The economics of user-generated content : impacts on consumer decision-making and consumption

Zhang, Yimiao

2019

Zhang, Y. (2019). The economics of user-generated content : impacts on consumer decision-making and consumption. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/86055

https://doi.org/10.32657/10220/48376

THE ECONOMICS OF USER-GENERATED CONTENT: IMPACTS ON CONSUMER DECISION-MAKING AND CONSUMPTION

ZHANG YIMIAO

NANYANG BUSINESS SCHOOL

2019

THE ECONOMICS OF USER-GENERATED CONTENT: IMPACTS ON CONSUMER DECISION-MAKING AND CONSUMPTION

ZHANG YIMIAO

NANYANG BUSINESS SCHOOL

A thesis submitted to the Nanyang Technological University

in partial fulfilment of the requirement for the degree of

Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

. . . May 14th 2019. . .   . .

Date   Zhang Yimiao

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

. . . May 14th 2019.. . .
Date

. . . .
Assoc Prof Goh Kim Huat

Authorship Attribution Statement

Please select one of the following; *delete as appropriate:

*(B) This thesis contains material from 1 paper accepted at conferences in which I am listed as an author.

Chapter 1 is accepted and presented as Zhang Yimiao, and Goh Kim Huat, "Attracting Versus Sustaining Attention in the Information Economy." *Workshop on E-Business*, Seoul, South Korea, December 10, 2017

The contributions of the co-authors are as follows:

- Assoc Prof Goh Kim Huat provided the initial project direction and revised the manuscript drafts.
- I prepared the manuscript drafts.
- I co-designed the study with A/Prof Goh Kim Huat and performed all the data collection, processing, and analysis work.

. . . May 14th 2019 . . . .

Date Zhang Yimiao

Acknowledgements

Finally, it is time to write acknowledgements. Pursuing PhD is a long journey and I really thank for those people who support me, help me, and accompany with me all along the journey.

First of all, I would like to express my greatest gratitude to my supervisor, Dr. Goh Kim Huat, for his patient, professional, and helpful guidance on my research and writing of this thesis. Thanks for the useful suggestions to my study and life in Singapore. Thanks for the understanding and support during the time period that I have to dedicate to my family.

Besides my supervisor, heartfelt thanks go to the rest of my thesis committee: Dr. Boh Wai Fong, Dr. Mo Jiahui, and Dr. Ng Sok Ling, Sharon, for their insightful and extensive comments and encouragement in the development of this thesis.

I am profoundly grateful to the faculty in Nanyang Business School, especially faculty in the Division of Information Technology & Operations Management, for their high quality teaching that broadened and strengthened my methodology and discipline knowledge.

In addition, I would like to thank the staff of Nanyang Business School, for their supportive work in regard to administration issues. My special thank gives to Mr. Ma Kai, for his help with data collection, data cleaning, and other responsive technical support.

My sincere thanks also goes to my friends in Nanyang Business School. It’s so good to study, play, and grow together with all of you. Especially, I would like to thank Mr. Chen Yi, Miss. Zhang Nila, and Mr. Lin Yan, for their insightful comments and suggestions to my research and thesis. Besides, I would like to thank my friends in China and other parts of the world, for their support and encouragement during my PhD life.

Last but not the least, it is my privilege to thank my family: my parents, for their care and support throughout my life; my husband Mr. He Ruoyun, for the constant encouragement and support in life and technical help in the data collection process of my thesis; my lovely daughter, He Yiru Elina, for the inspirations she brings to me; my parents-in-law, for their kind support on taking care of my daughter.

Table of Contents

SUMMARY

Nowadays, many of the online content is user-generated content. User-generated content is any type of content created and published by unpaid contributors (end users) mainly on social media platforms or on e-commerce platforms and is read, viewed, or consumed by other end users of the platforms. User-generated content may be in various formats, for example images, videos, blog posts, and online reviews. Users create user-generated content for many purposes, for instance, sharing daily life, making comment towards certain event or product/brand, and sharing some useful information to others.

The importance of user-generated content comes from the economic value it creates by attracting users' attention to consume the content and further shaping users' believe to make decisions. For example, sharing useful information in a video or in a blog post may attract a lot of audience and readers such that the traffic to the content may be converted to money in various ways such as advertisement income and readers' direct reward. Endorsing or critiquing a product in a blog-post or in an online review may affect other consumers' purchase decision or change of decision. Therefore, it is essential for us to understand the drivers and outcomes of consumption of user-generated content under different situations.

In this thesis, we conducted two studies to examine the drivers as well as outcomes of user-generated content consumption in different empirical settings. Study 1 examined the drivers of continued consumption of user-generated

content (online videos) and study 2 examined the outcomes of consumption of user-generated content (online reviews) in regard to consumer attitude change.

In study 1 (chapter 1), we examine the impact of user-generated content features on the continued consumption of the content from a dynamic perspective. Specifically, we use YouTube as the empirical setting to examine what are the factors that sustain attention of individuals in this information economy. We situate our study in an empirical context and hope to unpack the psychological underpinnings which lead to greater focus and attention. We apply theory of optimal stimulation level (OSL) to explore why users stay on with particular user-generated content. The results reveal that consistency in conversation pace and variety in content topic are positively related to sustained attention.

