relatively high overall star ratings.

Since the analysis is intended to verify alternative explanations, only model 3 results are interpreted. Booking.com, overall model is significant (R^2 = .002, F (3, 2034089) = 1190.94, p < .001). The explanatory power of the regression models does not increase (model 1: R^2 = .002, Adjusted R^2 = .002, model 2: R^2 = .002, Adjusted R^2 = .002, model 3: R^2 = .002, Adjusted R^2 = .002). Based on the coefficients output, collinearity statistics, all VIF values are of less than 1.329, so multicollinearity symptoms are not observed. The Durbin-Watson statistic is .466, showing a value of less than 1, which is not considered suitable for the assumption of residual independence (Field, 2013). This study will verify the residual independence assumption problem through additional negative binomial regression analysis. Specifically, even though total number of reviews by hotels (β = .010, t = 13.08, p < .001) and the overall star ratings by hotels (β = -.009, t = -11.59, p < .001) influence perceived helpfulness of the reviews, star ratings still significantly influences perceived helpfulness of the reviews (β = -.034, t = -43.80, p < .001).

Moreover, I conducted additional negative binomial regression analysis (Gardner, Mulvey & Shaw, 1995) with the same DV, IV and CVs as hierarchical regression analysis (see Table 14). The results of the Omnibus test shows that goodness-of-fit for the negative binomial regression model is significant ($\chi2 = 5380.394$, $p < .001$), and the Pearson Chi-Square is 1.613, showing a value close to 1, which is considered a suitable model that does fit the data. Second, as shown in results of parameter estimates, analysis finds that the main effects of star ratings ($B = -.1$, SE = .0018, Wald = 3015.485, $p < .001$), overall ratings ($B = -.056$, SE = .004, Wald = 197.333, $p < .001$) and number of reviews ($B = .000$, SE = .000, Wald = 208.621, $p < .001$) are significant. Thus, the higher the star ratings, the lower the helpfulness votes. Third, this research finds that the number of helpfulness votes decreases by 9.6% with a one-point increase in star ratings ($OR = .904$) suggesting the star ratings of online reviews still have a negative relationship with the perceived helpfulness of the reviews.

As for Agoda.com (see Table 15), as an additional test of the scarcity effects explanation, hierarchical regression analysis was conducted (Hair et al., 2013), with star ratings as the independent variable, perceived helpfulness of the reviews as the dependent variable, total number of reviews by hotels and the overall star ratings by hotels as the control variables, controlling for the following scarcity effects markers: relatively high number of positive reviews and relatively high overall star ratings. Once again, since the analysis is intended to verify alternative explanations, only model 3 results are interpreted. As for model 3 for Agoda.com, overall model is significant ($R^2 = .004$, F (3, 60262) = 77.495, $p < .001$). The explanatory power of the regression models does not increase (model 1: $R^2 = .004$, Adjusted $R^2 = .004$, model 2: $R^2 = .004$, Adjusted $R^2 = .004$, model 3: $R^2 = .004$, Adjusted $R^2 = .004$). VIF values are all about 1.266, meaning that multicollinearity symptoms are not observed. The Durbin-Watson statistic is 1.876, which is considered suitable for the assumption of residual independence (Field, 2013). Specifically, even though the overall star ratings by hotels ($\beta = -.016$, $t = -3.519$, $p < .001$) influence perceived helpfulness of the reviews, star ratings still significantly influences perceived helpfulness of the reviews ($\beta = -.054$, $t = -3.519$, $p < .001$).

Overall, the results confirm that even if the scarcity effects have a minor effect, negativity bias, negative reviews still convey greater perceived helpfulness of the reviews than did positive reviews. However, as scarcity effects have a minor effect, further analysis is needed for the interaction effect as the scarcity effects failed to fully explain the results.

Table 13: Results of scarcity effects: hierarchical regression for Booking.com

Model	IV	Unstandardized Coefficients		Standardized Coefficients	t	P value
		B	Std. Error	Beta		
1	(Constant)	.090	.001		107.849***	.000
	Star Ratings	-.006	.000	-.039	-55.980***	.000
F (1,2034091) =3133.772 (p<.001), R^2=.002, Adjusted R^2=.002						
2	(Constant)	.085	.001		97.362***	.000
	Star Ratings	-.005	.000	-.037	-52.242***	.000
	Number of Reviews	.000	.000	.012	17.438***	.000
F (2,2034080) =1719.158 (p<.001), R^2=.002, Adjusted R^2=.002						
3	(Constant)	.104	.002		56.505***	.000
	Star Ratings	-.005	.000	-.034	-43.796***	.000
	Number of Reviews	.000	.000	.010	13.083***	.000
	Overall Rating	-.003	.000	-.009	-11.588***	.000
F (3,2034089) =1190.941 (p<.001), R^2=.002, Adjusted R^2=.002						

Table 14: Results of scarcity effects: Negative binomial regression analysis for Booking.com

	B	**S. E.**	**Wald**	**P value**	***OR***	**95% CI**
Star Ratings	-.100	.0018	3015.485***	.000	.904	(.901 ~.908)
Overall Ratings	-.056	.0040	197.333***	.000	.946	(.938~.953)
Number of Reviews	.000	.000	208.621***	.000	1.000	(1.000~1.000)
(Intercept)	-1.933	.0294	4322.849***	.000	. .145	(.137~.153)
Pearson $\chi 2$=1.613, Omnibus Test: Likelihood Ratio $\chi 2$=5380.394 (*p*<.001)						

Table 15: Results of scarcity effects: hierarchical regression for Agoda.com

Model	IV	Unstandardized Coefficients		Standardized Coefficients	t	P value
		B	Std. Error	Beta		
1	(Constant)	.066	.003		22.024***	.000
	Star Ratings	-.006	.000	-.060	-14.822***	.000
F (1,60264) =219.697 (p<.001), R^2=.004, Adjusted R^2=.004						
2	(Constant)	.067	.003		21.139***	.000
	Star Ratings	-.006	.000	-.060	-14.824***	.000
	Number of Reviews	-.000	.000	-.002	-.602	.547
F (2,60263) =110.028 (p<.001), R^2=.004, Adjusted R^2=.004						
3	(Constant)	.089	.007		12.637***	.000
	Star Ratings	-.005	.000	-.054	-11.881***	.000
	Number of Reviews	-.000	.000	-.005	-1.164	.244
	Overall Rating	-.003	.001	-.016	-3.519***	.000
F (3,60262) =77.495 (p<.001), R^2=.004, Adjusted R^2=.004						

4.6.2 Cultural diversity

Another possible alternative explanation is the cultural diversity in reviewers across both visited cities and reviewers' nationalities. Cultural diversity is the concept people use to identify themselves with others. The perception that a person is different from me can be based on values, cultural norms and demographic factors, for example, race, gender and age (Ferris, Frink & Galang, 1993). Alternatively, consumers who visited and posted reviews in certain city or consumers of different nationalities may systematically differ from those reviewed who visited in another city. To test these possibilities, this research conducted regression analyse based on both the eight different most popular travel destinations in the US and eighty-two different reviewers' nationalities on Booking.com. Agoda.com was excluded from the analysis due to an insufficient amount of data. As for different visited cities, two out of eight cities, New York and Las Vegas, made up approximately half of the total reviews (see Table 16). As for different reviewers' nationalities, the top ten reviewers' nationalities, USA, United Kingdom, Brazil, France, Germany, Italy, Canada, Argentina, Spain, and Australia made up approximately seventy per cent of the overall total reviews. Of more than two-hundred reviewers' nationalities, the analysis was based on eighty-two reviewers' nationalities (N = 1,965,339, Percent = 96.52%, see Table 19), except for reviewers' nationalities where the sample size was insufficient due to lack of sample size.

The results of simple regression analysis based on different visited cities confirm that star ratings significantly influences perceived helpfulness of the reviews for all eight most popular travel destinations in the US (Boston: β = -.052, t = -16.26, p < .001, Chicago: β = -.027, t = -9.80, p < .001, Las Vegas: β = -.022, t = -13.38, p < .001, Los Angeles: β = -.045, t = -21.02, p < .001, Miami: β = -.030, t = -11.17, p < .001, New York: β = -.049, t = -38.11, p < .001, Orlando: β = -.028, t = -13.00, p < .001, San Francisco: β = -.033, t = -16.51, p < .001). Moreover, the results of negative binomial regression analysis (Gardner, Mulvey & Shaw, 1995) based on different visited cities confirm that star ratings significantly influence perceived helpfulness of the reviews for all eight most popular travel destinations in the US (Boston: B = -.194, SE = .0109, Wald = 315.571, p < .001, OR = .824, Chicago: B = -.120, SE = .0104, Wald = 132.403, p < .001, OR = .887, Las Vegas: B = -.072, SE = .0045, Wald = 257.648, p < .001, OR = .931, Los Angeles: B = -.131, SE = .005, Wald = 678.248, p < .001, OR = .877, Miami: B = -.088, SE = .0068, Wald = 166.073, p < .001, OR = .916, New York: B = -.131, SE = .0026, Wald = 2562.489, p < .001, OR = .877, Orlando: B = -.075, SE = .0044, Wald = 283.844, p < .001, OR = .928, San Francisco: B = -.101, SE = .0052, Wald = 375.071, p < .001, OR = .904). Overall, when the reviewers were divided into visited cities and negativity bias, negative reviews still convey greater perceived helpfulness of the reviews than do positive reviews. Table 16, 17 and 18 show a summary of results for Booking.com and Figure 11 shows the geo-chart of cultural diversity based on different visited cities.

Table 16: Frequency of eight most popular travel destinations in the US for Booking.com

City	Frequency	Valid Percent	Cumulative Percent
Boston	95884	4.7	4.7
Chicago	130103	6.4	11.1
Las Vegas	375320	18.4	29.5
Los Angeles	218387	10.7	40.3
Miami	143137	7.0	47.3
New York	611093	30.0	77.3
Orlando	215076	10.6	87.9
San Francisco	247260	12.1	100.0
Total	2036260	100.0	

Table 17: Results of cultural diversity based on different visited cities: Regression for Booking.com

City	IV	Unstandardized Coefficients		Standardized Coefficients	t	P value
		B	Std. Error	Beta		
Boston	(Constant)	.066	.003		24.184***	.000
	Star Ratings	-.005	.000	-.052	-16.259***	.000
F (1,95882) =264.352 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.343						
Chicago	(Constant)	.043	.002		18.102***	.000
	Star Ratings	-.003	.000	-.027	-9.803***	.000
F (1,130101) =96.092 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.460						
Las Vegas	(Constant)	.056	.002		34.419***	.000
	Star Ratings	-.003	.000	-.022	-13.380***	.000
F (1,375319) =179.013 (p<.001), R^2=.0005, Adjusted R^2=.0005, Durbin-Watson=.379						
Los Angeles	(Constant)	.096	.003		38.202***	.000
	Star Ratings	-.007	.000	-.045	-21.015***	.000
F (1,218385) =441.637 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.496						
Miami	(Constant)	.060	.002		25.509***	.000
	Star Ratings	-.003	.000	-.030	-11.173***	.000

F (1,143135) =124.841 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.358						
New York	(Constant)	.131	.002		69.465***	.000
	Star Ratings	-.009	.000	-.049	-38.111***	.000
F (1,611091) =1452.428 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.530						
Orlando	(Constant)	.086	.003		32.284***	.000
	Star Ratings	-.004	.000	-.028	-13.000***	.000
F (1,215074) =168.992 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.414						
San Francisco	(Constant)	.071	.002		35.250***	.000
	Star Ratings	-.004	.000	-.033	-16.510***	.000
F (1,247258) =272.575 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.382						

Table 18: Results of cultural diversity based on different visited cities: Negative binomial regression analysis for Booking.com

City	**IV**	***B***	**S. E.**	**Wald**	**P value**	***OR***	**95% CI**
Boston	(Intercept)	-2.246	.0862	678.245***	.000	.106	(.089~.125)
	Star Ratings	-.194	.0109	315.571***	.000	.824	(.807~.842)
Pearson $\chi2$=1.199, Omnibus Test: Likelihood Ratio $\chi2$=288.063 (p<.001)							
Chicago	(Intercept)	-2.917	.0855	1162.793***	.000	.054	(.046~.064)
	Star Ratings	-.120	.0104	132.403***	.000	.887	(.869~.905)
Pearson $\chi2$=1.374, Omnibus Test: Likelihood Ratio $\chi2$=123.713 (p<.001)							
Las Vegas	(Intercept)	-2.800	.0355	6207.147***	.000	.061	(.057~.065)
	Star Ratings	-.072	.0045	257.648***	.000	.931	(.923~.939)
Pearson $\chi2$=1.463 Omnibus Test: Likelihood Ratio $\chi2$=248.945 (p<.001)							
Los Angeles	(Intercept)	-2.114	.0385	3020.694***	.000	.121	(.112~.130)
	Star Ratings	-.131	.0050	678.248***	.000	.877	(.869~.886)
Pearson $\chi2$=1.545, Omnibus Test: Likelihood Ratio $\chi2$=645.157 (p<.001)							
Miami	(Intercept)	-2.688	.0529	2582.685***	.000	.068	(.061~.075)
	Star Ratings	-.088	.0068	166.073***	.000	.916	(.903~.928)
Pearson $\chi2$=1.359, Omnibus Test: Likelihood Ratio $\chi2$=160.147 (p<.001)							
	(Intercept)	-1.798	.0198	8219.114***	.000	.166	(.159~.172)

New York	Star Ratings	-.131	.0026	2562.489***	.000	.877	(.873~.882)
Pearson χ2=1.794, Omnibus Test: Likelihood Ratio χ2=2447.520 (p<.001)							
Orlando	(Intercept)	-2.366	.0353	4496.961***	.000	.094	(.088~.101)
	Star Ratings	-.075	.0044	283.844***	.000	.928	(.920~.936)
Pearson χ2=1.716, Omnibus Test: Likelihood Ratio χ2=273.363 (p<.001)							
San Francisco	(Intercept)	-2.499	.0395	4001.655***	.000	.082	(.076~.089)
	Star Ratings	-.101	.0052	375.071***	.000	.904	(.895~.913)
Pearson χ2=1.407, Omnibus Test: Likelihood Ratio χ2=361.663 (p<.001)							

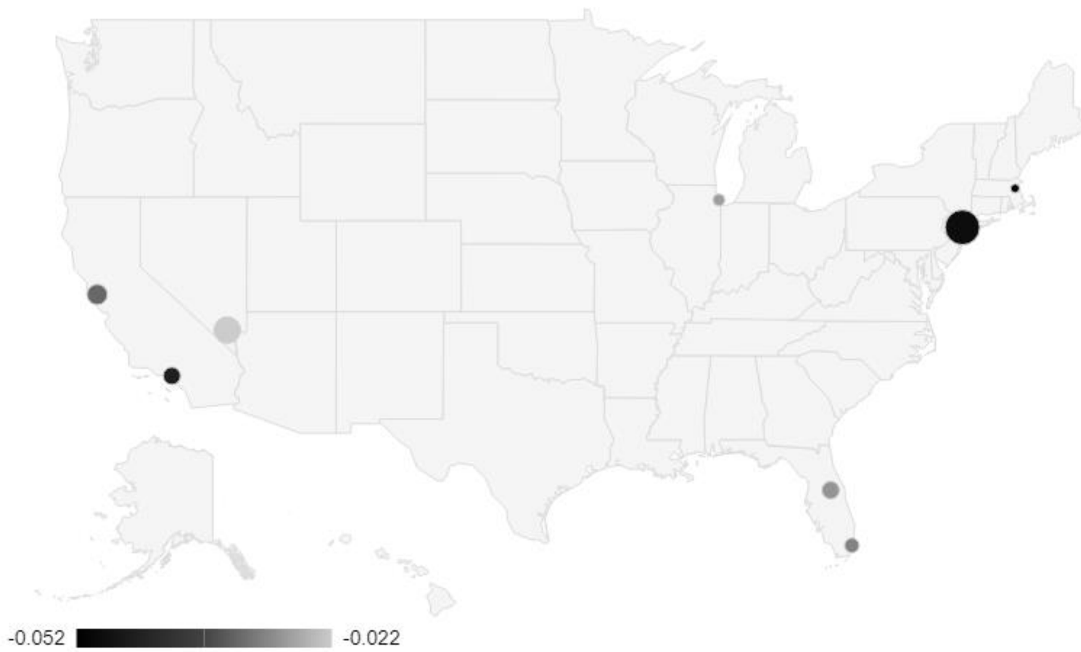

Figure 11: Geo chart of cultural diversity based on different visited cities (circle size = frequency, saturation = Beta)

The results of simple regression analysis based on different reviewers' nationalities similarly confirm that star ratings significantly influence perceived helpfulness of the reviews for all eighty-two reviewers' of different nationality (all of the β <= -.016, all of the p < .05) suggesting that even if the reviewers are divided by reviewers' nationalities, negativity bias, negative reviews still convey greater perceived helpfulness than do positive. Detailed results are summarised in Appendix A. Moreover, the results of a negative binomial regression analysis (Gardner, Mulvey & Shaw, 1995) based on reviewers' nationalities again confirm that star ratings significantly influences perceived helpfulness of the reviews for all eighty-two different reviewers' nationalities (all of the $B <= -.042$, all of the $p < .113$), suggesting that even if the reviewers were divided into reviewers' nationalities, negativity bias, negative reviews still convey greater perceived helpfulness than positive. Detailed results are summarised in Appendix B. Further analysis is not needed for the interaction effect as cultural diversity failed to explain the results. Table 19 shows a summary of results for Booking.com and Figure 12 shows the geo chart of cultural diversity based on different reviewers' nationalities.