In study 2 (chapter 2), we refer to theory in psychology and marketing to explore the impact of consumer online review on consumer post-purchase attitude change. Using a proprietary dataset from hotel industry and public user-generated online reviews, we examine how the numeric and textual information of post-purchase review affect hotel room cancellation behavior. The results show negativity bias and different trust levels on hedonic and utilitarian comment when consumers make post-purchase decision.

This thesis contributes to user-generated content literature from two perspectives. Firstly, we point out content features that affect users' continued consumption of user-generated content from a dynamic perspective in the social media context. Secondly, we explore the impact of consumption of user-generated online review on consumer post-purchase attitude change. We believe this thesis also has important practical implications for the design of user-generated content and for the management of user-generated online review.

CHAPTER 1: Sustaining Attention for User-Generated Content in the Information Economy

1. Introduction

The exponential growth of the information economy brought about immense competition within the attention economy. Attention is defined as "focused mental engagement on a particular item of information" (Davenport and Beck 2001). In the attention economy theory, attention is a resource possessed by a person and this resource is constantly facing competition and scarcity (Falkinger 2008; Goldhaber 1997; Pedrycz and Chen 2013). This argument is premised on two tenets. First, attention is a resource. Specifically, in the information age, the wealth of information consumes the attention of its recipients (Simon 1971) to the extent whereby attention is touted by some as the new currency in the information age (Davenport and Beck 2001). Consumers are spoilt by choices in information content and content that are able to attract and maintain the attention of individuals will be able to use this window of opportunity to shape their beliefs. Naturally, the ability to shape beliefs on issues ranging from what brand of soap to buy to which politician to vote for translate into value which the content provider or carrier can monetize (Goldhaber 1997). Second, attention is scarce. The attention possessed by any person is limited (Davenport and Beck 2001; Falkinger 2008). It is impossible for people to "pay attention" to every item, or every piece of information in the world. As a result, information competes for individual's attention (Falkinger 2008). For example, when you load your

Facebook page, you are not able to check out every post, every advertisement, or every notice which are competing for your attention. In sum, getting attention is important to whose business relies on it, but people's attention is limited.

With the features of the attention economy in mind, there are two notable phenomena. First, the amount of available information is rapidly increasing. In the information age, people are able to explore vast amount of information online. There are portal websites that publish news as well as different kinds of useful information on a plethora of topics such as education, health, technology, and entertainment. In addition, the last decade witnesses the rise of social media, from the likes of knowledge-driven Wikipedia to the likes of relationship-driven Facebook. A report in March 2018 shows that there are 4.021 billion social media users globally, increased by 13% since January 2017 (Chaffey 2018). Incredible amount of information is generated in social media every minute. For example, 7599 Tweets are send and 779 Instagram photos are uploaded in one second (InternetLiveStats 2017).

Second, people's attention span is short and getting decreasing. A study conducted by National Center for Biotechnology Information, U.S. National Library of Medicine in July 2016 shows that people's attention span decreased from 12 seconds in 2000 to 8.25 seconds in 2015. Brown (2000) argued that exposure to the digital world diminishes our attention span in work, entertainment and learning. Other studies examining people's online browsing behavior also arrived at similar conclusions. Liu (2005) surveyed peoples' reading behavior in the digital environment and found that online users spent less time on in-depth reading and concentrated reading, compared to the time spent ten years ago. A subsequent empirical Web usage study revealed further statistics of user's internet viewing and reading behavior (Weinreich et al. 2008). About 50% of webpage views within the empirical sample lasted less than 12 seconds, with an average number of 430 words on these webpages, and only 4% of the views lasted more than 10 minutes.

Juxtaposing the tenets of the attention economy with the phenomena of declining attention within a climate of growing information, begs the following question. What attracts and more importantly what sustains individual's attention in the information economy? Proponents of the attention economy will suggest that this will be an uphill task. Given that attention is scarce and declining, attracting and sustaining attention will become increasingly more difficult with more avenues of competition from different information sources. Plenty of research efforts have been made to examine the factors that attract users' attention in the online context (Bakshy et al. 2012; Brodersen et al. 2012; Cha et al. 2010; Cheng et al. 2008; De Vries et al. 2012; Figueiredo et al. 2011; Susarla et al. 2012; Szabo and Huberman 2010; Zhou et al. 2010). However, limited attention in literature has been paid to sustaining attention. Current literature examined the effect of static elements or features of online content such as position of an item, text content features (valence, color, and topic), and style of content (pictures, links, and status) on the sustained attention (Daugherty and Hoffman 2014; Hoffman and Daugherty 2013; Menon et al. 2016; Mou and Shin 2018; Muñoz-Leiva et al.