Table 19: Frequency of eighty different reviewers' nationalities

Nationality	**Frequency**	**Percent**	**Nationality**	**Frequency**	**Percent**
Algeria	543	.03	Kazakhstan	1724	.08
Argentina	59832	2.94	Kenya	510	.03
Armenia	279	.01	Kuwait	3493	.17
Australia	48057	2.36	Latvia	1203	.06
Austria	13951	.69	Macao	260	.01
Bangladesh	440	.02	Macedonia	267	.01
Belarus	746	.04	Malaysia	1914	.09
Belgium	17665	.87	Martinique	1075	.05
Bolivia	1697	.08	Mexico	27966	1.37
Brazil	89468	4.39	Monaco	284	.01
Burkina Faso	59	.00	Morocco	1115	.05
Cameroon	66	.00	Mozambique	133	.01
Canada	61326	3.01	Netherlands	32258	1.58
Chile	20053	.98	New Zealand	9983	.49
China	39784	1.95	Niger	1811	.09
Colombia	21076	1.04	Norway	7411	.36
Costa Rica	5771	.28	Panama	4271	.21
Cyprus	959	.05	Peru	6616	.32
Czech Republic	8432	.41	Philippines	3772	.19
Denmark	10632	.52	Poland	11535	.57
Dominica	30	.00	Portugal	7695	.38
Dominican Republic	3140	.15	Puerto Rico	6915	.34
Ecuador	7387	.36	Qatar	2338	.11
Egypt	2207	.11	Romania	3341	.16

El Salvador	1095	.05	Russia	23811	1.17
France	87394	4.29	Saudi Arabia	11221	.55
Gabon	71	.00	Slovakia	2745	.13
Germany	82134	4.03	South Africa	6199	.30
Greece	3757	.18	South Korea	15521	.76
Grenada	111	.01	Spain	56158	2.76
Guatemala	2829	.14	Sri Lanka	562	.03
Hong Kong	4976	.24	Sweden	18968	.93
Hungary	6248	.31	Switzerland	35426	1.74
Iceland	3381	.17	Taiwan	9895	.49
India	7809	.38	Turkey	12083	.59
Iraq	255	.01	Turkmenistan	40	.00
Ireland	19975	.98	Ukraine	4233	.21
Israel	27968	1.37	United Arab Emirates	8381	.41
Italy	71250	3.50	United Kingdom	115339	5.66
Japan	24579	1.21	USA	737258	36.21
Jordan	727	.04	Venezuela	11450	.56
			Total	1965339	96.52

Figure 12: Geo chart of cultural diversity based on different reviewers' nationalities (saturation = Beta)

4.7 Analysis of Hypothesis 2

The purpose of the second analysis is to test the hypothesis that the type of information moderates the impact of review valence on the perceived helpfulness of the reviews. Specifically, when the type of information is one-sided, review valence has a strong negative impact on the perceived helpfulness of the reviews. However, when the type of information is two-sided, the negative impact of review valence on the perceived helpfulness of the reviews is relatively weak. With the setting to test out the prediction, I analysed online hotel reviews generated by consumers on Booking.com, excluding zero-length reviews.

4.7.1 Data Set

To investigate the two-way interaction effect, this research divided Booking.com reviews (N = 2,036,260) into one-sided information reviews (N = 338,490) and two-sided (N = 872,685) minus zero-length reviews (N = 825,085). Agoda.com data was excluded from the second analysis as it does not provide positive and negative aspects. As for Booking.com, there are separate columns that write positive and negative aspects when consumers write hotel reviews. When consumers have written reviews on one of the positive or negative aspects, they were classified as one-sided information, whereas consumers who wrote reviews on both positive and negative aspects are classified as two-sided information.

4.7.2 The moderation analysis results

To formally test the too good to be true effect, PROCESS macro (Model 1, 5000 bootstrap samples) was utilised (Hayes, 2018), wherein the moderation analysis was conducted, with star ratings as independent variable, the type of information as moderating variable, and perceived helpfulness of the reviews as the dependent variable.

The analysis yields a significant two-way interaction effect between star ratings and the type of information on perceived helpfulness of the reviews (β = .0022, se = .0002, t = 12.0379, p <.001, 95% CI = (.0018, .0025). When probing the pattern for the interaction effect, further analysis indicates that when the information is two-sided, the negative impact of review valence on the perceived helpfulness of the reviews is relatively weak (β = -

.0045, se = .0002, t = -22.4996, p < .001, 95% CI = (-.0049, -.0041)). In contrast, when the information is one-sided, review valence has a strong negative impact on the perceived helpfulness of the reviews (β = -.0088, se = .0003, t = -29.2420, p < .001, 95% CI = (-.0094, -.0082), suggesting that the type of information moderates the impact of review valence on the perceived helpfulness of the reviews. Figure 13 and Table 20 show a summary of the results for Booking.com.

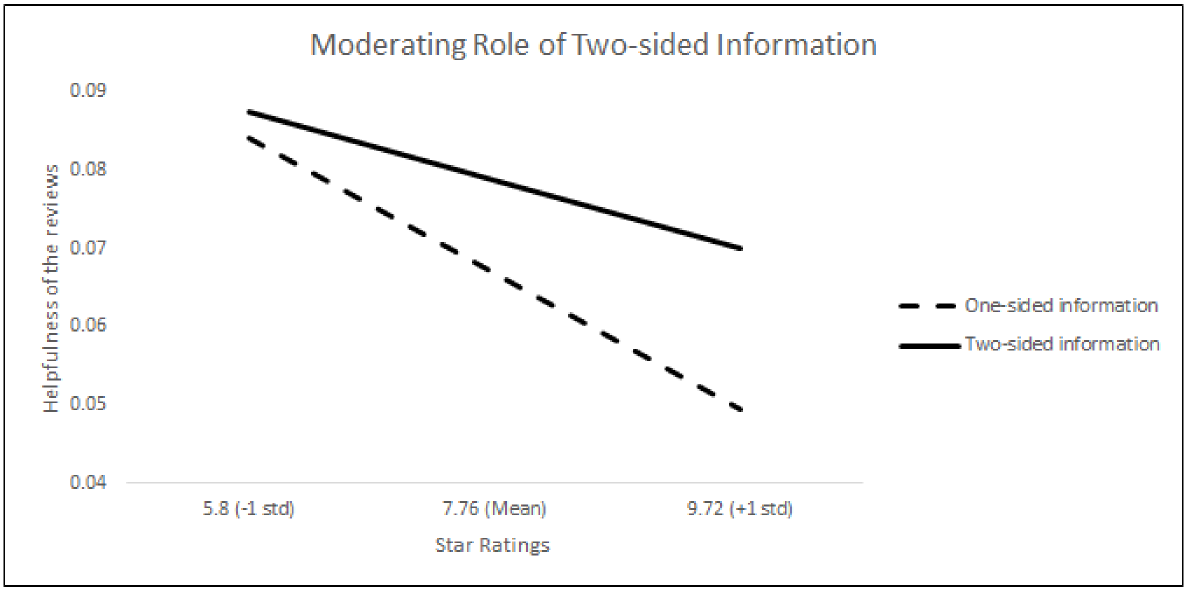

Figure 13: The moderating role of consumer scepticism, “too good to be true.”

Table 20: Results of second analysis: Hayes model 1 for Booking.com

Model	**coefficient**	**se**	**t**	**P**	**LLCI**	**ULCI**
constant	.1244	.0015	83.5340***	.0000	.1215	.1273
Star Ratings	-.0067	.0002	-36.7974***	.0000	-.0070	-.0063
Type of information	-.0109	.0015	-7.3469***	.0000	-.0139	-.0080
Interaction	.0022	.0002	12.0379***	.0000	.0018	.0025
Moderator	**Effect**	**se**	**t**	**P**	**LLCI**	**ULCI**
One-sided	-.0088	.0003	-29.2420***	.0000	-.0094	-.0082
Two-sided	-.0045	.0002	-22.4996***	.0000	-.0049	-.0041

4.8 Alternative Explanation for Hypothesis 2

4.8.1 Scarcity Effects

To further test the scarcity effects in a two-way interaction effect, PROCESS macro (Model 1, 5000 bootstrap samples) was utilised (Hayes, 2018), wherein the moderation analysis was conducted, with star ratings as independent variable, the type of information as moderating variable, perceived helpfulness of the reviews as the dependent variable, and total number of reviews by hotels and the overall star ratings by hotels as covariates.

The analysis yields a significant two-way interaction effect between star ratings and the type of information on perceived helpfulness of the reviews (β = .0023, se = .0002, t = 12.6005, p <.001, 95% CI = (.0019, .0026)). When probing the pattern for the interaction effect, further analysis indicates that when the information is two-sided the negative impact of review valence on the perceived helpfulness of the reviews is relatively weak (β = -.0032, se = .0002, t = -14.8195, p <..001, 95% CI = (-.0036, -.0028)). In contrast, when the information is one-sided, review valence has a strong negative impact on the perceived helpfulness of the reviews (β = -.0078, se = .0003, t = -24.9437, p < .001, 95% CI = (-.0084, -.0071), showing the pattern of results remains the same and the scarcity effect fails to explain the results. Figure 14 and Table 21 show a summary of the results for Booking.com.

Figure 14: The moderating role of consumer scepticism, “Too good to be true!”

Table 21: Results of scarcity effects: Hayes model 1 for Booking.com

Model	coefficient	se	t	P	LLCI	ULCI
constant	.1448	.0031	46.4021***	.0000	.1387	.1509
Star Ratings	-.0055	.0002	-27.8182***	.0000	-.0059	-.0051
Type of information	-.0119	.0015	-7.9943***	.0000	-.0148	-.0090
Interaction	.0023	.0002	12.6005***	.0000	.0019	.0026
Overall Ratings	-.0041	.0004	-10.0728***	.0000	-.0049	-.0033
Number of reviews	.0000	.0000	11.7241***	.0000	.0000	.0000
Moderator	**Effect**	**se**	**t**	**P**	**LLCI**	**ULCI**
One-sided	-.0078	.0003	-24.9437***	.0000	-.0084	-.0071
Two-sided	-.0032	.0002	-14.8195***	.0000	-.0036	-.0028

4.9 Analysis of Hypothesis 3

The purpose of the third analysis is to test the hypothesis that systematic information processing moderates the impact of review valence on the perceived helpfulness of the reviews. Specifically, when the length of comments is shorter, review valence has a strong negative impact on the perceived helpfulness of the reviews. However, when the length of comments is longer, review valence has a positive impact on the perceived helpfulness of the reviews.

4.9.1 Data Set

To investigate the two-way interaction effect, I analysed two full data sets: one composed of 2,036,260 online reviews ($M\_{length\_of\_comments} = 122.67$, Min = 0, Max = 5,869, SD = 204.94), and a second composed of 60,266 online reviews ($M\_{length\_of\_comments} = 231.234$, Min = 0, Max = 2,139, SD = 231.23). The length of the comment was measured based on the number of characters.

4.9.2 The moderation analysis results

To formally test the systematic information processing, PROCESS macro (Model 1, 5000 bootstrap samples) was utilised (Hayes, 2018), where star ratings is the independent variable, length of comments is the moderating variable and perceived helpfulness of the reviews is the dependent variable.

The analysis shows that Booking.com has a significant two-way interaction effect between star ratings and length of comments on perceived helpfulness of the reviews (β = .00002, se = .00001, t = 40.0252, p <.001, 95% CI = (.00002, .00002)). When probing the pattern for the interaction effect, further analysis indicates that when length of comment is shorter with 0, the negative impact of review valence on the perceived helpfulness of the reviews is relatively strong (β = -.004, se = .0001, t = -34.2079, p <.001, 95% CI = (-.0043, -.0038)). In contrast, when length of comments is longer with 328, review valence has a positive impact on the perceived helpfulness of the reviews (β = .0017, se = .0001, t = 12.6859, p < .001, 95% CI = (.0014, .0019), suggesting that systematic (vs. heuristic) information processing moderates the impact of review valence on the perceived helpfulness of the reviews. Figure 15 and Table 22 show a summary of the results for Booking.com.

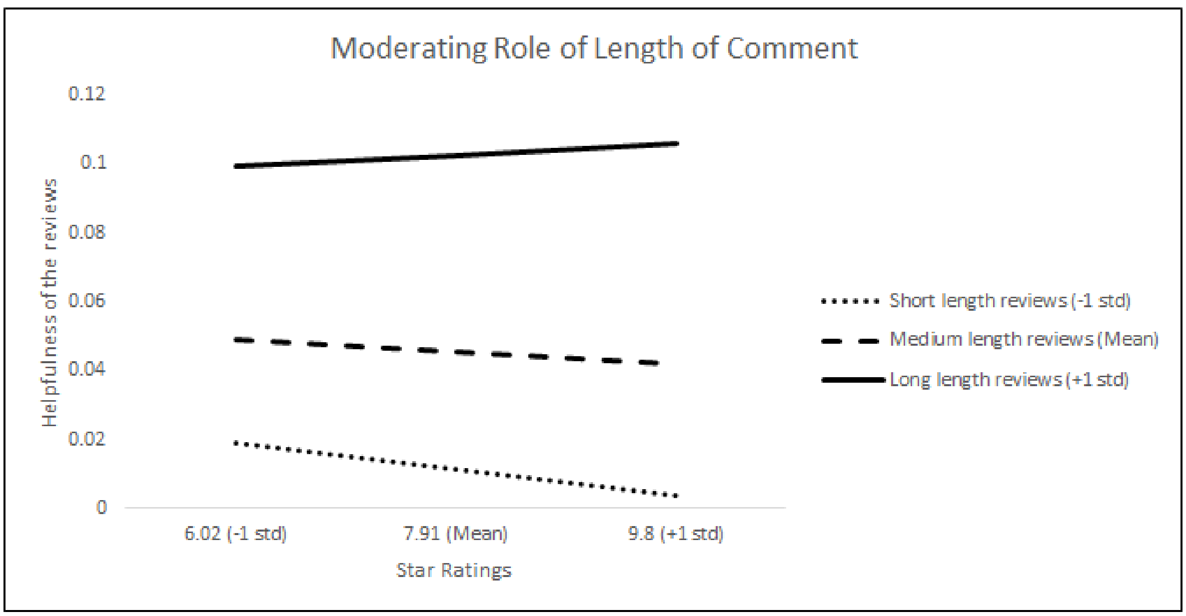

Figure 15: The moderating role of systematic (vs. heuristic) information processing

Table 22: Results of third analysis: Hayes model 1 for Booking.com

Model	**coefficient**	**se**	**t**	**P**	**LLCI**	**ULCI**
constant	.04332311	.00096861	44.72731547***	.0000	.04142468	.04522155
Star Ratings	-.00404537	.00011826	-34.20790822***	.0000	-.00427715	-.00381359
Length of comments	.00014115	.00000327	43.19494563***	.0000	.00013475	.00014756
Interaction	.00001740	.00000043	40.02519711***	.0000	.00001654	.00001825
Moderator	**Effect**	**se**	**t**	**P**	**LLCI**	**ULCI**
0	-.00404537	.00011826	-34.20790822***	.0000	-.00427715	-.00381359
123	-.00191143	.00010180	-18.77623008***	.0000	-.00211095	-.00171190
328	.00165383	.00013037	12.68585059***	.0000	.00139831	.00190934

This shows that Agoda.com has a significant two-way interaction effect between star ratings and length of comments on perceived helpfulness of the reviews (β = .0001, se = .00001, t = -8.486, p < .001, 95% CI = (-.00002, -.00001)). However, when seeking the interaction effect, further analysis indicated that when length of comments is shorter than 38, the negative impact of review valence on the perceived helpfulness of the reviews is relatively weak (β = -.0026, se = .0005, t = -5.3552, p < .001, 95% CI = (-.0035, -.0016)). In contrast, when the length of comments is longer than 385 characters, the negative impact of review valence on the perceived helpfulness of the reviews is relatively strong (β = -.007, se = .0005, t = -15.4341, p < .001, 95% CI = (-.0079, -.0061). This suggests that systematic information processing does, indeed, moderate the impact of review valence on the perceived helpfulness of the reviews.

Contrary to expectations, this finding was unexpected and suggests that, in terms of Agoda.com, with the presence of consumer scepticism, too good to be true systematic information processing could not weaken the negativity bias. One possible explanation is that long reviews filled with only positive content may be considered more like fake reviews than short reviews filled with only positive content, based on rational thinking. In contrast, long reviews filled with only negative content may be considered less like fake reviews than short reviews filled with only negative content. Thus, when online reviews are one-sided, systematic information processing strengthens consumer scepticism.

Overall, the analysis of both sites supports a significant two-way interaction effect between both star ratings and length of comments on perceived helpfulness of the reviews. In the post-hoc test, however, the opposite result is obtained. The systematic information processing hypothesis is supported only in Booking.com, which provides two-sided information. Thus, a more detailed analysis of the type of information and systematic information processing is reported in the fourth analysis. Figure 16 and Table 23 show a summary of the results for Agoda.com.

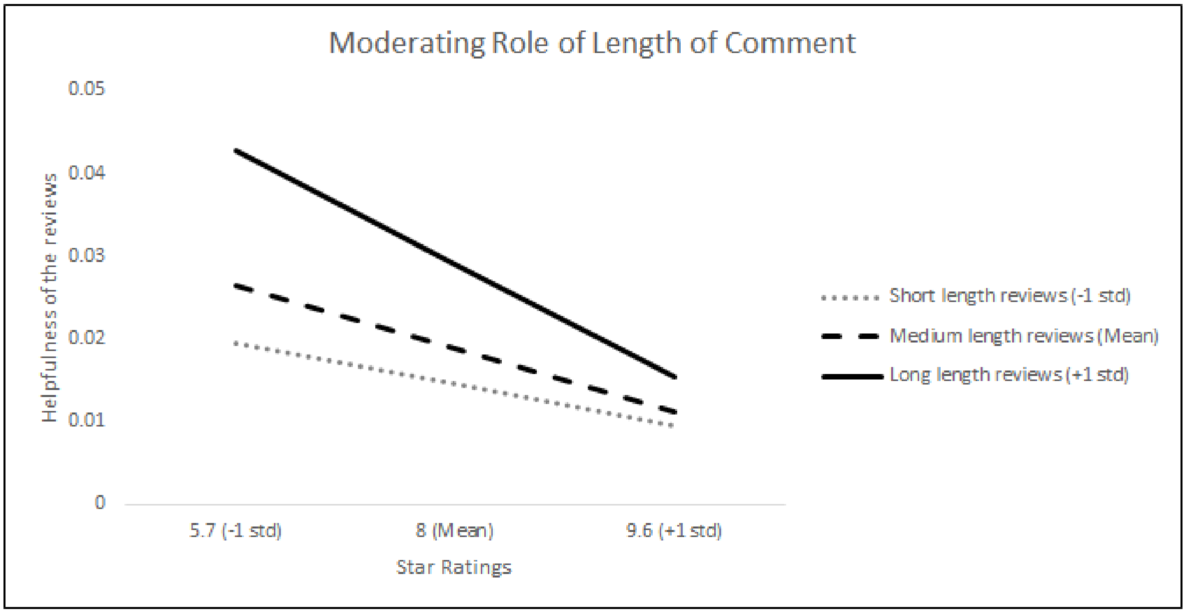

Figure 16: The moderating role of systematic (vs. heuristic) information processing

Table 23: Results of third analysis: Hayes model 1 for Agoda.com

Model	**coefficient**	**se**	**t**	**P**	**LLCI**	**ULCI**
constant	.02887687	.00404483	7.13920262 ***	.000	.02094899	.03680475
Star Ratings	-.00208825	.00051671	-4.04142647 ***	.000	-.00310101	-.00107550
Length of comments	.00013942	.00001120	12.45185598 ***	.000	.00011747	.00016136
Interaction	-.00001274	.00000151	-8.41855528 ***	.000	-.00001570	-.00000977
Moderator	**Effect**	**se**	**t**	**P**	**LLCI**	**ULCI**
38	-.00257223	.00048033	-5.35516263 ***	.000	-.00351367	-.00163079
141	-.00388405	.00040697	-9.54389480 ***	.000	-.00468171	-.00308640
385	-.00699168	.00045300	-15.43410283 ***	.000	-.00787956	-.00610379

4.10 Alternative Explanation for Hypothesis 3

4.10.1 Scarcity Effects

To further test the scarcity effects in two-way interaction effect, PROCESS macro (Model 1, 5000 bootstrap samples) was again utilised (Hayes, 2018). Moderation analysis was conducted, with star ratings as an independent variable, length of comments as the moderating variable, perceived helpfulness of the reviews as the dependent variable, and the total number of reviews by hotels and the overall star ratings by hotels as covariates.