2018; Navalpakkam et al. 2012; Xu and Zhang 2018), which is operationalized with fixation duration using eye-tracking method. However, much of the online content is displayed in a dynamic way such as online videos or audios. In literature, there is lack of attention paid to the dynamic perspectives of online content. As the actual online content is more fruitful than what have been examined in literature from a static perspective, it is important to go beyond the static features and look into the dynamics of online content. In this study, we take a dynamic perspective. We explore how the dynamics of content topic and content delivery impact the sustained attention.

User-generated content is any type of content created and published by unpaid users of a system. Nowadays, in the online context, we see plenty of user-generated content (UGC), especially on social media platforms. There are many formats of UGC, such as blog posts, tweets, videos, and pictures. The one common component in UGC is the linguistic content that appears in blog posts, tweets, audios and videos. In this study, we use video as research context to study how the delivery of linguistic content matters to sustaining attention. We focus on informational and conversation based videos where the linguistic content would be good representation of video content. In particular, we take exploratory steps and use YouTube as the empirical setting to examine what are the factors that sustain attention of individuals in this information economy. We situate our study in an empirical context and hope to unpack the psychological underpinnings which lead to greater focus and attention. We apply theory of optimal stimulation level (OSL) to explore why users stay on with particular social media content. The results reveal that it is not how much linguistic content is delivered (in fact more is less), but how it is delivered. Consistency in content delivery pace and variety in content topic are found to be strong factors sustaining the viewers' attention.

We believe the results of this study will benefit both theory and practice in various ways. From an academic standpoint, it provides greater empirical clarity among the works in the attention economy. We contribute to literature by adding knowledge to sustaining attention when the content is dynamically displayed. Specifically, we show that consistency in informational content delivery pace and variety in content topic are positively related to continued interests and attention. From a practice standpoint, the ability to understand what makes information content tick will allow marketers to better tailor their content to ensure greatest informational reach. We will review the literature on attention in social media in section 2 and set up the hypotheses in section 3. In section 4, we outline the research context and the methodology employed followed up by the results in section 5. Section 6 and 7 discuss the contributions and limitations, respectively. Section 8 concludes.

2. Literature Review

Attention in Social Media

Prior literature has shown that individuals pay attention to a piece of online information in a two-stage process (Figueiredo et al. 2011; Liu 2005; Szabo and Huberman 2010). In the first stage, a user is attracted by a piece of information because of some reason. For instance, highlighted words, prominent position, or an attractive picture. Once a user is attracted, in the second stage, he or she spends some time with this piece of information by reading the content (sometimes watching a video). Therefore, online content creators are facing the challenges of 1) attracting user attention and 2) sustaining user attention. For different types of online content creators, some may pursue click volume (such as static online ads) while others may pursue users' attention span (e.g. minimum elapsed time for online video advertisement). Thus, it is critical to distinguish the different factors that drive attention attraction and attention sustaining. However, current literature only emphasized the importance to study the mechanism of attention attraction, but very limited attention is paid attention sustaining.

Attention is a scarce resource under attention economy theory as individuals possess limited capabilities in maintaining attention. Psychology literature suggests that people have limited attention because of cognitive bottlenecks when processing incoming stimuli (Broadbent 1958; Deutsch and Deutsch 1963; Norman 1968; Treisman 1960; Treisman 1964). The investigation of human attention dates back to the 1960s and bottleneck models (e.g. Broadbent's Model, Treisman's Model, and Deutsch & Deutsch Model) are classic theories in the research on attracting attention. Collectively, they suggest that there will always exist is a bottleneck in the information processing process and human's limited capacity to pay attention is because of the existence of this cognitive bottleneck. The human brain is a "system" to process the stimuli received from the environment and although a lot of stimuli may enter the "system" simultaneously, not all of them can go through the bottleneck of the “system” effectively for processing (i.e. not all the stimuli can be processed and be paid attention to). For example, when an individual looks at an array of video links on YouTube, visually, he is seeing all images and links, but cognitively, one can only process only a few of them due to this processing bottleneck.

Many factors help online content go through the cognitive bottleneck. Existing literature shows both intrinsic and extrinsic features of online content drive users’ attention. The intrinsic features of online content, such as position on a website, content quality, and publisher location, have significant impact on attracting attention. Specifically, a brand post on the top of the page, and vivid and interactive brand post are more likely to attract attention (De Vries et al. 2012). Prior studies have also tried to classify YouTube videos into various archetypes such as viral video, quality video, and junk video according to the videos’ early view pattern in order to predict future viewership (Figueiredo et al. 2011; Pinto et al. 2013). Extant literature also points out the effect of geographic locality on YouTube video view count. About 50% of videos receive more than 70% of their total view count from a single country, however, social sharing widens a video’s geographic audience (Brodersen et al. 2012). The effect of online content position is also reflected on the extrinsic driver, recommendation systems. In general, recommendation systems provide popular content or relevant content. By providing popular content, recommendation systems let popular online content appear at salient positions, making popular getting more popular (Szabo and Huberman 2010). By providing content that is related to the referent video a user is watching, recommendation system on YouTube is an important factor that drives video view count and this supplements the view instances from direct searching of a video (Figueiredo et al. 2011; Zhou et al. 2010). In addition to recommendation systems, social factors are also important extrinsic features that drive users' attention. Social network size of the YouTube videos and YouTube users positively contribute to video view count. That is, view count of a focal video is highly correlated with view count of videos in the same playlist (Cheng et al. 2008). Further, YouTube user friends' network size has significant impact on the growth of view count over time (Susarla et al. 2012). The effect of social network on attracting attention is strengthened by influencers within social networks, such that online content posted or reposted by influencers attracts more attention (Cha et al. 2010).