The analysis for Booking.com yields a significant two-way interaction effect between star ratings and length of comments on perceived helpfulness of the reviews (β = .00002, se = .00001, t = 40.5943, p <.001, 95% CI = (.00002, .00002)). Further analysis indicates that when length of comments is shorter, with 0, the negative impact of review valence on the perceived helpfulness of the reviews is relatively strong (β = -.0029, se = .0001, t = -22.9986, p < .001, 95% CI = (-.0032, -.0027)). In contrast, when length of comments is longer than 328, review valence has a positive impact on the perceived helpfulness of the reviews (β = .0029, se = .0001, t = 20.6957, p < .001, 95% CI = (.0026, .0031). This indicates that the pattern of results remains the same and that scarcity effects failed to explain the results. Thus, further analysis is not needed for the three-way interaction effect as both the two-way interaction of scarcity effects failed to explain the results. Figure 17 and Table 24 show a summary of the results for Booking.com.

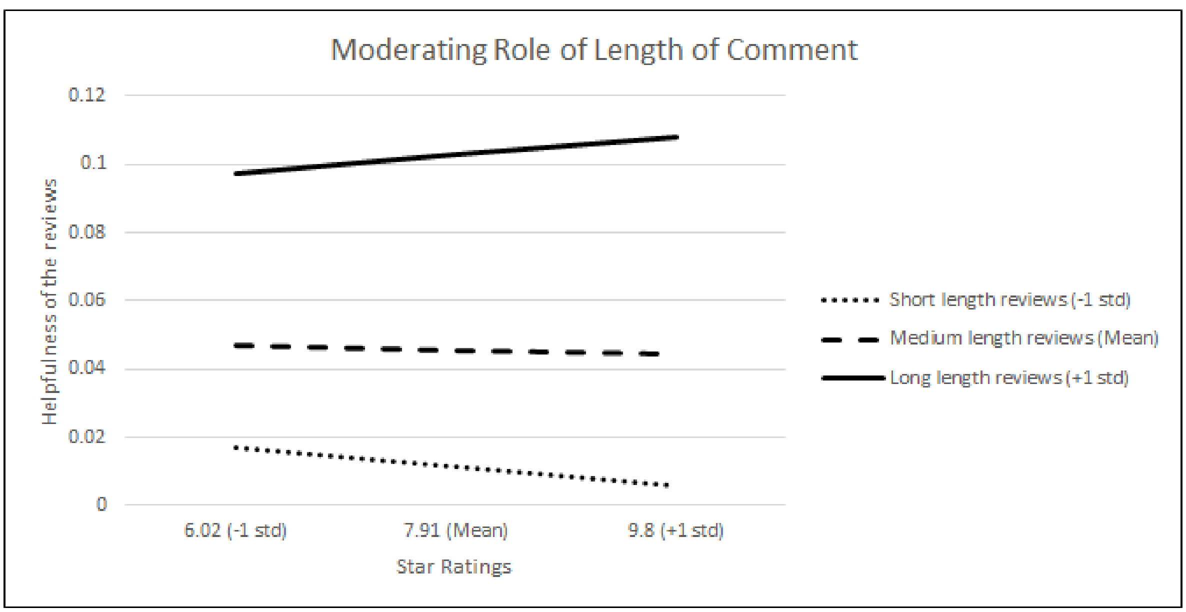

Figure 17: The moderating role of systematic (vs. heuristic) information processing

Table 24: Results of scarcity effects: Hayes model 1 for Booking.com

Model	coefficient	se	t	P	LLCI	ULCI
constant	.0692	.0019	36.9290***	.0000	.0656	.0729
Star Ratings	-.0029	.0001	-22.9986***	.0000	-.0032	-.0027
Length of comments	.0001	.0000	42.7486***	.0000	.0001	.0001
Interaction	.0000	.0000	40.5943***	.0000	.0000	.0000
Overall Ratings	-.0046	.0002	-19.4075***	.0000	-.0051	-.0042
Number of reviews	.0000	.0000	12.2922***	.0000	.0000	.0000
Moderator	**Effect**	**se**	**t**	**P**	**LLCI**	**ULCI**
0	-.0029	.0001	-22.9986***	.0000	-.0032	-.0027
123	-.0007	.0001	-6.6968***	.0000	-.0010	-.0005
328	.0029	.0001	20.6957***	.0000	.0026	.0031

4.11 Analysis for Hypothesis 4

The purpose of addressing Hypothesis 4 is to test how both consumer scepticism and systematic information processing moderate the impact of review valence on the perceived helpfulness of the reviews. Specifically, when information is two-sided review valence has a positive impact on the perceived helpfulness of the reviews only in a longer comment scenario. Alternatively, when information is one-sided, then review valence has a negative impact on the perceived helpfulness of the reviews for both shorter and longer comments.

4.11.1 Data Set

To investigate the three-way interaction effect, I divided Booking.com reviews (N = 2,036,260, $M_{length\_of\_comments}$ = 122.67, Min = 0, Max = 5,869, SD = 204.94) into one-sided information reviews (N = 338,490, $M_{length\_of\_comments}$ = 117.37, Min = 1, Max = 2,003, SD = 117.37), and two-sided (N = 872,685, $M_{length\_of\_comments}$ = 240.69, Min = 2, Max = 5,869, SD = 240.69); when zero length reviews are removed N = 825,085. Agoda.com data was excluded from the analysis as it does not provide positive and negative aspects.

4.11.2 The moderation analysis results

To formally test both the too good to be true effect and the systematic information processing, PROCESS macro (Model 3, 5000 bootstrap samples) was again utilised (Hayes, 2018). Star ratings is the independent variable, both the type of information and length of comments are moderating variables, and perceived helpfulness of the reviews is again the dependent variable. The analysis yields a significant three-way interaction effect between star ratings and the type of information and length of comments on perceived helpfulness of the reviews (β = .000004, se = .000001, t = 4.8377, p <.001, 95% CI = (.000003, .000006)).

Specifically, when information is one-sided and comments are shorter with 1, the negative impact of review valence on the perceived helpfulness of the reviews is relatively strong (β = -.0072, se = .0004, t = -18.7927, p <.001, 95% CI = (-.0079, -.0064)). In contrast, when information is one-sided and length of comments is longer with 437, review valence has a marginally negative impact on the perceived helpfulness of the reviews (β = -.0009, se = .0006, t = -1.677, p = .0935, 95% CI = (-.002, -.0002). This suggests that when the information is one-sided, review valence has a negative impact on the perceived helpfulness of the reviews in both shorter and longer comments.

Furthermore, results of Agoda.com in the analysis of Hypothesis 3 also supports Hypothesis 4. When information is two-sided and length of comments is shorter with 1, the negative impact of review valence on the perceived helpfulness of the reviews is relatively strong (β = -.007, se = .0003, t = -26.1028, p <.001, 95% CI = (-.0075, -.0065)). In contrast, when the information is two-sided and length of comments is longer with 437, review valence has a positive impact on the perceived helpfulness of the reviews (β = .0029, se = .0002, t = 12.3513, p < .001, 95% CI = (.0025, .0034), again suggesting that when information is two-sided, review valence has a positive impact on the perceived helpfulness of the reviews only for longer comments. Table 25 and Figure 18 show a summary of the results for Booking.com.

However, as shown in Table 25, one concern with the findings is that the effect sizes appear very small (absolute value (all 'Effect') < .01). Generally, the larger the sample size, the smaller the expected effect size (Burmeister & Aitken, 2012; Schnack & Kahn, 2016). Therefore, in the case of big data, the effect size might be very small. Nevertheless, Schnack and Kahn (2016) argued that the accuracy of machine learning techniques (e.g., deep learning) can be used as a proxy of statistical effect size. Thus, machine learning techniques are further analysed to address the small statistical effect sizes observed in findings.

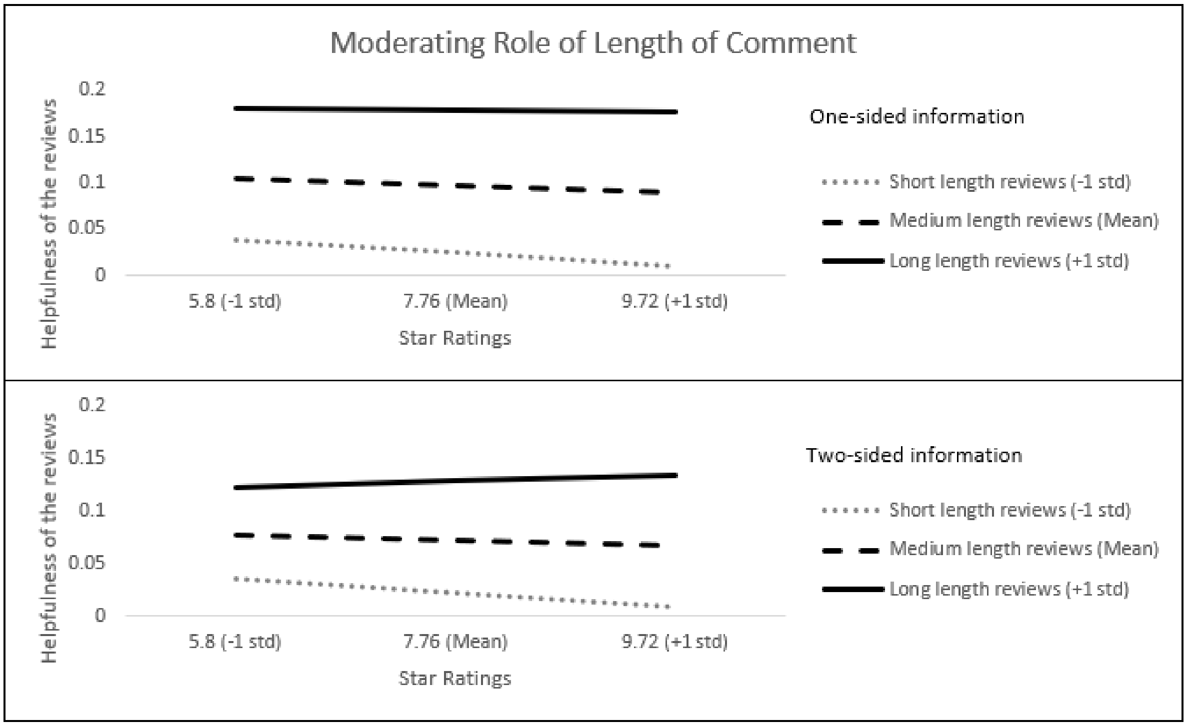

Figure 18: The moderating role of both consumer scepticism, "Too good to be true!" and systematic (vs. heuristic) information processing

Table 25: Results of fourth analysis: Hayes model 3 for Booking.com

Model	**coefficient**	**se**	**t**	**P**	**LLCI**	**ULCI**
constant	.7826046	.00191320	40.90544419 ***	.0000	.07451065	.08201028
Star Ratings	-.00710977	.00023380	-30.40897847 ***	.0000	-.00756802	-.00665152
Length of comments	.00015521	.00000661	23.49020240 ***	.0000	.00014226	.00016816
Interaction 1	.00001854	.00000088	21.08519841 ***	.0000	.00001681	.00002026
Type of information	-.00204810	.00191320	-1.07050911	.2844	-.00579792	.00170171
Interaction 2	.00006997	.00023380	.29926000	.7647	-.00038828	.00052822

Interaction 3	-.00008803	.00000661	-13.32324348 ***	.0000	-.00010098	-.00007508
Interaction 4	.00000425	.00000088	4.83768892 ***	.0000	.00000253	.00000598
Moderator	**Effect**	**se**	**T**	**P**	**LLCI**	**ULCI**
1, one-sided	-.00716545	.00038129	-18.79267954 ***	.0000	-.00791277	-.00641814
1, two-sided	-.00701701	.00026882	-26.10276944 ***	.0000	-.00754389	-.00649013
206, one-sided	-.00423392	.00031701	-13.35596897 ***	.0000	-.00485524	-.00361260
206, two-sided	-.00233975	.00020094	-11.64420249 ***	.0000	-.00273358	-.00194592
437, one-sided	-.00093367	.00055674	-1.67703777 *	.0935	-.00202485	.00015752
437, two-sided	.00292580	.00023688	12.35131538 ***	.0000	.00246152	.00339008

Product terms key:
Interaction 1: Star Ratings x Length of comments
Interaction 2: Star Ratings x Type of information
Interaction 3: Length of comments x Type of information
Interaction 4: Star Ratings x Length of comments x Type of information

Chapter 5: Artificial intelligence: Machine learning for Big data

5.1 Introduction

Artificial intelligence analysis was conducted to achieve multiple purposes. Even though the age of big data is swiftly coming, traditional statistical data analysis is limited in analysing these large amounts of data. In order to analyse big data efficiently, the application of novel technology machine learning techniques is essential (Lei et al., 2016). To this end, this study selected deep learning as the analytical method among numerous machine learning techniques. Because deep learning is one of the exceptional machine learning technologies developed to this time. It has been particularly successful in several areas, such as voice recognition, image analysis, and natural language processing (Zhang et al., 2018). The multiple purposes are as follows; firstly, in order to make the results of analysis more robust by addressing a small statistical effect size, the effect size of the IVs is not only measured with statistical methods but also a novel machine learning method, namely artificial neural network (Schnack & Kahn, 2016). Secondly, in order to measure and apply advanced machine learning techniques, the effect size of the IVs is measured by comparing both traditional machine learning technique, namely bi-logistic regression, using single-layer perceptron and deep-learning using multi-layer perceptron.

5.2 Neural networks

Machine learning techniques are a subfield of artificial intelligence technology and are given the name of techniques that modifies behaviour through learning to maximize the likelihood of reaching a specific goal. The three main algorithms of machine learning are supervised learning, unsupervised learning, and reinforcement learning. First, the supervised learning algorithm learns by modifying the target parameters by comparing the output with the correct answer. Second, the unsupervised learning algorithm learns everything from the data without the correct answer and finds and automatically classifies the features that are common to the groups. Finally, reinforcement learning algorithms proceed with learning with reward and punishment information about whether to continue or stop learning (Bonaccorso, 2017). Even though the big data revolution is delivering new value through insight discovery and better decision making, realizing this potential requires novel machine learning techniques that can extract value from vast amounts of data. Because traditional machine learning and statistical techniques were developed in different eras before big data appeared, the unique characteristics of big data became an obstacle to analysis (L'heureux et al., 2017). Thus, this study analyses big data using artificial neural networks, which is just one of the novel techniques of the supervised learning algorithm.

The original name for a neural network is an artificial neural network. It imitates the learning method of the human brain and it repeatedly learns the given data through the machine learning algorithm. This is an analysis algorithm that is used to find patterns and characteristics in data and generalize them to predict future results (LeCun, Bengio & Hinton, 2015). Since the calculation method of the existing computers was sequential operation processing, according to a flowchart, it was not able to carry out the self-learning method that artificial intelligence demanded. However, as the neural network algorithm has been developed, it has become possible to design the algorithm to find its own solution through learning. In other words, while conventional analysis methods have been processed sequentially by serial processing according to a given flowchart, a neural network is a self-modifying and repetitive learning method finding a better solution by parallel processing much like a human brain operation (Chollet, 2018).

The way in which the human brain processes information in parallel is roughly understood. The unit of the cell that constitutes the human brain is called a neuron. It consists of four parts: the soma, synapse, and axon including dendrite and nucleus. Approximately one hundred billion neurons are present in the human cerebral cortex, and each neuron is linked to another by a synapse. Neurons activate or deactivate other neurons connected by electrical stimulation of the synapse. In the process of information exchange between neurons, the signals inputted through the axon protrusions of several different neurons arrive at the dendrites of specific neurons. All input signals are converted into pulse signals, and when the sum of these signals reaches a certain threshold, one signal is transmitted to the other neuron through the axon. Learning is accomplished through this series of activities of all neurons. The human brain has a conceptually simple structure composed of neurons as above, but can perform complex tasks despite its simplicity of processing. The neural network analysis algorithm is implemented by machine learning of the connection relationship between neurons as above (Beale & Jackson, 1990).

5.2.1 Logistic Regression

Logistic regression is one of the analysis methods of early machine learning. Logistic regression analysis has the same as linear regression analysis, where the relationship between dependent and independent variables is expressed as a mathematical function and used in future prediction models. The main difference is that the dependent variable in linear regression is continuous and of logistic regression is categorical. Because binomial dependent variables are primarily used, the sigmoid function is used to separate the result by 0 for values less than 0 and 1 for values greater than 1. Therefore, it can be seen as a classification technique because the results are divided into specific categories (Tu, 1996).

However, despite the advantages of logistic regression, which can easily represent analytical results with mathematical functions, there is the disadvantage of using a single-layer perceptron (SLP) in terms of neural networks (Yoldaş, Tez & Karaca, 2012). This simplified single-layer perceptron has poor predictive power because of the eXclusive OR, XOR problem and the accuracy of the model is degraded. As shown in Figure 19, it is not possible to distinguish between a single line or a mathematical expression in order to distinguish between grey dots and black dots in a simple XOR problem (Goodfellow, Bengio & Courville, 2016). Thus, to distinguish whether a review is helpful or not, logistic regression is easy to understand by providing a predictive model as a mathematical function. As a result of the logistic regression analysis using the independent variables used in this study, the actual result prediction accuracy and the predictive mathematical function can be verified.

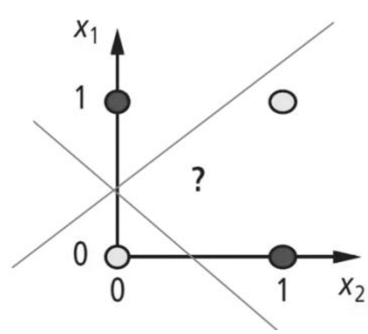

Figure 19. eXclusive OR, XOR problem

5.2.2 Deep Learning

To solve the XOR problem in SLP, the neural network can be composed of several nodes in the hidden layer. This is traditionally called multi-layer perceptron, MLP, but is more often now called deep learning. In SLP, the classification of XOR problem was mathematically impossible, but in deep learning, which uses a hidden layer, classification became mathematically possible. Therefore, it is possible to improve the accuracy of predictions by enabling the classification of data of complex structures such as big data. Deep learning, like logistic regression, is also used as an analytical tool for categorical data that can be used for classification and prediction (Tu, 1996).

The basic structure and process of a neural network are useful to know. Each object that receives the initial data can be seen as a neuron, and is the input layer. Through the hidden layer, the data of each neuron is collected through stimulation of weight. Supervised learning, used for this study, is the most common form of machine learning; it lets the machine knows what the correct answer is and trains it until they find the optimized model (LeCun, Bengio & Hinton, 2015). In logistic regression, the predictive model can be expressed as a mathematical function, whereas the hidden layer of deep learning is a 'black box,' so the process of summing and transmitting the information is not shown. In the output layer, the hidden layer is sent to output with summed information as a result. A function model is used to transfer values from neuron to neuron, which model is called an active function. Like logistic regression, deep learning usually uses the sigmoid function, but it also uses various active functions, namely hyperbolic tangent, softmax and maxout functions, for improved deep learning. In the output layer, the back-propagation process is repeated, where new results are compared to existing results and delivered to the hidden layer to compensate for the weight. If the outcome of the output layer reaches a certain level and is no longer improved, the iterative learning process is stopped, and the structure of the iterative process of this neural network is called perceptron. Through perceptron, deep learning using MLP is more accurate than classifications of logistic regression, so it can be distinguished more clearly whether a review is helpful or not (Goodfellow, Bengio, & Courville, 2016). However, since the prediction model is a black box, there is a disadvantage that only classifications using computer modelling calculation can be applied. Deep learning analysis using the independent variables used in this study can be used to verify the predicted accuracy of the results, which is higher than that of logistic regression (Tu, 1996).

Deep learning with multi-layer perceptron technology has become the most advanced machine learning technologies available today. It has been particularly successful in areas such as voice recognition, image analysis, and natural language processing (Zhang et al., 2018). This is the technique I have selected to use in the following analysis, deep learning using multi-layer perceptron. Bi-logistic regression analysis was also performed to verify the difference (if any) in prediction effects between multi-layer perceptron and single-layer perceptron.

5.3 Analysis for verification of the combination of the causal variables

The purpose of this analysis is to verify how the combination of the causal variables, namely star ratings, the type of information, and length of comments, affect perceived helpfulness of the reviews. To do this, this research compares the results of bi-logistic regression, a traditional machine learning technique and the deep learning, the more novel machine learning technique, to see how accurately perceived helpfulness of the reviews can be predicted.

5.3.1 Data Set

To analyse by machine learning techniques, this research divided Booking.com reviews into 50% to 50% with all helpful reviews (N=73,262) and randomly selected not helpful reviews (N=73,262) excluding zero-length reviews. This is done as machine learning can make biased predictions when using an unbalanced data set. As both bi-logistic regression and deep learning are used when the dependent variable is categorical, perceived helpfulness of the reviews is coded as 0 (not helpful) and 1 (helpful). Agoda.com data was excluded from the fifth analysis as it does not provide positive and negative aspects.

5.3.2 The results of bi-logistic regression analysis

This thesis conducted a bi-logistic regression analysis (Hair et al., 2013) with the dependent variable set as perceived helpfulness of the reviews and the independent as star ratings, the type of information and length of comments.

Firstly, the results of the Hosmer and Lemeshow test show that goodness-of-fit for the logistic regression model is significant ($\chi2(8) = 1143.084$, $p < .001$), and the explanatory power of the logistic regression model is 7.6% (Nagelkerke $R_2 = .076$). Second, as shown in results of the significance test of the regression coefficients, the main effects of star ratings (B = -.051, SE = .003, Wald = 374.067, $p < .001$), the type of information (B = -.020, SE = .006, Wald = 10.248, $p = .001$), and length of comments (B = .002, SE = .000, Wald = 5664.841, $p < .001$) are all significant. Third, I find that the main effect of the constant (B = -.066, SE = .022, Wald = 8.669, $p = .003$) is also significant. The formula for predicting perceived helpfulness of the reviews is as follows: logit (perceived helpfulness of the reviews) = (-.051 * star ratings) + (.002 * the type of information) + (-.02 * length of comments) - .066, in that the predictive accuracy of the bi-logistic regression model is 59.9%. Tables 26 and 27 show a summary of the results for Booking.com.

Overall, when the accuracy is less than 60%, the machine learning effect size is considered small, and the statistical effect size of 0.4, is also small, based on Cohen's classification (Cohen, 2013; Schnack & Kahn, 2016).

Table 26: Results of fifth analysis: bi-logistic regression analysis for Booking.com

IV	*B*	S. E.	Wald	P value	*OR*	95% CI
Star Ratings	-.051	.003	374.067***	.000	.950	(.945~.955)
Type of information	.002	.000	5664.841***	.000	1.002	(1.002~1.002)
Length of comments	-.020	.006	10.248***	.001	.980	(.968~.992)
Constant	-.066	.022	8.669**	.003	.936	
-2LL=194576.145, Nagelkerke R Square=.076, Hosmer and Lemeshow Test: $\chi2=1143.084$($p<.001$)						

Table 27: Classification table of fifth analysis: bi-logistic regression analysis for Booking.com

Observed		**Predicted**		
		Helpfulness		**Percentage Correct**
		Not helpful	**Helpful**	
Helpfulness	Not helpful	54085	19177	73.80%
	Helpful	39637	33625	45.90%
Overall Percentage				59.90%

5.3.3 The deep learning analysis results

I conducted the deep learning analysis (IBM, 2019) with the dependent variable again perceived helpfulness of the reviews, with as star ratings and length of comments as covariates, and 'Factor' as the type of information. Continuous variables have been rescaled as standardized methods. As shown in Figure 20, a multi-layer perceptron model with double hidden layers was used; a hyperbolic tangent is used as the activation function. In the output layer, softmax is used as the activation function. 70% of the data was used to create the model, and 30% of the data was used to verify it (see Figure 20).

First, I find that receiver operating characteristic (ROC) of a perfect predictive model in the area under the curve (AUC) are all .651 and the fit of the model is poor ($.6 <$ ROC AUC $\leq .7$) (Shipitsyna et al., 2013). Second, the importance analysis of the independent variable shows that the length of comments is 100% important, star ratings are 14.9% important, and the type of information is 6.1% important.

Finally, as shown in Table 28, the predictive accuracy of the neural network model is 60.6%, which is 0.7% higher than the 59.9% of the bi-logistic regression analysis. Specifically, since the initial independent variable consisted 50/50 helpful and not helpful reviews, the accuracy of prediction was improved by 10.6% through the neural network model using three independent variables of star ratings, length of comments, and the type of information on perceived helpfulness of the reviews. Table 28 shows a summary of the results for Booking.com.

To formally compare statistical effect size within the same data set, PROCESS macro (Model 3, 5000 bootstrap samples) was utilised (Hayes, 2018), where moderation analysis is conducted, with star ratings as an independent variable, both the type of information and length of comments as moderating variables, and perceived helpfulness of the reviews as the dependent variable. As shown in Table 29, although effect sizes are bigger than the analysis for hypothesis 4 (absolute value (all 'Effect') < .01), the effect sizes of this data set still appear small (absolute value (all 'Effect') < .05).

Overall, when the accuracy is greater than 60% and less than 70%, the machine learning effect size is considered modest, and the statistical effect size is considered to be approximately 0.5, which is between small and medium based on Cohen's classification (Cohen, 2013; Schnack & Kahn, 2016). Thus, when the statistical effect size is estimated by the machine learning effect size, the effect size problem is not diagnosed.

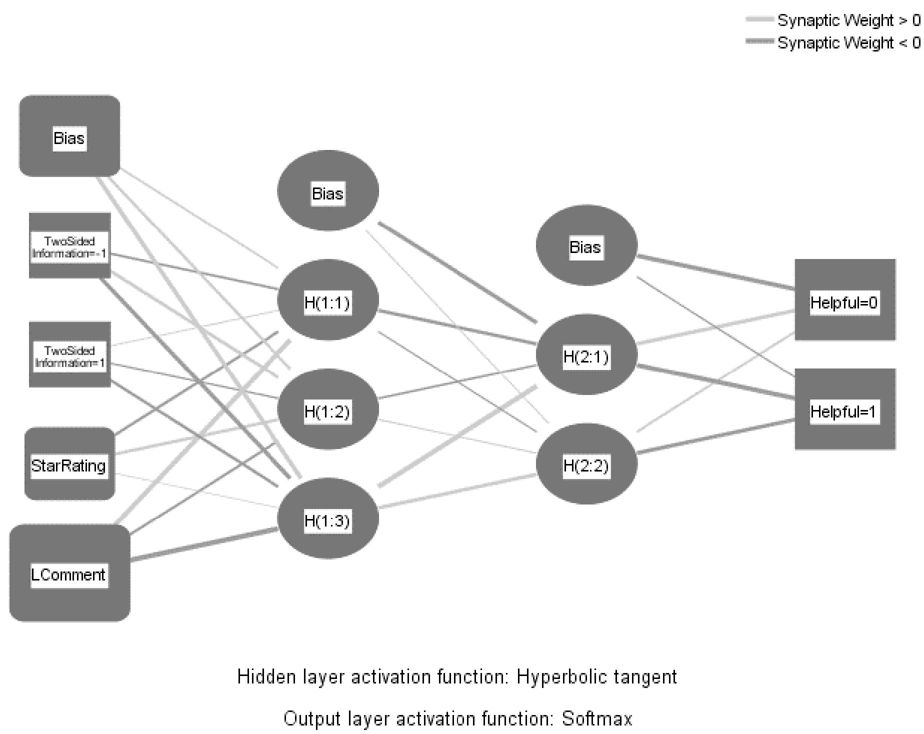

Figure 20: Multi-layer perceptron model with double hidden layers

Table 28: Results of fifth analysis: deep learning analysis for Booking.com

Sample	Observed	Predicted		
		Not helpful	Helpful	Percent Correct
Training	Not helpful	32141	19010	62.80%
	Helpful	21073	30341	59.00%
	Overall Percent	51.90%	48.10%	60.90%
Testing	Not helpful	13807	8304	62.40%
	Helpful	9033	12815	58.70%
	Overall Percent	52.00%	48.00%	60.60%

Table 29: Results of fifth analysis: Hayes model 3 for Booking.com

Model	coefficient	se	t	P	LLCI	ULCI
constant	.7325	.0200	36.5652	.0000	.6933	.7718
Star Ratings	-.0436	.0024	-17.8945	.0000	-.0484	-.0388
Length of comments	.0005	.0001	8.1495	.0000	.0004	.0006
Interaction 1	.0001	.0000	11.2268	.0000	.0001	.0001
Type of information	-.0901	.0239	-3.7658	.0002	-.1370	-.0432
Interaction 2	.0141	.0030	4.7799	.0000	.0083	.0199
Interaction 3	-.0002	.0001	-3.8934	.0001	-.0004	-.0001
Interaction 4	.0000	.0000	-2.3489	.0188	.0000	.0000
Moderator	**Effect**	**se**	**T**	**P**	**LLCI**	**ULCI**
4, one-sided	-.0432	.0024	-17.8909	.0000	-0.048	-0.0385
4, two-sided	-.0292	.0017	-17.4688	.0000	-0.0325	-0.0259
264, one-sided	-.0205	.0020	-10.2283	.0000	-0.0244	-0.0165
264, two-sided	-.0116	.0013	-9.2222	.0000	-0.0141	-0.0091
567, one-sided	.0061	.0035	1.7312	.0834	-0.0008	0.0129
567, two-sided	.0088	.0015	5.8876	.0000	0.0059	0.0118

Product terms key:
Interaction 1: Star Ratings x Length of comments
Interaction 2: Star Ratings x Type of information
Interaction 3: Length of comments x Type of information
Interaction 4: Star Ratings x Length of comments x Type of information

Chapter 6: General Discussion and Conclusion

6.1 Introduction

Having presented and elaborated on the results of five analyses and three alternative analyses, this chapter will present and highlight the findings and how they relate and link back to the main research question; 'When are Negative Reviews More Helpful than Positive Reviews to Customers?' This chapter goes on to discusses contributions, theoretical and managerial implications as well as limitations and future research directions.

6.2 Discussion

A major aim of my study was to examine identify factors that affect which reviews are perceived as helpful, as previous research has yielded different results. For instance, previous research suggested that review valence has a positive (Liu and Park, 2015), a negative (Willemsen et al., 2011), a U-shape (Forman, Ghose and Wiesenfeld, 2008), an inverted U-shape (Mudambi and Schuff, 2010) relationship with the perceived helpfulness of the reviews. Within the theoretical framework of negativity bias and loss aversion (Barkley-Levenson et al., 2013; Tversky & Kahnemank, 1992), negative information has a greater impact than the positive information. Findings also show that customers perceive negative reviews are more helpful than positive reviews. Furthermore, findings contribute an explanation of systemic moderators eliciting different outcomes.

Analysis uncovers a significant moderating role of narrative reviews difference (i.e., consumer scepticism, too good to be true and the systematic information processing). To make the results of analysis more robust, big data – consisting of over two million online hotel reviews – were collected in an attempt to obtain more reliable research results. Also, to address a small statistical effect size, the effect of the independent variables is not only measured in traditional machine learning method (bi-logistic regression), but also in the recent machine learning method (deep-learning) (Schnack & Kahn, 2016). These differences in perceived helpfulness of online reviews are important to the business of tourism, as the online reviews are an increasingly salient channel in that customers make hotel selection decisions (Vermeulen & Seegers, 2009).

The findings contribute to the literature in one more way, too. That is, by providing boundary conditions where review valence could have a positive or negative impact on the perceived helpfulness of the reviews.

Specifically, as illustrated in the results section, this research delivers insights regarding factors influencing perceived helpfulness of the reviews in the context of online hotel customer reviews. To this end, various analyses based on big data were conducted to identify the main research question: 'When are negative reviews more helpful than positive reviews to customers?'

First, despite the mixed findings of previous research, I expected that because of negativity bias and loss aversion, negative reviews would have a greater effect on the perceived helpfulness of the reviews than positive reviews. In the first analysis, regression analysis of both Booking.com and Agoda.com do show the negativity bias effect, supporting the hypothesis that 'the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews.' Thus, in the first answer to the main research question, negative reviews are more helpful than positive reviews to customers.

Second, I also expected that because of consumer scepticism, the 'too good to be true' idea, that two-sided reviews would have a more significant effect on the perceived helpfulness of the reviews than one-sided reviews. Specifically, two-sided information not only improves source credibility and believability more than one-sided information, but also positively affects the formation of attitudes of customers by reducing consumer scepticism of positive reviews over negative. In the second analysis, a significant two-way interaction effect between star ratings and the type of information on perceived helpfulness of the reviews in Booking.com is found, supporting the hypothesis that the type of information moderates the impact of review valence on the perceived helpfulness of the reviews. Again, more specifically, when reviews are one-sided, review valence has a strong negative impact on the perceived helpfulness of the reviews. However, when reviews are two-sided, the negative impact of review valence on the perceived helpfulness of the reviews is relatively weak. Thus, in answer to the main research question, two-sided negative reviews are less likely to be more helpful to customers than two-sided positive reviews.

My third expectation was that because of induced systematic information processing, longer reviews will have a more significant effect on the perceived helpfulness of the reviews than shorter reviews. Specifically, the helpfulness of negative hotel reviews will be more significantly influenced by strengthened loss aversion through the peripheral route of heuristic information processing. In contrast, the helpfulness of negative hotel reviews will be less influenced by weakened loss aversion through the central route of systematic information processing. In the third analysis, a significant two-way interaction effect between star ratings and length of comments on perceived helpfulness of the reviews in Booking.com is found, supporting the hypothesis that systematic information processing does indeed moderate the impact of review valence on the perceived helpfulness of the reviews. Thus, when the length of reviews is shorter, review valence has a strong negative impact on their perceived helpfulness. However, when reviews are longer, review valence has a positive impact on their perceived helpfulness. However, the opposite result is obtained for Agoda.com. In this case, I assume that different mechanisms work depending on whether two-sided information is provided (Booking.com) or only one-sided information is provided (Agoda.com). Nevertheless, Hypothesis 3 is supported in the Booking.com data, which is the main source for this study. Thus, in answer to the main research question, longer positive reviews are more helpful than longer negative reviews to customers.

Fourth, after analysing the literature, I expected that that only when reviews are two-sided with longer comments might the negativity bias and loss aversion become eliminated, and result in the helpfulness of positive reviews being more significantly influenced. In the fourth analysis, a significant three-way interaction effect between star ratings and the type of information and length of comments on perceived helpfulness of the reviews in Booking.com is found, supporting the hypothesis that both length of comments and type of information moderate the impact of review valence on the perceived helpfulness of the reviews.

Specifically, when reviews contain two-sided information, my analysis suggests that review valence has a positive relationship with the perceived helpfulness of the reviews with longer comments, whereas, review valence has a negative impact on the perceived helpfulness of the reviews in shorter comments. Contrarily, when reviews are one-sided, review valence has a negative impact on the perceived helpfulness of the reviews in both shorter and longer comments. Thus, in answer to the main research question, only two-sided and longer positive reviews are more helpful than two-sided and longer negative reviews to customers.