Sustaining Attention

Sustaining attention is important in practice. The direct benefit of sustaining user attention comes from the attention to the advertisements. For example, online videos, the advertisements could be in many formats. For instance, ads at the beginning of a video, ads that play in the middle of a video by interrupting the video, pop-up ads in the video picture area without interrupting, and ads that embedded in the video content. For pop-up ads and embedded ads, sustaining people's attention to watch the video longer is essentially important, as they often appear in the middle or even at the end of a video.

For online content creators, there is also indirect benefit of sustaining people's attention, which is building user loyalty. User loyalty is long term benefit for online content creators as it is the key to maintain substantial attention from the audience and thereafter generating monetary value. Users who spend more time on a piece of online content are more like to subscribe/follow the content creator and visit other content of the content creator in the future, because they enjoy the content and think it is valuable to do so.

Nevertheless, in literature, there is limited research focusing on the factors that sustain individual's attention. On social media platforms, structural difference in message presentation (use of image versus text-based content), brand utility and message valence matter to sustained attention (Daugherty and Hoffman 2014; Hoffman and Daugherty 2013). Specifically, non-luxury brand e-word of mouth (eWOM) received more sustained attention than luxury brand eWOM and negative message sustained more attention than positive and neutral message. When including message presentation structure in the analysis, for luxury brand, negative text-based eWOM sustained more attention while for non-luxury brand, negative image-based eWOM sustained more attention. Other important content text features are content topic, valance and color. In Navalpakkam et al. (2012)'s study, what content to deliver was found to be important to sustained attention and it is operationalized as content topic. In a study on the colored tags, which are the summary of high-frequency keywords of online reviews, color and valence of the colored tags were important to sustained attention (Xu and Zhang 2018). Position and saliency of online content were also found to be important features for sustained attention (Menon et al. 2016; Navalpakkam et al. 2012). On Facebook page that displays clothing products, the position and saliency of price are important factors affecting sustained attention (Menon et al. 2016). On web page that displays news articles, position and saliency of the article play important role in impacting sustained attention (Navalpakkam et al. 2012). For online retailers, in addition to the position and saliency of price, pricing strategy (time scarcity) was proved to be important factor that sustains individual's attention. Time scarcity refers to "time-left-to-buy" for the product. Research shows high-level time scarcity increases sustained attention (Mou and Shin 2018). Some studies explored the sustained attention in the advertisement context. An ad banner on a web page was paid at a low level of awareness supported by less sustained attention (Muñoz-Leiva et al. 2018). Similar results of the effects of position, brand, picture and text features on sustained attention in the online environment were revealed in printed advertisement in earlier studies (Garcia et al. 2000; Pieters and Wedel 2004).

In sum, current literature examined the effect of static elements or features of online content such as position of an item, text content features (valence, color, and topic), and style of content (pictures, links, and status) on the sustained attention, which is operationalized with fixation duration using eye-tracking method (Daugherty and Hoffman 2014; Hoffman and Daugherty 2013; Menon et al. 2016; Mou and Shin 2018; Muñoz-Leiva et al. 2018; Navalpakkam et al. 2012; Xu and Zhang 2018). However, there is lack of attention paid to the dynamic perspectives of online content. As the actual online content is more fruitful than what have been examined in literature from a static perspective, it is important to go beyond the static features and look into the dynamics of online content.

Table 1.1: Literature on Sustained Attention

Study	Context	The static element/feature
(Garcia et al. 2000)	Printed advertisement	Position of the illustration
(Pieters and Wedel 2004)	Printed advertisement	Brand, pictorial and text of advertisement
(Navalpakkam et al. 2012)	News article	Position, saliency and topic of the news items
(Hoffman and Daugherty 2013)	Social media platform: Pinterest	Image versus text-based elements, brand utility and message valence of eWOM
(Daugherty and Hoffman 2014)	Social media platform: Pinterest	Message valence and brand type of eWOM
(Menon et al. 2016)	Social media platform: Facebook page	Price (position and saliency) of products
(Mou and Shin 2018)	Online retailer platform	Time scarcity: “time-left-to-buy” for the product
(Muñoz-Leiva et al. 2018)	Hotel’s blog, Facebook, TripAdvisor	Appearance of ad banner
(Xu and Zhang 2018)	E-commerce platform	Color and valence of colored tags (summary of high-frequency keywords of online reviews)

3. Theory and Hypotheses

To look at what influences continued interest and attention, we will have to examine how individuals perceive environmental stimuli over time.