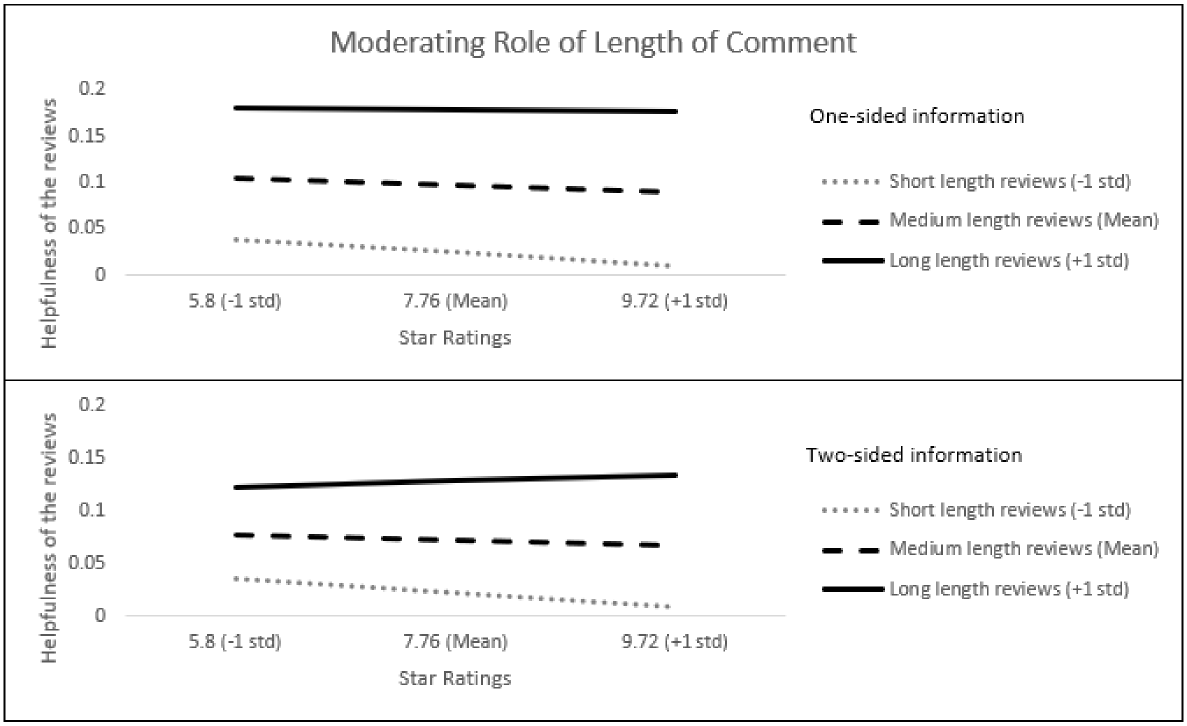

Figure 21: The moderating role of both consumer scepticism, “Too good to be true!” and systematic (vs. heuristic) information processing

My fifth hypothesised expectation was that analysing the effect of the independent variables with a machine learning method would make the results of the analysis more robust by addressing the small statistical effect size issue. In order to measure and apply advanced machine learning techniques, the effect of the independent variables was measured by comparing both traditional machine learning technique (using single-layer perceptron) and a newer machine learning technique (using multi-layer perceptron, or deep learning). In the fifth analysis, the predictive accuracy of the deep learning model is 60.6%, which is .7% higher than the 59.9% of the bi-logistic regression analysis, suggesting that when the three important variables used in this study are put into deep-learning, AI can predict whether a review will be helpful or not to customers with approximately 60% accuracy. Thus, in answer to the main research question, AI can predict whether a positive or negative review will be helpful with 60.6% accuracy. Also, when the statistical effect size is estimated by the machine learning effect size, the effect size problem is not an issue.

Finally, I expected that alternative explanations would fail to explain the main results. For scarcity effects, both hierarchical regressions analyse and significant two-way interaction effects suggest that the pattern of results remains the same and so the scarcity effects fails to explain the results. As for cultural diversity, even if the reviewers were divided into eight visited cities or eighty-two reviewers' nationalities, negativity bias, negative reviews still convey greater perceived helpfulness of the reviews than positive reviews. Thus, in answer to the main research question, despite many positive reviews, dividing into different reviewers' nationalities or dividing into different visited cities, negative reviews are perceived to be more helpful than positive reviews to customers.

6.3 Theoretical contribution

In the remainder of this research, the major theoretical and practical contributions of this current work are highlighted, along with reflections of potential limitations. The first contribution is to the ever-increasing stream of literature on online customer reviews and eWOM, providing new insights into what factors influence perceived helpfulness of reviews. While most studies have provided mixed findings (Liu & Park, 2015; Mudambi & Schuff 2010; Schindler & Bickart 2012; Schlosser, 2011; Willemsen et al. 2011), this study provides systemic moderators that elicit different outcomes, using a big dataset with over two million online customer reviews. I also examined the combination of the causal variables with a machine learning method, artificial neural network, in order to make the results of analysis more robust. I believe this is the first comprehensive study that explores the impacts of star rating on the perceived helpfulness of the reviews using such mixed methods; a combination of statistical, traditional and novel machine learning methods.

The second contribution is that the findings highlight that there is a negativity bias and loss aversion regarding the perceived helpfulness of the reviews. Notably, when analysed together with alternative theories, the findings document a negative association between the review valence and the perceived helpfulness of the reviews. The findings consistently highlight that negativity bias and loss aversion in the perceived helpfulness of the reviews can be found in (i) both Booking.com and Agoda.com, (ii) eight different visited cities, (iii) eighty-two different reviewers' nationalities, and (iv) a majority of positive reviews and a minority of negative reviews. This implies that because of negativity bias, loss aversion, customers are putting more weight on negative reviews on the Internet (in the hotel booking situation, at least).

The third contribution is the findings show that there is a moderating role for 'too good to be true' consumer scepticism, in the perceived helpfulness of the reviews. The findings document that when the level of consumer scepticism decreases the negativity bias and loss aversion are also weakened. The results also show that there is a moderating role for systematic information processing regarding the perceived helpfulness of the reviews. The findings document that when the level of heuristic information processing decreases, then negativity bias and loss aversion also weaken. So, when both consumer scepticism, and heuristic information processing decrease, the negativity bias and loss aversion are eliminated (or minimised) resulting in conversion to a positivity bias.

Arguably, this is the first research to suggest the significant moderating variables of consumer scepticism and heuristic information processing that could alter the perceived helpfulness of the reviews. That is, the findings shed light upon the boundary conditions where review valence could have a positive or negative impact on the perceived helpfulness of the reviews. Also, this research facilitates a higher predictive effect than surveys or small sample sizes used in previous studies by utilizing big data (Siegel, 2013).

6.4 Managerial implications

This thesis has a potential impact on practice in the travel and hospitality sectors. Online customer reviews can be a valuable tool for managers to not only navigate customer preferences efficiently, but also to navigate what information related to their hotel is helpful to consumers efficiently. Compared with traditional standard methods of measuring customer awareness of helpful information related to hotels, hotels that use online reviews as a performance measurement tool can extract more meaningful information for real-life situations. Also, artificial intelligence analysis of big data with online reviews enables managers to design customer segmentation and predictive model that traditional surveys cannot provide (Al-Jarrah et al., 2015). Traditional survey-based methods can also provide a valid source of information. However, such surveys have limitations because they require time, limited sample selection and costs (Stamolampros et al., 2019), and often even generate false information from jaded, survey-saturated, respondents. This study provides insight into how managers can leverage their own big data from online reviews to understand how they can deliver helpful information related to their hotels to encourage potential customers to make hotel selection decisions.

At the same time, this study has found that, generally, negative reviews are more helpful than positive reviews to customers. Contrary to traditional beliefs, if managers encourage their customers to write only positive reviews with positive content, they may not be considered by customers as helpful as negative reviews. Thus, if managers do not take any action, consumers will continue to assess negative reviews as more helpful and adversely affect the hotel's decision-making process for potential consumers. Fortunately, this study suggests that the negativity bias and loss aversion tendency of current consumers can be turned into positivity bias. If managers encourage their customers to write positive reviews with details including negative aspects, potential customers will evaluate positive reviews as helpful. This will ultimately have a positive impact on the hotel selection process of potential customers.

This study can provide greater benefits to customers who want to evaluate hotels. Consumers inevitably give recommendations to and receive recommendations from others. When customers recommend a hotel to others and want to have a real impact on their choices, listing only good experiences will not be considered a helpful review for others. To improve their recommendations for others, this research recommends providing as much detail as possible in their online reviews including negative information. In contrast, when customers do not recommend a hotel to others and want to have a real impact on their choices, listing only bad experiences will be considered a helpful review for others. To improve their recommendations for others, this research recommends providing as much detail as possible in their online reviews excluding positive information.

In a similar vein, when making their hotel choice self-awareness can lead to a better decision. If a potential customer is aware that they tend to value negative reviews more than positive, perhaps they can consciously accord equal weight to both positive and negative reviews and thus make a better choice.

Ultimately, this study can provide critical clues to increasing hotel booking rates by increasing helpful reviews. Chen et al. (2008) found that the helpful reviews that received helpfulness votes had a positive effect on sales. Thus, hotels with many helpful reviews that received helpfulness votes could have higher hotel reservation rates. As seen in the findings of this study, in order for a review to receive helpfulness votes: 1) both positive and negative information must be mentioned and 2) the length of the comment must be relatively long. Giving it a nudge may be the best way to effect change. For instance, when designing a hotel evaluation form, this can be achieved with a simple design change; nudge: 1) providing both advantages and disadvantages input boxes rather than only one comment input box. 2) the minimum number of input character limits should be required for both advantages and disadvantages input boxes. These nudges would allow hotels to get balanced and longer online reviews from consumers with both positive and negative information, which would eventually play a positive role in increasing the hotel booking rates.

6.5 Limitations and future research directions

Despite the potential usefulness of research using big data, this study also has general limitations that derive directly from the nature of online reviews. Several biases are set in the literature that analyses online reviews data, such as self-selection and response biases (Stamolampros et al., 2019). Also, online reviews can be manipulated by the temptation to increase economic profits (Liu & Glance, 2012). This study did not have access to source data that stores sensitive personal information of customers and can only collect published online reviews. This only allows me to control the published factors and not the non-published factors including various demographic factors such as names, gender, age, and income level.

In order to overcome the above limitations, spatial regression discontinuity design for big data was used to empirically compare the differences between groups, as in an experiment. Furthermore, large sample size can be considered to better represent the entire population rather than outlier results of the overall variation result in unbiased results. Apart from the general limitations shown above, some more specific limitations were found during the analysis and are highlighted in the following text.

6.5.1 The inconsistent finding of the moderating role of systematic information processing

One unanticipated finding was that PROCESS macro analysis used to test the moderating role of systematic information processing showed that a significant two-way interaction effect between star ratings and length of comments on perceived helpfulness of the reviews are significant, but that an opposite pattern exists for Booking.com and Agoda.com. This finding is unexpected and suggests that the moderating role of systematic information processing may enhance or weaken negativity bias and loss aversion in the perceived helpfulness of the reviews. Thus, future work could be extended utilising extended meaningful moderator variables to address this inconsistent finding.

6.5.2 Main effect of star ratings on the perceived helpfulness of the reviews.

As shown in mixed findings, opposite results were shown depending on the type of experience product. In this study, however, only the hotel selection experience was analysed. Before meaningful generalisations can be made, it is necessary to verify whether the analysis results of the products of various experience product lines are consistent with this study. Also, another unanticipated finding was that regression analysis used to predict the negativity bias shows that the main effects of star ratings on the perceived helpfulness of the reviews are significant but very small for both Booking.com and Agoda.com. This finding was unexpected and suggests that star ratings may not have a great impact on perceived helpfulness of the reviews. Thus, future work could be extended utilising extended meaningful independent variables to address this potential issue.

6.5.3 Different magnitude of negativity bias based on different gender, cities and nationalities

As mentioned previously, control over the published factors and not the non-published factors are limiting for this kind of research. However, this study recommends that non-published factors may be inferred by convergence with other data sets. For instance, when comparing the name and gender recorded in the official census data with the reviewers’ name, the reviewer's gender can usually be categorized as male, female and unisex (GenderChecker, 2019). This study suggests that different gender may be used as extended meaningful independent variables.

Secondly, as shown in the cultural diversity analysis, the results based on eight different visited cities confirm that the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews. Interestingly, there were also differences in the magnitude of negativity bias based on different cities. This may simply be a sampling issue but could suggest that different cities can be categorized by characteristics and used as extended meaningful independent variables.

Third, as shown in the cultural diversity analysis, the results based on eighty-two reviewers' nationalities confirm that the star ratings of online reviews have a negative relationship with the perceived helpfulness of the reviews. Interestingly, though, there were also differences in the magnitude of negativity bias based on different nationalities. Whether this is, again, a random sampling error or that different cities can also be categorized by characteristics such as language and used as extended meaningful independent variables is if considerable interest.

6.5.4 Most powerful combination of the causal variables

Together these results provide important insights into how to analyse big data. This work suggests that big data analysis requires not only theory-based traditional approach but also data-driven decision-making approach.

(i) Theory-based traditional approaches can analyse the data according to the theory after establishing the hypothesis through literature review. The advantage is that the research is based on theory. The disadvantage is that effective analysis of big data itself is difficult.

(ii) A data-driven decision-making approach can analyse the most effective causal variables through artificial intelligence: The advantages here is that it makes good use of large data, but the difficulty is that it sometimes becomes hard to explain in theoretical terms (Al-Jarrah et al., 2015).

In the medical sector, for example, the accuracy of diagnosis (in very narrow application areas) using big data-based machine learning has begun to rise higher than that of human medical doctors (Obermeyer & Emanuel, 2016). Thus, it is vital for brand-owners and marketers to identify most important factors that affect which reviews are perceived as helpful using data-driven decision-making approach. This study suggests that qualitative comparative analysis (fs-QCA) to determine the success recipes, providing a combination of causal variables that have the greatest effect on the perceived helpfulness of consumer reviews. fs-QCA is a set-theoretic approach based on the assumption that the various solutions can be equally effective in achieving the final effect. The combination of the cause conditions identified in the success recipes can effectively describe the case (Díaz-Fernández et al., 2019). The predictive and fit validity of the success recipe could be confirmed by machine learning. This study suggests that future research should go through a transformational shift from expert systems to artificial intelligence. Thus, future work should be extended utilising data-driven decision-making approach to address this potential research.

6.6 Conclusion

This study was conducted based on the theoretical framework of negativity bias and loss aversion (Barkley-Levenson et al., 2013; Tversky & Kahnemank, 1992). That negative online reviews have a greater impact than positive has been strongly supported. It seems reasonably certain that customers perceived negative hotel reviews as more helpful than positive reviews. This work enables a higher predictive effect than surveys or small sample sizes used in previous studies by utilizing big data. Furthermore, this study finds systemic moderators eliciting different outcomes. Findings uncover significant moderating roles for narrative review difference (i.e., consumer scepticism, "too good to be true" and the systematic information processing) and that only two-sided and longer positive reviews are more helpful than two-sided and longer negative reviews to customers.

By conducting this research, I hope to have contributed a little more certainty in the body of research around the value and perception of hotel reviews. Hopefully, too, the use of big data and deep learning may inspire other researchers to probe yet deeper into this domain and to then generalise the work into a wider consideration of different domains.

References

Akehurst, G. (2009). User generated content: The use of blogs for tourism organisations and tourism consumers. *Service Business*, 3(1), 51–61.

Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015). Efficient machine learning for big data: A review. *Big Data Research*, 2(3), 87-93.

Anderson, E. W. (1998). Customer satisfaction and word of mouth. *Journal of Service Research*, *1*(1), 5-17.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 243-268.

Barkley-Levenson, E. E., Van Leijenhorst, L., & Galván, A. (2013). Behavioral and neural correlates of loss aversion and risk avoidance in adolescents and adults. *Developmental Cognitive Neuroscience*, 3, 72-83.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323-370.

BBC News. (2019). TripAdvisor defends itself in fake reviews row. Retrieved 23 November 2019, *from https://www.bbc.com/news/business-49605457*

Beale, R., & Jackson, T. (1990). Introduction to neural networks. *Bristol*.

Bei, L. T., Chen, E. Y., & Widdows, R. (2004). Consumers' online information search behavior and the phenomenon of search vs. experience products. *Journal of Family and Economic Issues*, 25(4), 449-467.

Belch, G. E. (1981). An examination of comparative and noncomparative television commercials: The effects of claim variation and repetition on cognitive response and message acceptance. *Journal of Marketing Research*, 18(3), 333-349.

Bonaccorso, G. (2017). Machine learning algorithms. *Packt Publishing Ltd*.

Booking Holdings. (2019). Booking Holdings to Make Fourth Quarter 2018 Earnings Press Release Available on Company's Investor Relations Website on Wednesday, February 27 | Booking Holdings. *Retrieved from http://ir.Bookingholdings.com/index.php/ news-releases/news-release-details/Booking-holdings-make-fourth-quarter-2018-earnings-press-release*

Booking.com. (2019). Booking.com: The largest selection of hotels, homes, and holiday rentals. *Retrieved from https://www.Booking.com/reviews.en-gb.html*

Burmeister, E., & Aitken, L. M. (2012). Sample size: How many is enough?. *Australian Critical Care*, 25(4), 271-274.

Bulchand-Gidumal, J., Melián-González, S., & López-Valcárcel, B. G. (2011). Improving hotel ratings by offering free Wi-Fi. *Journal of Hospitality*, 2(3), 235–245.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, 76(5), 839.

Cameron, A. C., & Trivedi, P. K. (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics*, *46*(3), 347-364.

Chatterjee, P. (2001). Online reviews: do consumers use them? *ACR Proceedings*, pp. 129.134.

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.

Cheung, M. Y., Luo, C., Sia, C. L., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *International Journal of Electronic Commerce*, 13(4), 9-38.

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.

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.

Chollet, F. (2018). Deep learning with Python. Shelter Islands: Manning.

CNBC. (2019). Prime Day is coming: Tips to spot a fake review on Amazon. *Retrieved 8 August 2019, from https://www.cnbc.com/2019/07/12/prime-day-tips-for-spotting-a-fake-reviews-on-amazon.html*

Cox, C., Burgess, S., Sellitto, C., & Buultjens, J. (2009). The role of user-generated content in tourists' travel planning behavior. *Journal of Hospitality Marketing & Management*, 18(8), 743–764.

Crowley, A. E., & Hoyer, W. D. (1994). An integrative framework for understanding two-sided persuasion. *Journal of Consumer Research*, 20(4), 561-574.

De Langhe, B., Fernbach, P. M., & Lichtenstein, D. R. (2015). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *Journal of Consumer Research*, 42(6), 817-833.

Dellarocas, C., & Narayan, R. (2006). A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, *21*(2), 277-285.

Díaz-Fernández, M. C., Rosario, M., & Simonetti, B. (2019). Top Management Team diversity and high performance: an integrative approach based on Upper Echelons and Complexity Theory. *European Management Journal*.

Diener, E., & Diener, C. (1996). Most people are happy. *Psychological science*, 7(3), 181-185.

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

Duverger, P. (2013). Curvilinear effects of user-generated content on hotels' market share: a dynamic panel-data analysis. *Journal of Travel Research*, 52(4), 465-478.

Dwivedi, M., Shibu, T. P., & Venkatesh, U. (2007). Social software practices on the internet: Implications for the hotel industry. *International Journal of Contemporary Hospitality Management*, 19(5), 415–426.

Etgar, M., & Goodwin, S. A. (1982). One-sided versus two-sided comparative message appeals for new brand introductions. *Journal of Consumer Research*, 8(4), 460-465.

Ferris, G. R., Frink, D. D., & Galang, M. C. (1993). Diversity in the Workplace: The Human Resources Management Challenges. *Human Resource Planning*, 16(1).

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.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. *London: sage*.

Filieri, R., & McLeay, F. (2014). E-WOM and accommodation: An analysis of the factors that influence travellers' adoption of information from online reviews. *Journal of Travel Research*, 53(1), 44-57.

Frías, D. M., Rodriguez, M. A., & Castaneda, J. A. (2008). Internet vs. travel agencies on pre-visit destination image formation: An information processing view. *Tourism Management*, 29(1), 163-179.

Gauri, D. K., Bhatnagar, A., & Rao, H. R. (2008). Role of word of mouth in online store loyalty. Commun. *ACM*, 51(3), 89-91.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, over dispersed Poisson, and negative binomial models. *Psychological Bulletin*, *118*(3), 392.

Gelman, A., & Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. *Cambridge University Press*.

GenderChecker. (2019). GenderChecker - The number one database for checking male, female and unisex names. *Retrieved 25 August 2019, from https://genderchecker.com/*

Golden, L. L., & Alpert, M. I. (1987). Comparative analysis of the relative effectiveness of one-and two-sided communication for contrasting products. *Journal of Advertising*, 16(1), 18-68.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. *MIT Press*.

Gretzel, U., Fesenmaier, D. R., Lee, Y. J., & Tussyadiah, I. (2010). 11 Narrating travel experiences: the role of new media [J]. *Tourist Experience: Contemporary Perspectives*, 171.

Gu, B., & Ye, Q. (2014). First step in social media: Measuring the influence of online management responses on customer satisfaction. *Productions and Operations Management*, 23(4), 570–582.

Guijarro-Berdiñas, B., Fontenla-Romero, O., Pérez-Sánchez, B., & Fraguela, P. (2007). A linear learning method for multilayer perceptron's using least-squares. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 365-374). Springer, Berlin, Heidelberg.

Hart, C., & Blackshaw, P. (2006). Internet Inferno-One customer can take down your company, but you can turn the potential nightmare into a boon. *Marketing Management*, *15*(1), 18.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2013). Multivariate data analysis: Pearson new international edition. *Pearson Higher Ed*.

Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis second edition: A regression-based approach. *New York, NY: Guilford Press*

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and social psychology Review*, 10(1), 47-66.

Hastak, M., & Park, J. W. (1990). Mediators of message sidedness effects on cognitive structure for involved and uninvolved audiences. *ACR North American Advances*.

Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The effects of positive and negative online customer reviews: do brand strength and category maturity matter? *Journal of Marketing*, 77(6), 37-53.

IBM. (2019). IBM SPSS Neural Networks. *Retrieved 26 August 2019, from http://www.spss.com.hk/software/statistics/neural-networks/*

Internet Live Stats. (2019). 1 Second - Internet Live Stats. *Retrieved 25 August 2019, from https://www.internetlivestats.com/one-second/*

Ito, K. (2014). Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. *American Economic Review*, 104(2), 537-63.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, 75(4), 887.

Ito, T., & Cacioppo, J. (2005). Variations on a human universal: Individual differences in positivity offset and negativity bias. *Cognition & Emotion*, 19(1), 1-26.

Jeong, E., & Jang, S. S. (2011). Restaurant experiences triggering positive electronic word-of-mouth (eWOM) motivations. *International Journal of Hospitality Management*, 30(2), 356–366.

Jeong, M. Y., & Jeon, M. H. M. (2008). Customer reviews of hotel experiences through consumer generated media (CGM). *Journal of Hospitality Marketing & Management*, 17(1), 121–138.

Kahneman, D. (2011). Thinking, fast and slow. New York, *NY: Farrar, Straus and Giroux*.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1), 193-206.

Kamins, M. A. (1989). Celebrity and noncelebrity advertising in a two-sided context. *Journal of Advertising Research*.

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), 29-39.

Kamins, M. A., & Marks, L. J. (1987). Advertising puffery: The impact of using two-sided claims on product attitude and purchase intention. *Journal of Advertising*, 16(4), 6-15.

Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5), 6000–6010.

Keele, L. J., & Titiunik, R. (2015). Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1), 127-155.

Kiley Hamlin, J., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, 13(6), 923-929.

Kim, E. E. K., Mattila, A. S., & Baloglu, S. (2011). Effects of gender and expertise on consumers' motivation to read online hotel reviews. *Cornell Hospitality Quarterly*, 52(4), 399–406.

Kim, S. J., Maslowska, E., & Malthouse, E. C. (2018). Understanding the effects of different review features on purchase probability. *International Journal of Advertising*, 37(1), 29-53.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, 27(3), 313-327.

L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *IEEE Access*, 5, 7776-7797.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, *6*(70), 1.

Lau, K. N., Lee, K. H., & Ho, Y. (2005). Text mining for the hotel industry. *Cornell Hotel and Restaurant Administration Quarterly*, 46(3), 344-362.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436.

Lee, C. C., & Hu, C. (2004). Analyzing hotel customers' e-complaints from an internet complaint forum. *Journal of Travel & Tourism Marketing*, 17(2–3), 167–181.

Lee, J., Park, D. H., & 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.

Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. X. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, 63(5), 3137-3147.

Levy, S. E., Duan, W., & Boo, S. (2013). An analysis of one-star online reviews and responses in the Washington, D. C., lodging market. *Cornell Hospitality Quarterly*, 54(1), 49–63.

Li, J., & Zhan, L. (2011). Online persuasion: How the written word drives WOM: Evidence from consumer-generated product reviews. *Journal of Advertising Research*, 51(1), 239-257.

Li, X., Wong, W., Lamoureux, E. L., & Wong, T. Y. (2012). Are linear regression techniques appropriate for analysis when the dependent (outcome) variable is not normally distributed? *Investigative Ophthalmology & Visual Science*, *53*(6), 3082-3083.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468.

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

Lu, X., Ba, S., Huang, L., & Feng, Y. (2013). Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. *Information Systems Research*, 24 (3), 596–612.

Lynn, M. (1989). Scarcity effects on desirability: Mediated by assumed expensiveness? *Journal of Economic Psychology*, 10(2), 257-274.

Mariani, M. M., & Borghi, M. (2018). Effects of the Booking. com rating system: Bringing hotel class into the picture. *Tourism Management*, 66, 47-52.

Marr, B. (2015). Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance. *John Wiley & Sons*.

Marr, B. (2018). How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. *Retrieved 27 May 2020, from https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#4580335360ba*

Mauri, A. G., & Minazzi, R. (2013). Web reviews influence on expectations and purchasing intentions of hotel potential customers. *International Journal of Hospitality Management*, 34, 99–107.

Mellinas, J. P., María-Dolores, S. M. M., & García, J. J. B. (2015). Booking. com: The unexpected scoring system. *Tourism Management*, 49, 72-74.

Miao, L., Kuo, P. J., & Lee, B. Y. (2011). Consumers' responses to ambivalent online hotel reviews: The role of perceived source credibility and pre-decisional disposition. *International Journal of Hospitality Management*, 30(1), 178–183.

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

Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. In Proceedings of the 21st international conference on World Wide Web (pp. 191-200). *ACM*.

Mulpuru, S. (2007). How damaging are negative customer reviews. *Retrieved from Forrester: http://www. forester. com/Research/Document/Excerpt/0*, 7211(40649), p.00

O'Connor, P. (2010). Managing a hotel's image on TripAdvisor. *Journal of Hospitality Marketing & Management*, 19(7), 754–772.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, 375(13), 1216.

Öğüt, H., & Onur Taş, B. K. (2012). The influence of internet customer reviews on the online sales and prices in hotel industry. *The Service Industries Journal*, 32(2), 197–214.

Paas, L. J., & Morren, M. (2018). Please do not answer if you are reading this: Respondent attention in online panels. *Marketing Letters*, 29(1), 13-21.

Pantelidis, I. S. (2010). Electronic meal experience: A content analysis of online restaurant comments. *Cornell Hospitality Quarterly*, 51(4), 483–491.

Papathanassis, A., & Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach. *Tourism Management*, 32(2), 215-224.

Park, D. H., 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.

Pechmann, C. (1992). Predicting when two-sided ads will be more effective than one-sided ads: The role of correlational and correspondent inferences. *Journal of Marketing Research*, 29(4), 441-453.

Pekar, V., & Ou, S. (2008). Discovery of subjective evaluations of product features in hotel reviews. *Journal of Vacation Marketing*, 14(2), 145–155.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In Communication and persuasion (pp. 1-24). *Springer, New York, NY*.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, 41(5), 847.

Purnawirawan, N., De Pelsmacker, P., & Dens, N. (2012). Balance and sequence in online reviews: How perceived usefulness affects attitudes and intentions. *Journal of Interactive Marketing*, 26(4), 244-255.

Racherla, P., Connolly, D. J., & Christodoulidou, N. (2013). What determines consumers' ratings of service providers? An exploratory study of online traveller reviews. *Journal of Hospitality Marketing & Management*, 22(2), 135–161.

Richins, M. L. (1984). Word of mouth communication as negative information. *ACR North American Advances*.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296-320.

Sawyer, A. G. (1973). The effects of repetition of refutational and supportive advertising appeals. *Journal of Marketing Research*, 10(1), 23-33.

Schindler, R. M., & Bickart, B. (2005). Published word of mouth: Referable, consumer-generated information on the Internet. *Online Consumer Psychology: Understanding and Influencing Consumer Behavior in the Virtual World*, 32, 35-61.

Schlosser, A. E. (2011). Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments. *Journal of Consumer Psychology*, 21(3), 226-239.

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, 32(5), 608-621.

Schnack, H. G., & Kahn, R. S. (2016). Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. *Frontiers in Psychiatry,* 7, 50.

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing*, 21(4), 76-94.

Settle, R. B., & Golden, L. L. (1974). Attribution theory and advertiser credibility. *Journal of Marketing Research*, 11(2), 181-185.

Shipitsyna, E., Roos, A., Datcu, R., Hallén, A., Fredlund, H., Jensen, J. S., ... & Unemo, M. (2013). Composition of the vaginal microbiota in women of reproductive age–sensitive and specific molecular diagnosis of bacterial vaginosis is possible? *PloS One*, 8(4), e60670.

Siegel, E. (2013). Predictive analytics: The power to predict who will click, buy, lie, or die. *John Wiley & Sons*.

Simonson, I., & Rosen, E. (2014). What marketers misunderstand about online reviews. *Harvard Business Review*, 92(1), 7.

Smith, R. E., & Hunt, S. D. (1978). Attributional processes and effects in promotional situations. *Journal of Consumer Research*, 5(3), 149-158.

Sparks, B. A., & Browning, V. (2010). Complaining in cyberspace: The motives and forms of hotel guests' complaints online. *Journal of Hospitality Marketing & Management*, 19(7), 797–818.

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel Booking intentions and perception of trust. *Tourism Management*, 32(6), 1310-1323.

Sridhar, S., & Srinivasan, R. (2012). Social influence effects in online product ratings. *Journal of Marketing*, 76(5), 70-88.

Stamolampros, P., Korfiatis, N., Kourouthanassis, P., & Symitsi, E. (2019). Flying to quality: Cultural influences on online reviews. *Journal of Travel Research*, 58(3), 496-511.

Statista. (2019). Topic: Online travel market. *Retrieved 17 August 2019, from https://www.statista.com/topics/2704/online-travel-market/*

Statistics Solutions. (2019). Assumptions of Logistic Regression - Statistics Solutions. Retrieved 18 September 2019, from https://www.statisticssolutions.com/assumptions-of-logistic-regression/

Stayman, D., Hoyer, W., & Leon, R. (1987). Attribute importance in discounting product features in advertising. *In American Marketing Association Summer Educator's Conference in Toronto, Canada*.

Stringam, B. B., & Gerdes Jr, J. (2010). An analysis of word-of-mouse ratings and guest comments of online hotel distribution sites. *Journal of Hospitality Marketing & Management*, 19(7), 773-796.

Stringam, B. B., Gerdes Jr, J., & Vanleeuwen, D. M. (2010). Assessing the importance and relationships of ratings on user-generated traveller reviews. *Journal of Quality Assurance in Hospitality & Tourism*, 11(2), 73–92.

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

The Social Skinny. (2019). 100 social media statistics for 2012 | The Social Skinny. *Retrieved from http://thesocialskinny.com/100-social-media-statistics-for-2012/*

Tu, J. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11), 1225-1231.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297-323.

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism management*, *30*(1), 123-127.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. *Sage Publications*.

Wetzer, I. M., Zeelenberg, M., & Pieters, R. (2007). Never eat in that restaurant, I did!": Exploring why people engage in negative word of mouth communication. *Psychology & Marketing*, 24(8), 661–680.

Whitaker, S. D. (2018). Big data versus a survey. *The Quarterly Review of Economics and Finance*, 67, 285-296.

Willemsen, L. M., Neijens, P. C., Bronner, F., & De Ridder, J. A. (2011). "Highly recommended!" The content characteristics and perceived usefulness of online consumer reviews. *Journal of Computer-Mediated Communication*, 17(1), 19-38.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97-107.

Yacouel, N., & Fleischer, A. (2012). The role of cyberdiaries in reputation building and price premiums in the online hotel market. *Journal of Travel Research*, 51(2), 219–226.

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180-182.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveller behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, 27(2), 634–639.

Ye, Q., Li, H., Wang, Z., & Law, R. (2012). The influence of hotel price on perceived service quality and value in e-tourism: An empirical investigation based on online traveller reviews. *Journal of Hospitality & Tourism Research*, Advance online publication.

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications*, 36(3), 6527–6535.

Yoldaş, Ö., Tez, M., & Karaca, T. (2012). Artificial neural networks in the diagnosis of acute appendicitis. *The American Journal of Emergency Medicine*, 30(7), 1245-1247.

Yoo, K. H., & Gretzel, U. (2008). What motivates consumers to write online travel reviews? *Information Technology & Tourism*, 10(4), 283–295.

Yoo, K. H., & Gretzel, U. (2010). Antecedents and impacts of trust in travel-related consumer-generated media. *Information Technology & Tourism*, 12(2), 139–152.

Yoo, K. H., & Gretzel, U. (2011). Influence of personality on travel-related consumer-generated media creation. *Computers in Human Behavior*, 27(2), 609-621.

Zhang, J. J., & Mao, Z. (2012). Image of all hotel scales on travel blogs: Its impact on customer loyalty. *Journal of Hospitality Marketing & Management*, 21(2), 113–131.

Zhang, J. Q., Craciun, G., & Shin, D. (2010). When does electronic word-of-mouth matter? A study of consumer product reviews. *Journal of Business Research*, 63(12), 1336-1341.

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, 42, 146-157.

Zhang, Z., Ye, Q., & Law, R. (2011). Determinants of hotel room price: An exploration of traveller's hierarchy of accommodation needs. *International Journal of Contemporary Hospitality Management*, 23(7), 972–981.

Zhang, Z., Ye, Q., Law, R., & Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29(4), 694–700.

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, 38(6), 7674–7682.

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.

Zikopoulos, P., & Eaton, C. (2011). Understanding big data: Analytics for enterprise class hadoop and streaming data. *McGraw-Hill Osborne Media*.

Appendices

Appendix A. Results of cultural diversity based on different reviewers' nationalities: Regression analysis for Booking.com

Nationality	IV	Unstandardized Coefficients		Standardized Coefficients	t	P value
		B	Std. Error	Beta		
Algeria	(Constant)	.217	.042		5.197***	.000
	Star Ratings	-.022	.005	-.176	-4.167***	.000
F (1,541) =17.361 (p<.001), R^2=.031, Adjusted R^2=.029, Durbin-Watson=.744						
Argentina	(Constant)	.102	.006		17.512***	.000
	Star Ratings	-.005	.001	-.030	-7.329***	.000
F (1,59830) =53.719 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.350						
Armenia	(Constant)	.130	.044		2.935**	.004
	Star Ratings	-.012	.006	-.132	-2.211**	.028
F (1,277) =4.888 (p=.028), R^2=.017, Adjusted R^2=.014, Durbin-Watson=.698						
Australia	(Constant)	.072	.005		14.284***	.000
	Star Ratings	-.004	.001	-.027	-5.960***	.000
F (1,48055) =35.516 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.256						
Austria	(Constant)	.052	.009		5.927***	.000
	Star Ratings	-.002	.001	-.018	-2.139**	.032
F (1,13949) =4.577 (p=.032), R^2=.0003, Adjusted R^2=.0003, Durbin-Watson=.250						
Bangladesh	(Constant)	.098	.035		2.836**	.005
	Star Ratings	-.009	.005	-.095	-1.992**	.047
F (1,438) =3.968 (p=.047), R^2=.009, Adjusted R^2=.007, Durbin-Watson=.693						
Belarus	(Constant)	.242	.058		4.155***	.000
	Star Ratings	-.021	.007	-.106	-2.909**	.004
F (1,744) =8.463 (p=.004), R^2=.011, Adjusted R^2=.01, Durbin-Watson=.745						
Belgium	(Constant)	.098	.008		12.432***	.000
	Star Ratings	-.008	.001	-.059	-7.865***	.000
F (1,17663) =61.854 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.279						
Bolivia	(Constant)	.141	.026		5.451***	.000
	Star Ratings	-.013	.003	-.097	-3.992***	.000
F (1,1695) =15.936 (p<.001), R^2=.009, Adjusted R^2=.009, Durbin-Watson=.647						
Brazil	(Constant)	.158	.009		18.409***	.000
	Star Ratings	-.005	.001	-.016	-4.895***	.000
F (1,89466) =23.959 (p<.001), R^2=.0003, Adjusted R^2=.0003, Durbin-Watson=.682						

Burkina Faso	(Constant)	.357	.123		2.893**	.005
	Star Ratings	-.043	.017	-.319	-2.546**	.014
	F (1,57) =6.48 (p=.014), R^2=.102, Adjusted R^2=.086, Durbin-Watson=1.382					
Cameroon	(Constant)	.392	.164		2.385**	.020
	Star Ratings	-.048	.022	-.264	-2.189**	.032
	F (1,64) =4.792 (p=.032), R^2=.07, Adjusted R^2=.055, Durbin-Watson=1.935					
Canada	(Constant)	.076	.004		17.328***	.000
	Star Ratings	-.005	.001	-.036	-8.814***	.000
	F (1,61324) =77.678 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.384					
Chile	(Constant)	.107	.009		11.543***	.000
	Star Ratings	-.006	.001	-.039	-5.473***	.000
	F (1,20051) =29.955 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.362					
China	(Constant)	.202	.010		20.457***	.000
	Star Ratings	-.015	.001	-.064	-12.709***	.000
	F (1,39782) =161.509 (p<.001), R^2=.004, Adjusted R^2=.004, Durbin-Watson=.544					
Colombia	(Constant)	.097	.008		12.355***	.000
	Star Ratings	-.006	.001	-.046	-6.692***	.000
	F (1,21074) =44.785 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.418					
Costa Rica	(Constant)	.096	.020		4.708***	.000
	Star Ratings	-.005	.002	-.028	-2.116**	.034
	F (1,5769) =4.479 (p=.034), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.983					
Cyprus	(Constant)	.186	.051		3.649***	.000
	Star Ratings	-.017	.006	-.091	-2.838**	.005
	F (1,957) =8.056 (p=.005), R^2=.008, Adjusted R^2=.007, Durbin-Watson=.978					
Czech Republic	(Constant)	.058	.011		5.543***	.000
	Star Ratings	-.004	.001	-.033	-3.031**	.002
	F (1,8430) =9.185 (p=.002), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.335					
Denmark	(Constant)	.064	.008		8.003***	.000
	Star Ratings	-.005	.001	-.051	-5.274***	.000
	F (1,10631) =27.81 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.594					
Dominica	(Constant)	.797	.330		2.413**	.023
	Star Ratings	-.090	.042	-.378	-2.158**	.040
	F (1,28) = (p=.04), R^2=.143, Adjusted R^2=.112, Durbin-Watson=1.445					
Dominican Republic	(Constant)	.128	.017		7.614***	.000
	Star Ratings	-.012	.002	-.102	-5.747***	.000
	F (1,3138) =33.032 (p<.001), R^2=.01, Adjusted R^2=.01, Durbin-Watson=.439					
Ecuador	(Constant)	.098	.012		7.978***	.000
	Star Ratings	-.007	.002	-.055	-4.712***	.000
	F (1,7385) =22.207 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.421					