We are constantly bombarded by different types of environmental stimuli every day which vary in terms of their level of excitement. As individuals we response differently to the same level of stimulus we received, for example, not all individuals will find riding a kids' rollercoaster exciting. Research has shown that people have a need for an appropriate stimulation level which is called optimal stimulation level (Leuba 1955; Raju 1980; Steenkamp and Baumgartner 1992) to maintain interest and attention. Optimal stimulation level is an individual characteristic that varies from person to person. If the environmental stimulus is lower than an individual's optimal stimulation level, he or she feels bored and want to increase the stimulation level by changing the environment stimulus. On the contrary, if the environmental stimulus is higher than an individual's optimal stimulation level, he or she feels uncomfortable and tend to reduce the stimulation level by changing the environment stimulus. Further, optimal stimulation level is a relatively stable individual trait and does not vary within the same individual in the short run.

In this study, we examine the factors drive the sustained attention from a dynamic perspective using online user-generated videos. Video differs from text or static image content due to its dynamic features. In a video, images and audio are constantly displayed to the audience in a predesigned pace and order. Specifically, we examine the delivery linguistic content in videos. Given that user-generated content can exist in various forms (e.g. Videos, audio blogs, short tweets, extended Facebook posts) in order to ensure that the research is more generalizable we chose linguistic content as it is the common element of different formats of user-generated content. As a first step to examine the factors on sustained attention from a dynamic perspective, it is important to focus on the fundamental element of user-generated content which is the linguistic content. User-generate content is created to deliver information or message, therefore, what is delivered and how it is delivered matter to the potential audience when we focus on the dynamics of the linguistic content. This argument is align with research on sustaining students' attention during a lecture that what material to deliver and how to deliver the material matter to sustaining students' attention (Keller 1987a; Keller 1987b).

As there are multiple aspects of linguistic content, specifically for "what" aspect, we choose the topic of the linguistic content to examine its impact on sustained attention. The operation of using topic as an indicator of linguistic content is in line with literature (Bauer et al. 2012; Hua et al. 2005). There are three major aspects about how to deliver the linguistic content, tone of speech (Ishii et al. 2003; Jones and Macken 1993), intensity (volume) of speech (Jones et al. 1990; Mattys et al. 2012) and pace of speech (Grobe et al. 1973; Kormos and Dénes 2004). In this study, we examine the pace of delivering the linguistic content to on sustained attention. Intensity of speech is adjustable by the audience for online videos therefore we didn't examine that factor. Tone of speech is also a possible factor that impact audiences' sustained attention. We understand it is a limitation of this study that we didn't examine other aspect of delivering linguistic content. It is further discussed in the limitation section.

For user-generated videos in social media, we argue that the temporal stimulation level influences users' sustained attention. When a social media user views a particular content, sustained viewing is contingent upon the pace of which the linguistic content is being delivered. Linguistic content delivery pace doesn't equal to speech speed. It is about pause between words and sentences. Consistent linguistic content delivery pace means consistent length of pauses. Based on the theory of optimal stimulation level, maintaining stimulation consistently within a reasonable range is essential to prevent the user from being under or over-stimulated, and hence continue viewing. Consistency in content delivery pace helps to keep the stimulation level at a certain range, which is an ideal range for the people who decide to watch this video. Inconsistent informational linguistic content pace means long block of silence occurs when delivering information, which results in drop in stimulation level. In social media, users usually discontinue viewing because of a drop in stimulation level. For example, a professor is giving an interesting lecture but interrupted by some technical fault. In this case, the consistent informational linguistic content pace is interrupted and students' attention are likely to be momentarily lost. Therefore, we expect that consistency in informational linguistic content pace facilitates sustained attention.

Hypothesis 1 (H1): On social media platforms, consistency in informational linguistic content pace is positively related to sustained attention.

Prior research in consumer behavior has shown that variety in general increases stimulation level. Theory of exploratory consumer behavior states that consumers have a desire for exploration which is exhibited by variety seeking behavior

(Joachimsthaler and Lastovicka 1984; Raju 1980; Steenkamp and Baumgartner 1992). Variety seeking behavior has "the capacity to lead to exciting and novel experiences, to relief from boredom, and to satisfy one's desire for knowledge and the urge of curiosity" (Baumgartner and Steenkamp 1996). Variety in general satisfies consumers' cognitive stimulation needs by increasing stimulation level (Baumgartner and Steenkamp 1996; Raju 1980; Steenkamp and Baumgartner 1992). In our research context, video content variety refers to the variety in linguistic content topics. Linguistic content of a video reflects overall video content of conversation based videos. When the video content topic varies across some sub-topics within a high level theme, we expect the stimulation level is maintained over time because of changing topics, and then user's attention is sustained. We have the above expectation because variety promotes or at least maintain the stimulation level which is in line with psychology and attention literature (Baumgartner and Steenkamp 1996; Joachimsthaler and Lastovicka 1984; Raju 1980; Steenkamp and Baumgartner 1992). Leading from these studies, one can conclude that when there is no topic variety, stimulation level drops after maintain discussion on the same topic for a while. Variety in topic prevents stimulation level from dropping below the optimal stimulation level. Users' interests are continually captured by the time-to-time changed topics when the various topics are within the boundary of video category. Further, psychology literature has also suggested the lack of variety during the consumption of goods (information goods in this case) often lead to satiation (Sevilla et al. 2019).