Egypt	(Constant)	.150	.018		8.091***	.000
	Star Ratings	-.015	.002	-.134	-6.331***	.000
	F (1,2205) =40.087 (p<.001), R^2=.018, Adjusted R^2=.017, Durbin-Watson=.524					
El Salvador	(Constant)	.115	.034		3.342***	.001
	Star Ratings	-.009	.004	-.066	-2.175**	.030
	F (1,1093) =4.729 (p=.03), R^2=.004, Adjusted R^2=.003, Durbin-Watson=.868					
France	(Constant)	.115	.004		28.278***	.000
	Star Ratings	-.008	.001	-.053	-15.581***	.000
	F (1,87392) =242.758 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.261					
Gabon	(Constant)	.284	.105		2.697**	.009
	Star Ratings	-.030	.014	-.260	-2.235**	.029
	F (1,69) =4.995 (p=.029), R^2=.068, Adjusted R^2=.054, Durbin-Watson=1.145					
Germany	(Constant)	.061	.004		17.460***	.000
	Star Ratings	-.003	.000	-.027	-7.617***	.000
	F (1,82132) =58.016 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.246					
Greece	(Constant)	.260	.025		10.267***	.000
	Star Ratings	-.025	.003	-.127	-7.832***	.000
	F (1,3755) =61.346 (p<.001), R^2=.016, Adjusted R^2=.016, Durbin-Watson=.555					
Grenada	(Constant)	.216	.077		2.792**	.006
	Star Ratings	-.023	.009	-.232	-2.491**	.014
	F (1,109) = (p=.014), R^2=.054, Adjusted R^2=.045, Durbin-Watson=.744					
Guatemala	(Constant)	.095	.018		5.216***	.000
	Star Ratings	-.007	.002	-.062	-3.287***	.001
	F (1,2827) =10.801 (p=.001), R^2=.004, Adjusted R^2=.003, Durbin-Watson=.531					
Hong Kong	(Constant)	.099	.015		6.506***	.000
	Star Ratings	-.007	.002	-.052	-3.693***	.000
	F (1,4974) =13.64 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.366					
Hungary	(Constant)	.103	.015		6.651***	.000
	Star Ratings	-.008	.002	-.051	-4.005***	.000
	F (1,6246) =16.036 (p<.001), R^2=.003, Adjusted R^2=.002, Durbin-Watson=.517					
Iceland	(Constant)	.092	.016		5.791***	.000
	Star Ratings	-.008	.002	-.077	-4.486***	.000
	F (1,3379) =20.122 (p<.001), R^2=.006, Adjusted R^2=.006, Durbin-Watson=.492					
India	(Constant)	.058	.009		6.293***	.000
	Star Ratings	-.003	.001	-.030	-2.656**	.008
	F (1,7807) =7.056 (p=.008), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.322					
Iraq	(Constant)	.139	.042		3.324***	.001
	Star Ratings	-.016	.005	-.177	-2.859**	.005
	F (1,253) =8.173 (p=.005), R^2=.031, Adjusted R^2=.027, Durbin-Watson=1.003					

Ireland	(Constant)	.098	.010		9.842***	.000
	Star Ratings	-.007	.001	-.040	-5.622***	.000
	F (1,19973) =31.606 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.844					
Israel	(Constant)	.084	.006		14.725***	.000
	Star Ratings	-.006	.001	-.052	-8.790***	.000
	F (1,27966) =77.271 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.313					
Italy	(Constant)	.112	.005		23.916***	.000
	Star Ratings	-.008	.001	-.051	-13.735***	.000
	F (1,71248) =188.654 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.359					
Japan	(Constant)	.092	.007		12.510***	.000
	Star Ratings	-.005	.001	-.034	-5.390***	.000
	F (1,24577) =29.047 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.296					
Jordan	(Constant)	.106	.030		3.580***	.000
	Star Ratings	-.009	.004	-.091	-2.458**	.014
	F (1,725) =6.041 (p=.014), R^2=.008, Adjusted R^2=.007, Durbin-Watson=.791					
Kazakhstan	(Constant)	.212	.038		5.547***	.000
	Star Ratings	-.016	.005	-.079	-3.285***	.001
	F (1,1722) =10.794 (p=.001), R^2=.006, Adjusted R^2=.006, Durbin-Watson=.537					
Kenya	(Constant)	.125	.036		3.490***	.001
	Star Ratings	-.012	.005	-.114	-2.587**	.010
	F (1,508) =6.695 (p=.01), R^2=.013, Adjusted R^2=.011, Durbin-Watson=.748					
Kuwait	(Constant)	.108	.019		5.700***	.000
	Star Ratings	-.007	.002	-.050	-2.978**	.003
	F (1,3491) =8.87 (p=.003), R^2=.003, Adjusted R^2=.002, Durbin-Watson=.568					
Latvia	(Constant)	.136	.037		3.636***	.000
	Star Ratings	-.011	.005	-.069	-2.384**	.017
	F (1,1201) =5.683 (p=.017), R^2=.005, Adjusted R^2=.004, Durbin-Watson=.67					
Macao	(Constant)	.353	.139		2.547**	.011
	Star Ratings	-.037	.018	-.128	-2.069**	.040
	F (1,258) =4.28 (p=.04), R^2=.016, Adjusted R^2=.013, Durbin-Watson=1.343					
Macedonia	(Constant)	.182	.058		3.113**	.002
	Star Ratings	-.017	.007	-.139	-2.286**	.023
	F (1,265) =5.228 (p=.023), R^2=.019, Adjusted R^2=.016, Durbin-Watson=.836					
Malaysia	(Constant)	.120	.026		4.658***	.000
	Star Ratings	-.011	.003	-.076	-3.338***	.001
	F (1,1912) =11.14 (p=.001), R^2=.006, Adjusted R^2=.005, Durbin-Watson=.493					
Martinique	(Constant)	.147	.041		3.572***	.000
	Star Ratings	-.011	.005	-.067	-2.195**	.028
	F (1,1073) =4.816 (p=.028), R^2=.004, Adjusted R^2=.004, Durbin-Watson=.557					

Mexico	(Constant)	.097	.006		16.026***	.000
	Star Ratings	-.007	.001	-.059	-9.803***	.000
	F (1,27964) =96.09 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.306					
Monaco	(Constant)	.249	.068		3.661***	.000
	Star Ratings	-.025	.009	-.171	-2.922**	.004
	F (1,282) =8.54 (p=.004), R^2=.029, Adjusted R^2=.026, Durbin-Watson=1.137					
Morocco	(Constant)	.135	.026		5.135***	.000
	Star Ratings	-.012	.003	-.106	-3.554***	.000
	F (1,1113) =12.628 (p<.001), R^2=.011, Adjusted R^2=.01, Durbin-Watson=.390					
Mozambique	(Constant)	.324	.115		2.807**	.006
	Star Ratings	-.033	.014	-.195	-2.280**	.024
	F (1,131) =5.196 (p=.024), R^2=.038, Adjusted R^2=.031, Durbin-Watson=.433					
Netherlands	(Constant)	.053	.005		10.152***	.000
	Star Ratings	-.003	.001	-.027	-4.829***	.000
	F (1,32256) =23.323 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.302					
New Zealand	(Constant)	.075	.012		6.308***	.000
	Star Ratings	-.004	.001	-.028	-2.749**	.006
	F (1,9981) =7.556 (p=.006), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.313					
Niger	(Constant)	.087	.020		4.287***	.000
	Star Ratings	-.006	.003	-.058	-2.480**	.013
	F (1,1809) =6.15 (p=.013), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.324					
Norway	(Constant)	.066	.011		6.201***	.000
	Star Ratings	-.005	.001	-.042	-3.650***	.000
	F (1,7409) =13.324 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.401					
Panama	(Constant)	.077	.016		4.729***	.000
	Star Ratings	-.005	.002	-.036	-2.371**	.018
	F (1,4269) =5.619 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.317					
Peru	(Constant)	.100	.014		7.093***	.000
	Star Ratings	-.007	.002	-.052	-4.243***	.000
	F (1,6614) =18.007 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.517					
Philippines	(Constant)	.092	.017		5.335***	.000
	Star Ratings	-.007	.002	-.055	-3.387***	.001
	F (1,3770) =11.475 (p=.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.673					
Poland	(Constant)	.103	.010		9.975***	.000
	Star Ratings	-.008	.001	-.061	-6.597***	.000
	F (1,11533) =43.526 (p<.001), R^2=.004, Adjusted R^2=.004, Durbin-Watson=.452					
Portugal	(Constant)	.141	.016		8.750***	.000
	Star Ratings	-.010	.002	-.056	-4.896***	.000
	F (1,7693) =23.969 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.379					

Puerto Rico	(Constant)	.071	.010		6.757***	.000
	Star Ratings	-.005	.001	-.045	-3.755***	.000
	F (1,6913) =14.1 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.296					
Qatar	(Constant)	.242	.061		3.989***	.000
	Star Ratings	-.023	.008	-.060	-2.925**	.003
	F (1,2336) =8.556 (p=.003), R^2=.004, Adjusted R^2=.003, Durbin-Watson=1.692					
Romania	(Constant)	.106	.018		5.889***	.000
	Star Ratings	-.008	.002	-.064	-3.704***	.000
	F (1,3339) =13.721 (p<.001), R^2=.004, Adjusted R^2=.004, Durbin-Watson=.398					
Russia	(Constant)	.347	.023		15.277***	.000
	Star Ratings	-.020	.003	-.048	-7.352***	.000
	F (1,23809) =54.045 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.85					
Saudi Arabia	(Constant)	.124	.011		11.055***	.000
	Star Ratings	-.008	.001	-.049	-5.212***	.000
	F (1,11219) =27.16 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.427					
Slovakia	(Constant)	.075	.019		3.933***	.000
	Star Ratings	-.006	.002	-.047	-2.460**	.014
	F (1,2743) =6.503 (p=.014), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.472					
South Africa	(Constant)	.075	.012		6.052***	.000
	Star Ratings	-.005	.002	-.039	-3.043**	.002
	F (1,6197) =9.262 (p=.002), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.348					
South Korea	(Constant)	.115	.009		12.714***	.000
	Star Ratings	-.008	.001	-.057	-7.144***	.000
	F (1,15519) =51.037 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.391					
Spain	(Constant)	.102	.006		17.242***	.000
	Star Ratings	-.006	.001	-.031	-7.442***	.000
	F (1,56156) =55.384 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.449					
Sri Lanka	(Constant)	.118	.041		2.889**	.004
	Star Ratings	-.011	.005	-.087	-2.071**	.039
	F (1,560) =4.29 (p=.039), R^2=.008, Adjusted R^2=.006, Durbin-Watson=.686					
Sweden	(Constant)	.041	.006		7.252***	.000
	Star Ratings	-.002	.001	-.025	-3.449***	.001
	F (1,18966) =11.892 (p=.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.253					
Switzerland	(Constant)	.074	.006		12.781***	.000
	Star Ratings	-.005	.001	-.035	-6.641***	.000
	F (1,35424) =44.101 (p<.001), R^2=.001, Adjusted R^2=.001, Durbin-Watson=.338					
Taiwan	(Constant)	.120	.015		8.260***	.000
	Star Ratings	-.009	.002	-.049	-4.874***	.000
	F (1,9893) =23.759 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.635					

Turkey	(Constant)	.083	.008		10.881***	.000
	Star Ratings	-.006	.001	-.054	-5.974***	.000
	F (1,12081) =35.691 (p<.001), R^2=.003, Adjusted R^2=.003, Durbin-Watson=.266					
Turkmenistan	(Constant)	1.471	.500		2.940**	.006
	Star Ratings	-.163	.064	-.382	-2.550**	.015
	F (1,38) =6.505 (p=.015), R^2=.146, Adjusted R^2=.124, Durbin-Watson=1.447					
Ukraine	(Constant)	.173	.033		5.197***	.000
	Star Ratings	-.011	.004	-.040	-2.576**	.010
	F (1,4231) =6.634 (p=.01), R^2=.002, Adjusted R^2=.001, Durbin-Watson=.872					
United Arab Emirates	(Constant)	.080	.011		7.104***	.000
	Star Ratings	-.004	.001	-.031	-2.830**	.005
	F (1,8379) =8.011 (p=.005), R^2=.010, Adjusted R^2=.001, Durbin-Watson=.435					
United Kingdom	(Constant)	.088	.003		27.054***	.000
	Star Ratings	-.005	.000	-.041	-14.007***	.000
	F (1,115337) =196.196 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.256					
USA	(Constant)	.068	.001		70.622***	.000
	Star Ratings	-.005	.000	-.046	-39.869***	.000
	F (1,737257) =1589.501 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.213					
Venezuela	(Constant)	.063	.008		8.233***	.000
	Star Ratings	-.005	.001	-.047	-5.070***	.000
	F (1,11448) =25.703 (p<.001), R^2=.002, Adjusted R^2=.002, Durbin-Watson=.267					

Appendix B. Results of cultural diversity based on different reviewers' nationalities: Negative binomial regression analysis for Booking.com

City	IV	*B*	S. E.	Wald	P value	*OR*	95% CI
Algeria	(Intercept)	-.526	.524	1.006	.316	.591	(.211~1.652)
	Star Ratings	-.381	.087	19.373***	.000	.683	(.577~.81)
Pearson χ2=1.173, Omnibus Test: Likelihood Ratio χ2=20.3(*p*<.001)							
Argentina	(Intercept)	-2.176	.073	886.125***	.000	.114	(.098~.131)
	Star Ratings	-.081	.009	77.54***	.000	.922	(.905~.939)
Pearson χ2=1.466, Omnibus Test: Likelihood Ratio χ2=75.098(*p*<.001)							
Armenia	(Intercept)	-1.297	.882	2.164	.141	.273	(.049~1.539)
	Star Ratings	-.298	.135	4.843**	.028	.742	(.569~.968)
Pearson χ2=1.013, Omnibus Test: Likelihood Ratio χ2=4.816(*p*=.028)							
Australia	(Intercept)	-2.526	.097	674.444***	.000	.08	(.066~.097)
	Star Ratings	-.079	.012	42.652***	.000	.924	(.903~.946)
Pearson χ2=1.219, Omnibus Test: Likelihood Ratio χ2=41.063(*p*<.001)							
Austria	(Intercept)	-2.883	.22	171.983***	.000	.056	(.036~.086)
	Star Ratings	-.065	.027	5.57**	.018	.937	(.888~.989)
Pearson χ2=1.23, Omnibus Test: Likelihood Ratio χ2=5.418(*p*=.02)							
Bangladesh	(Intercept)	-1.703	.804	4.492**	.034	.182	(.038~.88)
	Star Ratings	-.27	.128	4.439**	.035	.764	(.594~.981)
Pearson χ2=1.168, Omnibus Test: Likelihood Ratio χ2=4.445(*p*=.035)							
Belarus	(Intercept)	-.836	.511	2.683	.101	.433	(.159~1.179)
	Star Ratings	-.23	.068	11.326***	.001	.794	(.695~.908)
Pearson χ2=1.345, Omnibus Test: Likelihood Ratio χ2=10.654(*p*=.001)							
Belgium	(Intercept)	-1.907	.16	141.622***	.000	.149	(.108~.203)
	Star Ratings	-.184	.022	73.251***	.000	.832	(.798~.868)
Pearson χ2=1.23, Omnibus Test: Likelihood Ratio χ2=69.294(*p*<.001)							
Bolivia	(Intercept)	-1.299	.403	10.397***	.001	.273	(.124~.601)
	Star Ratings	-.266	.058	21.291***	.000	.767	(.685~.858)
Pearson χ2=1.285, Omnibus Test: Likelihood Ratio χ2=20.094(*p*<.001)							
Brazil	(Intercept)	-1.814	.045	1590.374***	.000	.163	(.149~.178)
	Star Ratings	-.042	.006	55.622***	.000	.958	(.948~.969)
Pearson χ2=2.352, Omnibus Test: Likelihood Ratio χ2=54.713(*p*<.001)							
Burkina Faso	(Intercept)	1.967	2.27	.751	.386	7.153	(.084~612.232)
	Star Ratings	-.826	.429	3.702*	.054	.438	(.189~1.016)
Pearson χ2=0.665, Omnibus Test: Likelihood Ratio χ2=4.982(*p*=.026)							
Cameroon	(Intercept)	3.331	2.782	1.433	.231	27.967	(.12~6531.966)