Therefore, we argue that informational linguistic content variety has positive effect on sustained attention.

Hypothesis 2 (H2): On social media platforms, linguistic content topic variety is positively related to sustained attention.

4. Methodology

The YouTube Context

We use YouTube as the research context in this study. YouTube is the leading video sharing social media platform in the world with over 1 billion users and 3.25 billion hours of viewership per month (YouTube 2016). On YouTube, each registered user has a YouTube account and these accounts can subscribe to other accounts as well as being subscribed by others. The act of subscription is just similar to "following" on Twitter or "being friends" on Facebook. To view videos, users can either browse the home page which lists the popular videos or the videos the user may be interested in basing on the browsing history or actively search for them. For registered users, popular videos uploaded by accounts they have subscribed to are also presented on their home page when they login.

The predominant information content provided on YouTube is video and similar to online articles on other social media platforms, as only limited information (such as title, content source, and publish time) is presented with the link. Users need to click the link in order to view the full content and like most information good, consumption of the good requires the investment of time and attention.

Data Collection Approach

We collect the data from YouTube between 22 July 2016 and 22 September 2016. We create Python scripts with YouTube data API to extract all the accounts from two categories, entertainment and sports. We choose these two categories because entertainment and sports contains a wide variety of videos whereby there is wide heterogeneity in terms of viewership lengths and popularity. More importantly, videos from these two account categories are majorly informational and conversation based. Notably, videos in our dataset don't have TV drama or movie that are story based, or pure music videos that are not conversation based. We captured all the videos published in these two categories by all the accounts from 22 July 2016 till 19 Aug 2016 (4 weeks duration) and for each video, we repeatedly collect all relevant data on a daily basis for at least one month. Daily collection is required as some data, such as number of views, number of likes, and number of comments is dynamic and keeps changing over time. Data from a total of 13784 videos is collected.

In order to investigate how the video conversation pace and video topics affect sustained attention, we extract the video captions generated by YouTube for the uploaded videos. We filter out videos with only one-sentence caption, as they commonly include only the video title and not the contents of the video. Given that the captions are automatically generated by YouTube system, we often observe that poorly-short videos with inferior sound quality are often without captions due to difficulty experienced in the captioning process. The final data set has 9492 unique videos with data collected over a period of 31 days. This data represents a sample of relatively well-designed, properly formatted, quality social media content, and this makes it an ideal sample for us to examine the factors that impact attention retention.

Text Mining

To obtain the proxy measures for coherence in video topic, we conduct topic modeling on the video captions. Topic modeling is a text mining technique that applies statistical methods to discover the topics that appear in a collection of documents. In this study, we adopt the Latent Dirichlet Allocation (LDA) topic model (Blei et al. 2003), which is commonly used in prior studies (Wei and Croft 2006; Weng et al. 2010; Zhao et al. 2011). The LDA topic modeling is conducted using R-Studio (Hornik 2015; Hornik and Grün 2011; RStudio 2015). The routine of text mining is as follows. First, the caption text is preprocessed by removing, stemming, and striping white space. The removing procedure removes non-English characteristics, numbers, stop words, and punctuation. Stemming is a procedure to reduce words to their root (dictionary) form (e.g. inflected words like running is stemmed into run). Striping white space deletes the redundant white space within a sentence or between paragraphs. The preprocessing procedure is to reduce redundancy of words for processing. The words are subsequently used to generate a document term matrix that is subjected to variational expectation-maximization (VEM) algorithm to fit the topic models. Words that commonly appeared alongside one another within a corpus of text are more likely to be lexically similar and hence likely to below to a similar topic. To illustrate this intuition, for text that discuss the topic of color, we are more likely to see the use of words such as red, blue compared to a topic which discusses taste where we are likely to see words such as sweet and sour appearing alongside one another. In running the algorithm, we define the total number of topics (*k*) will be generated and tried various iterations of *k*. Each discovered topic is represented by the most frequent words and topic modeling generates loadings of each document for each topic. A loading is the probability of the document belongs to the corresponding topic, and this is a similar score which measures the fit of the content with a particular topic. For every video caption, it is scored against all topics, *k*, and we use the loadings to calculate the degree of video topic vareity (operationalization details of this construct in the next section).

Variables

The dependent variable of this study is *sustained attention*. Sustained attention is the amount of attention a user pays to a video. Conceptually, the longer one spends watching a video, the longer the attention is being attributed. We use the ratio of daily average video view duration to the total video duration to measure *Sustained attention*. Average video view duration is the average time a video is watched by users. Therefore, the higher the ratio, the higher proportion a video is watched i.e. the longer attention is sustained with this video. We use the ratio rather than the absolute video view duration as videos vary in duration, and taking absolute duration will result in dependent variable that has different natural maximum durations.

In the video caption data, we have both the text and the timing of each word. To proxy the consistency in informational linguistic content pace, we measure consistency in *video conversation pace*. Here, we identify time interval between words. In general, time interval between words within a sentence is quite short, while time interval between the last word of a sentence and the first word of a subsequent sentence varies from short to long. A long time interval indicates no conversation in the video and this occurs in between sentences of captions. We measure the time duration of each time interval and compute the standard deviation of this durations throughout the video. A video that has a consistent video conversation pace will have a low standard deviation; while an inconsistently paced video with conversations punctuated between long blocks of silence will have high standard deviation. Given that the videos in this sample have substantial captions, we believe that these videos are conversational in nature and do not contain solely moving images. As a result, the conversational content represents an important part of the stimuli that is required to maintain the user's attention.

Another independent variable is *video topic variety*. To operationalize, we adopt the concept of Herfindahl-Hirschman Index (HHI) to calculate video topic variety. HHI is an economic concept to measure the industry competition and is calculated by the sum of the squares of the market shares of the firms within the industry (Rhoades 1993). Market shares are expressed as fractions and the HHI ranges from 0 to 1. High HHI represent the existence of monopoly while a low HHI presents a fragmented market. We calculate the video topic variety as follow. Using the topic modeling results describe in the previous section, every video, i, loads on each topic, j. This loading, L_{ij} represents the degree of which the contents of the video fall into the particular topic. Given that each topic constitute a collection of commonly occurred words which are lexically similar, a video which spread across multiple topics (i.e. fragment and low HHI score) is likely to has higher variety than another which is concentrated into fewer topics (i.e. similar to a concentrated market with high HHI). Here, define the video topic variety measure as V_i

$$V_i = \sum_{j=1}^{k} \left(\frac{L_{ij}}{\sum_{j=1}^{k} L_{ij}}\right)^2 \qquad (1.1)$$

where k is the total number of topics for all the videos. Intuitively, a higher V_i, suggests lower content variety and a lower V_i, suggests higher content variety. In our analysis, we use 60 as the number of topics (k = 60). Given that the dataset contains many videos on various topics, to cover these topics, a large k is more reasonable than a small k (Hong and Davison 2010). In this study, we have 9492 video caption documents. A meta-analysis on research using topic modeling shows that when the number of documents is large, 60 topics is the inflection point such that lower or more than 60 topics results in a lower semantic coherence

(Schmiedel et al. 2018). We also try various iterations of *k* from 40 to 80 (in intervals of 10), and find similar results for *k* from 40 to 80.

Control variables

In addition to audible stimulus, sustained attention is also affected by other factors such as video overall quality, image features, and publisher (channel) reputation and experience. To control for video overall quality, we control video cumulative view count (*video_view_cum*), video like count (*like*), video dislike count (*dislike*), video comment count (*comment*), and daily sharing count (*daily_sharing*). To control for video image features, we control the *color contrast* of video thumbnail. Color contrast is measured by the gradient of each pixel on red, green, blue (RGB) values in a video thumbnail. We calculate the gradients on R, G and B of each pixel in thumbnails and then get the variance of gradients of all pixels on R, G and B respectively. A higher gradient represent an image with more vibrate and striking variation in color tones. Finally, we take average of the three variance value to measure the color contrast. On video level, we also control for *video age*, *video duration*, video conversation ratio (*conv_ratio*), video tag count (*tag*), and *video category*. Video conversation ratio is the ratio of conversation duration to total video duration. Videos in the dataset are from 15 categories that is classified by YouTube. To control for publisher (account) reputation and experience, we control for *channel age*, channel subscriber count (*no_follower*), and channel video count (*channel video*). We also control for channel *location* that is the geographical location of a channel.

Analysis and Models

We propose that the sustained attention (*sust_attention*)[1] of the video is a function of the conversation pace (*conv_pace*) (H1), topic variety (*topic_variety*) (H2), and other exogenous time invariant control factors K_i and time variant control factors J_{it}.

$$Sust\_attention_{it} = \gamma_0 + \gamma_1 Conv\_pace_i + \gamma_2 Topic\_variety_i + \gamma_3 K_i + \gamma_4 J_{it} + U_i + W_t + \varepsilon_{it} \quad (1.2)$$

Where γ represent parameters of the model, ε represents stochasticity across video and time, *U* represents video level stochasticity, and *W* represents time level stochasticity.

In (1.2) we included a time-related error term *W* to partial out any stochasticity due to temporal differences (e.g. time periods with more online traffic due to seasonal differences). Similarly, we also included a video-level error term *U* to control for any video-level idiosyncrasies which were not captured by our exogenous variables. We estimate (2) using random effects GLS regression to mitigate any heteroscedasticity issues.