	Star Ratings	-1.473	.901	2.675	.102	.229	(.039~1.34)
Pearson χ2=0.226, Omnibus Test: Likelihood Ratio χ2=11.785(*p*=.001)							
Canada	(Intercept)	-2.395	.084	820.716***	.000	.091	(.077~.107)
	Star Ratings	-.112	.011	110.514***	.000	.894	(.875~.913)
Pearson χ2=1.413, Omnibus Test: Likelihood Ratio χ2=105.316(*p*<.001)							
Chile	(Intercept)	-2.083	.121	294.336***	.000	.125	(.098~.158)
	Star Ratings	-.099	.015	41.186***	.000	.905	(.878~.933)
Pearson χ2=1.416, Omnibus Test: Likelihood Ratio χ2=39.64(*p*<.001)							
China	(Intercept)	-1.252	.072	299.888***	.000	.286	(.248~.329)
	Star Ratings	-.166	.009	315.306***	.000	.847	(.832~.863)
Pearson χ2=2.008, Omnibus Test: Likelihood Ratio χ2=298.6(*p*<.001)							
Colombia	(Intercept)	-2.111	.125	287.34***	.000	.121	(.095~.155)
	Star Ratings	-.127	.016	62.129***	.000	.881	(.853~.909)
Pearson χ2=1.409, Omnibus Test: Likelihood Ratio χ2=59.107(*p*<.001)							
Costa Rica	(Intercept)	-2.214	.23	92.504***	.000	.109	(.07~.172)
	Star Ratings	-.091	.029	9.767**	.002	.913	(.862~.967)
Pearson χ2=2.333, Omnibus Test: Likelihood Ratio χ2=9.405(*p*=.002)							
Cyprus	(Intercept)	-.775	.535	2.1	.147	.461	(.161~1.314)
	Star Ratings	-.31	.074	17.687***	.000	.733	(.635~.847)
Pearson χ2=1.712, Omnibus Test: Likelihood Ratio χ2=16.773(*p*<.001)							
Czech Republic	(Intercept)	-2.596	.285	83.013***	.000	.075	(.043~.13)
	Star Ratings	-.13	.036	12.663***	.000	.878	(.818~.943)
Pearson χ2=1.442, Omnibus Test: Likelihood Ratio χ2=12.015(*p*=.001)							
Denmark	(Intercept)	-2.273	.24	89.564***	.000	.103	(.064~.165)
	Star Ratings	-.208	.033	38.591***	.000	.812	(.761~.867)
Pearson χ2=1.612, Omnibus Test: Likelihood Ratio χ2=36.546(*p*<.001)							
Dominica	(Intercept)	1.013	6.987	2.054	.152	22317.664	(.025~19774836973.419)
	Star Ratings	-2.05	1.293	2.512	.113	.129	(.01~1.624)
Pearson χ2=0.232, Omnibus Test: Likelihood Ratio χ2=7.916(*p*=.005)							
Dominican Republic	(Intercept)	-1.313	.315	17.332***	.000	.269	(.145~.499)
	Star Ratings	-.28	.044	40.297***	.000	.756	(.693~.824)
Pearson χ2=1.39, Omnibus Test: Likelihood Ratio χ2=37.844(*p*<.001)							
Ecuador	(Intercept)	-2.023	.21	92.887***	.000	.132	(.088~.2)
	Star Ratings	-.15	.027	30.172***	.000	.861	(.816~.908)
Pearson χ2=1.338, Omnibus Test: Likelihood Ratio χ2=28.479(*p*<.001)							
Egypt	(Intercept)	-.969	.319	9.21**	.002	.379	(.203~.709)
	Star Ratings	-.353	.051	47.565***	.000	.703	(.636~.777)
Pearson χ2=1.154, Omnibus Test: Likelihood Ratio χ2=48.286(p<.001)							
El Salvador	(Intercept)	-1.71	.564	9.195**	.002	.181	(.06~.546)

	Star Ratings	-.193	.075	6.597**	.01	.825	(.712~.955)
Pearson χ2=1.601, Omnibus Test: Likelihood Ratio χ2=6.259(p=.012)							
France	(Intercept)	-1.916	.058	1088.628***	.000	.147	(.131~.165)
	Star Ratings	-.137	.008	311.786***	.000	.872	(.859~.885)
Pearson χ2=1.323, Omnibus Test: Likelihood Ratio χ2=299.474(p<.001)							
Gabon	(Intercept)	.092	1.493	.004	.951	1.096	(.059~2.454)
	Star Ratings	-.456	.249	3.346*	.067	.634	(.389~1.033)
Pearson χ2=0.833, Omnibus Test: Likelihood Ratio χ2=3.655(p=.056)							
Germany	(Intercept)	-2.662	.082	1046.777***	.000	.07	(.059~.082)
	Star Ratings	-.09	.011	70.405***	.000	.914	(.895~.934)
Pearson χ2=1.24, Omnibus Test: Likelihood Ratio χ2=68.147(p<.001)							
Greece	(Intercept)	-.512	.215	5.702**	.017	.599	(.393~.912)
	Star Ratings	-.306	.031	98.076***	.000	.736	(.693~.782)
Pearson χ2=1.635, Omnibus Test: Likelihood Ratio χ2=95.603(p<.001)							
Grenada	(Intercept)	.619	1.855	.111	.739	1.858	(.049~7.467)
	Star Ratings	-.592	.288	4.238**	.04	.553	(.315~.972)
Pearson χ2=0.702, Omnibus Test: Likelihood Ratio χ2=4.174(p=.041)							
Guatemala	(Intercept)	-1.968	.357	30.423***	.000	.14	(.069~.281)
	Star Ratings	-.179	.048	13.87***	.000	.836	(.761~.919)
Pearson χ2=1.382, Omnibus Test: Likelihood Ratio χ2=13.123(p<.001)							
Hong Kong	(Intercept)	-2.042	.258	62.59***	.000	.13	(.078~.215)
	Star Ratings	-.146	.035	17.374***	.000	.864	(.806~.925)
Pearson χ2=1.366, Omnibus Test: Likelihood Ratio χ2=16.671(p<.001)							
Hungary	(Intercept)	-1.95	.255	58.279***	.000	.142	(.086~.235)
	Star Ratings	-.155	.033	22.553***	.000	.856	(.803~.913)
Pearson χ2=1.477, Omnibus Test: Likelihood Ratio χ2=21.188(p<.001)							
Iceland	(Intercept)	-1.475	.425	12.018***	.001	.229	(.099~.527)
	Star Ratings	-.303	.057	28.157***	.000	.739	(.661~.826)
Pearson χ2=1.496, Omnibus Test: Likelihood Ratio χ2=25.689(p<.001)							
India	(Intercept)	-2.739	.215	162.543***	.000	.065	(.042~.098)
	Star Ratings	-.087	.029	9.001**	.003	.917	(.866~.97)
Pearson χ2=1.301, Omnibus Test: Likelihood Ratio χ2=8.756(p=.003)							
Iraq	(Intercept)	-.173	1.137	.023	.879	.841	(.091~7.813)
	Star Ratings	-.587	.216	7.394**	.007	.556	(.364~.849)
Pearson χ2=0.855, Omnibus Test: Likelihood Ratio χ2=8.89(p=.003)							
Ireland	(Intercept)	-2.061	.139	220.905***	.000	.127	(.097~.167)
	Star Ratings	-.135	.017	60.716***	.000	.873	(.844~.904)
Pearson χ2=1.882, Omnibus Test: Likelihood Ratio χ2=56.947(p<.001)							
Israel	(Intercept)	-2.182	.119	338.34***	.000	.113	(.089~.142)

	Star Ratings	-.151	.015	97.552***	.000	.86	(.834~.886)
Pearson χ2=1.254, Omnibus Test: Likelihood Ratio χ2=91.503(*p*<.001)							
Italy	(Intercept)	-1.916	.066	844.762***	.000	.147	(.129~.167)
	Star Ratings	-.148	.009	277.894***	.000	.863	(.848~.878)
Pearson χ2=1.558, Omnibus Test: Likelihood Ratio χ2=265.565(*p*<.001)							
Japan	(Intercept)	-2.271	.111	416.445***	.000	.103	(.083~.128)
	Star Ratings	-.089	.015	36.501***	.000	.915	(.888~.941)
Pearson χ2=1.28, Omnibus Test: Likelihood Ratio χ2=35.502(*p*<.001)							
Jordan	(Intercept)	-1.74	.594	8.584**	.003	.176	(.055~.562)
	Star Ratings	-.229	.087	6.913**	.009	.796	(.671~.943)
Pearson χ2=1.171, Omnibus Test: Likelihood Ratio χ2=6.669(*p*=0.01)							
Kazakhstan	(Intercept)	-1.232	.307	16.138***	.000	.292	(.16~.532)
	Star Ratings	-.158	.041	14.842***	.000	.854	(.788~.925)
Pearson χ2=1.35, Omnibus Test: Likelihood Ratio χ2=14.282(*p*<.001)							
Kenya	(Intercept)	-1.438	.707	4.139**	.042	.237	(.059~.949)
	Star Ratings	-.271	.104	6.833**	.009	.762	(.622~.934)
Pearson χ2=1.023, Omnibus Test: Likelihood Ratio χ2=6.484(*p*=.011)							
Kuwait	(Intercept)	-2.042	.251	66.363***	.000	.13	(.079~.212)
	Star Ratings	-.12	.034	12.761***	.000	.887	(.83~.947)
Pearson χ2=1.512, Omnibus Test: Likelihood Ratio χ2=12.295(*p*<.001)							
Latvia	(Intercept)	-1.551	.505	9.434**	.002	.212	(.079~.57)
	Star Ratings	-.195	.068	8.286**	.004	.823	(.72~.94)
Pearson χ2=1.61, Omnibus Test: Likelihood Ratio χ2=7.823(*p*=.005)							
Macao	(Intercept)	-.057	.779	.005	.942	.945	(.205~4.349)
	Star Ratings	-.367	.116	10.004**	.002	.693	(.552~.87)
Pearson χ2=1.55, Omnibus Test: Likelihood Ratio χ2=9.745(*p*=.002)							
Macedonia	(Intercept)	-.997	.858	1.349	.246	.369	(.069~1.985)
	Star Ratings	-.275	.124	4.886**	.027	.759	(.595~.969)
Pearson χ2=1.048, Omnibus Test: Likelihood Ratio χ2=4.742(*p*=.029)							
Malaysia	(Intercept)	-1.452	.456	10.128***	.001	.234	(.096~.573)
	Star Ratings	-.256	.064	16.214***	.000	.774	(.683~.877)
Pearson χ2=1.436, Omnibus Test: Likelihood Ratio χ2=15.106(*p*<.001)							
Martinique	(Intercept)	-1.529	.523	8.535**	.003	.217	(.078~.605)
	Star Ratings	-.175	.07	6.217**	.013	.84	(.732~.963)
Pearson χ2=1.26, Omnibus Test: Likelihood Ratio χ2=5.953(p=.015)							
Mexico	(Intercept)	-1.994	.114	305.643***	.000	.136	(.109~.17)
	Star Ratings	-.161	.015	118.688***	.000	.852	(.827~.877)
Pearson χ2=1.271, Omnibus Test: Likelihood Ratio χ2=111.494(*p*<.001)							
Monaco	(Intercept)	-.298	.782	.146	.703	.742	(.16~3.436)

	Star Ratings	-.378	.121	9.774**	.002	.685	(.54~.868)
Pearson χ2=1.152, Omnibus Test: Likelihood Ratio χ2=9.851(*p*=.002)							
Morocco	(Intercept)	-1.447	.454	10.167***	.001	.235	(.097~.573)
	Star Ratings	-.241	.067	12.852***	.000	.786	(.689~.897)
Pearson χ2=1.127, Omnibus Test: Likelihood Ratio χ2=12.506(*p*<.001)							
Mozambique	(Intercept)	.536	1.284	.174	.676	1.709	(.138~21.163)
	Star Ratings	-.46	.194	5.608**	.018	.631	(.431~.924)
Pearson χ2=1.011, Omnibus Test: Likelihood Ratio χ2=6.143(*p*=.013)							
Netherlands	(Intercept)	-2.764	.148	346.981***	.000	.063	(.047~.084)
	Star Ratings	-.105	.019	29.154***	.000	.9	(.867~.935)
Pearson χ2=1.278, Omnibus Test: Likelihood Ratio χ2=28.118(*p*<.001)							
New Zealand	(Intercept)	-2.466	.223	121.885***	.000	.085	(.055~.132)
	Star Ratings	-.085	.028	9.49**	.002	.918	(.869~.969)
Pearson χ2=1.275, Omnibus Test: Likelihood Ratio χ2=9.15(*p*=.002)							
Niger	(Intercept)	-2.153	.436	24.402***	.000	.116	(.049~.273)
	Star Ratings	-.151	.059	6.56**	.01	.86	(.766~.965)
Pearson χ2=1.085, Omnibus Test: Likelihood Ratio χ2=6.246(*p*=.012)							
Norway	(Intercept)	-2.402	.281	73.014***	.000	.091	(.052~.157)
	Star Ratings	-.156	.037	17.484***	.000	.856	(.795~.921)
Pearson χ2=1.314, Omnibus Test: Likelihood Ratio χ2=16.605(*p*<.001)							
Panama	(Intercept)	-2.384	.328	52.708***	.000	.092	(.048~.175)
	Star Ratings	-.109	.041	6.925**	.009	.897	(.827~.973)
Pearson χ2=1.249, Omnibus Test: Likelihood Ratio χ2=6.617(*p*=.01)							
Peru	(Intercept)	-1.987	.234	72.369***	.000	.137	(.087~.217)
	Star Ratings	-.158	.031	26.109***	.000	.854	(.804~.907)
Pearson χ2=1.552, Omnibus Test: Likelihood Ratio χ2=24.687(*p*<.001)							
Philippines	(Intercept)	-1.994	.328	36.946***	.000	.136	(.072~.259)
	Star Ratings	-.181	.044	16.978***	.000	.834	(.765~.909)
Pearson χ2=1.421, Omnibus Test: Likelihood Ratio χ2=15.99(*p*<.001)							
Poland	(Intercept)	-1.797	.188	90.876***	.000	.166	(.115~.24)
	Star Ratings	-.202	.025	63.547***	.000	.817	(.778~.859)
Pearson χ2=1.458, Omnibus Test: Likelihood Ratio χ2=59.842(p<.001)							
Portugal	(Intercept)	-1.689	.186	82.469***	.000	.185	(.128~.266)
	Star Ratings	-.142	.025	32.797***	.000	.867	(.826~.911)
Pearson χ2=1.433, Omnibus Test: Likelihood Ratio χ2=31.529(*p*<.001)							
Puerto Rico	(Intercept)	-2.433	.246	97.629***	.000	.088	(.054~.142)
	Star Ratings	-.124	.031	16.44***	.000	.883	(.832~.938)
Pearson χ2=1.175, Omnibus Test: Likelihood Ratio χ2=15.321(*p*<.001)							
Qatar	(Intercept)	-.777	.244	10.115***	.001	.46	(.285~.742)

	Star Ratings	-.269	.036	56.122***	.000	.764	(.712~.82)
Pearson χ2=3.483, Omnibus Test: Likelihood Ratio χ2=55.23(*p*<.001)							
Romania	(Intercept)	-1.895	.316	35.993***	.000	.15	(.081~.279)
	Star Ratings	-.171	.042	16.812***	.000	.843	(.777~.915)
Pearson χ2=1.317, Omnibus Test: Likelihood Ratio χ2=15.779(*p*<.001)							
Russia	(Intercept)	-.901	.067	179.873***	.000	.406	(.356~.463)
	Star Ratings	-.101	.008	142.834***	.000	.903	(.889~.919)
Pearson χ2=2.742, Omnibus Test: Likelihood Ratio χ2=139.063(*p*<.001)							
Saudi Arabia	(Intercept)	-1.943	.123	250.609***	.000	.143	(.113~.182)
	Star Ratings	-.104	.017	38.485***	.000	.901	(.872~.931)
Pearson χ2=1.453, Omnibus Test: Likelihood Ratio χ2=37.525(*p*<.001)							
Slovakia	(Intercept)	-2.202	.447	24.242***	.000	.111	(.046~.266)
	Star Ratings	-.173	.058	8.81**	.003	.841	(.75~.943)
Pearson χ2=1.552, Omnibus Test: Likelihood Ratio χ2=8.268(*p*=.004)							
South Africa	(Intercept)	-2.409	.258	87.108***	.000	.09	(.054~.149)
	Star Ratings	-.112	.033	11.288***	.001	.894	(.837~.954)
Pearson χ2=1.257, Omnibus Test: Likelihood Ratio χ2=10.835(p=0.001)							
South Korea	(Intercept)	-1.909	.123	241.251***	.000	.148	(.117~.189)
	Star Ratings	-.142	.017	72.165***	.000	.868	(.84~.897)
Pearson χ2=1.405, Omnibus Test: Likelihood Ratio χ2=69.453(*p*<.001)							
Spain	(Intercept)	-2.161	.07	954.193***	.000	.115	(.1~.132)
	Star Ratings	-.09	.009	93.183***	.000	.914	(.897~.931)
Pearson χ2=1.725, Omnibus Test: Likelihood Ratio χ2=90.687(*p*<.001)							
Sri Lanka	(Intercept)	-1.355	.866	2.448	.118	.258	(.047~1.408)
	Star Ratings	-.273	.122	4.999**	.025	.761	(.599~.967)
Pearson χ2=1.217, Omnibus Test: Likelihood Ratio χ2=4.825(*p*=.028)							
Sweden	(Intercept)	-3.035	.212	204.546***	.000	.048	(.032~.073)
	Star Ratings	-.104	.028	14.148***	.000	.901	(.854~.951)
Pearson χ2=1.209, Omnibus Test: Likelihood Ratio χ2=13.635(*p*<.001)							
Switzerland	(Intercept)	-2.386	.122	380.072***	.000	.092	(.072~.117)
	Star Ratings	-.122	.016	58.019***	.000	.885	(.858~.914)
Pearson χ2=1.348, Omnibus Test: Likelihood Ratio χ2=55.65(*p*<.001)							
Taiwan	(Intercept)	-1.806	.182	98.299***	.000	.164	(.115~.235)
	Star Ratings	-.154	.024	41.271***	.000	.857	(.818~.898)
Pearson χ2=1.805, Omnibus Test: Likelihood Ratio χ2=39.262(*p*<.001)							
Turkey	(Intercept)	-2.252	.154	213.282***	.000	.105	(.078~.142)
	Star Ratings	-.14	.022	41.711***	.000	.869	(.833~.907)
Pearson χ2=1.239, Omnibus Test: Likelihood Ratio χ2=40.368(*p*<.001)							
Turkmenistan	(Intercept)	1.696	.935	3.291*	.07	5.451	(.873~34.058)

	Star Ratings	-.531	.17	9.763**	.002	.588	(.421~.82)
Pearson χ2=0.894, Omnibus Test: Likelihood Ratio χ2=13.156(*p*<.001)						
Ukraine	(Intercept)	-1.593	.212	56.384***	.000	.203	(.134~.308)
	Star Ratings	-.106	.027	15.281***	.000	.899	(.853~.949)
Pearson χ2=2.456, Omnibus Test: Likelihood Ratio χ2=14.691(*p*<.001)						
United Arab Emirates	(Intercept)	-2.432	.181	181.465***	.000	.088	(.062~.125)
	Star Ratings	-.079	.024	10.78***	.001	.924	(.881~.969)
Pearson χ2=1.4, Omnibus Test: Likelihood Ratio χ2=10.508(*p*=.001)						
United Kingdom	(Intercept)	-2.24	.058	148.915***	.000	.106	(.095~.119)
	Star Ratings	-.114	.007	241.671***	.000	.892	(.879~.905)
Pearson χ2=1.257, Omnibus Test: Likelihood Ratio χ2=229.662(*p*<.001)						
USA	(Intercept)	-2.464	.023	11616.961***	.000	.085	(.081~.089)
	Star Ratings	-.132	.003	202.256***	.000	.876	(.871~.881)
Pearson χ2=1.303, Omnibus Test: Likelihood Ratio χ2=1898.462(*p*<.001)						
Venezuela	(Intercept)	-2.441	.226	116.814***	.000	.087	(.056~.136)
	Star Ratings	-.159	.029	3.45***	.000	.853	(.806~.903)
Pearson χ2=1.209, Omnibus Test: Likelihood Ratio χ2=28.154(*p*<.001)						