5. Results and Discussion

The regression results are shown in Table 1.1. As hypothesized in H1, we find negative and significant relationship between video conversation pace value and sustained attention (coeff: -3.126; p-value < 0.001). A high standard deviation in

[1] Sust_attention is a portion of two variables and the value is between 0 and 1. In the regression, we multiply the value of sust_attention by 100, so the unit of the variable sust_attention becomes %.

conversation pace indicates long block of silence exists when delivering information. When the standard deviation in the conversation pace increases (inconsistent conversation pace), we see a drop in the proportion of the video being viewed. In other words, higher consistency in conversation pace helps to maintain attention, supporting Hypothesis 1. We also find that topic variety value has negatively significant effect on sustained attention (coeff: -1.863; p-value < 0.01). Remembering the calculation of topic variety, a high value of the variable means low topic variety while a low value means a high topic variety. The result shows, increasing in topic variety value decreases sustained attention. That is to say, lower topic variety leads to less sustained attention, or higher topic variety leads to more sustained attention, supporting Hypothesis 2.

Table 1.2. Regression Results to Examining Drivers of Sustained Attention

DV: Sustained Attention	
Variables	Coefficients (Std. Err.)
Conv_pace	**-3.126***** (0.159)
Topic_variety value	**-1.863**** (0.592)
Channel age	0.001*** (0.000)
Control for channel location	<suppressed for brevity>
No_follower	1.09e-06*** (0.000)
Channel video	4.84e-06 (0.000)
Video age	-0.183*** (0.002)
Video duration	-0.018*** (0.000)
Conv_ratio	-18.353*** (1.027)
Color contrast	2.08e-5 (0.000)
Tag	-0.041*** (0.011)
Video_view_cum	-6.40e-07 (0.000)
Like	-3.97e-5* (0.000)
Dislike	-1.472e-4*** (0.000)
Comment	-1.082e-4 (0.000)
Daily_sharing	0.003*** (0.000)
Control for video category	<suppressed for brevity>

Note: * p<0.05; ** p<0.01; *** p<0.001

In addition to the hypotheses, we observe some other interesting results. We find that video age is negatively and significantly related to sustained attention, which suggests many videos may contain popular topics the population is interested in but these interests wane over time. We also observe that longer video duration relates to shorter sustained attention, providing further evidence that people have difficulty to keep continued interests and attention with online content. Further, we find that quantity of speech content, which is operationalized as conversation ratio, is negatively related to sustained attention. This result indicates that it is not about how much linguistic content is delivered. In fact, the result shows more is less.

The estimates from equation (2) suggest that consistency in informational linguistic content pace and variety in content topic are important to sustained attention. Consistent video conversation pace keeps the stimulation at a certain level and maintains the interest level of the viewer. When video conversation pace is not consistent, users will experience periods of intense content followed by slow, monotonous content, resulting in a drop in stimulation level. Further, changing the linguistic content topic time-to-time is another way to maintain the stimulation level.

The results of this study only hold for user generated content as this is the scope of this study. We are not clear about other forms of video content as there might be some fundamental difference between user-generated content and professionally produced content. Actually, the results of this study are not applicable to all user generated content. Due to the research design, we only focus on information based videos in this study. Therefore, user generated videos with pure images and purely playing music are not within the scope of this study.

6. Contribution to Literature and Practice

In this study, we use YouTube as the research context to take exploratory steps to examine what are the factors that sustain attention of individuals from a dynamic perspective in this information economy.

This study contributes to the existing conversation of attention theory in psychology particularly within the social media context. Existing literature only looks at factors that contribute to attracting attention, explains what kind of stimuli can get through the information processing bottleneck, with limited discussion on what contributes to sustaining attention of the stimuli beyond the bottleneck from a static perspective (Daugherty and Hoffman 2014; Hoffman and Daugherty 2013; Menon et al. 2016; Mou and Shin 2018; Muñoz-Leiva et al. 2018; Navalpakkam et al. 2012; Xu and Zhang 2018). This study identified some factors that impact on attention retention in the social media context from a dynamic perspective and our results suggested that consistency in informational content delivery pace and content topic variety are critical to sustaining attention.

Literature on consumer behavior research shows variety promotes interests and stimulation level. However, the studies were in early time using questionnaire (Joachimsthaler and Lastovicka 1984; Raju 1980; Steenkamp and Baumgartner

1992) or laboratory experiment (Baumgartner and Steenkamp 1996). In the new age of social media with the phenomenon of information overload that people are exposed to plenty of information and people's attention is so easily being distracted, whether the argument is still valid is not clear.. Our study contribute to literature by examining the effect of variety in the social media context. We believe that the results of this study are not just limited to linguistic content in videos, but can be generalized to other text-based user generated contents or audio recordings on social media platforms.

This study also contributed to the practice. As discussed before, due to the limited attention possessed by people, there exist severe competitions among social media online content creators. Getting attention from the population is important for social media online content creators as monetary returns come along with attention (Goldhaber 1997). The results of this study provide implications for online content creators by identifying factors that can sustain users' attention, especially for YouTube video uploaders. Video content providers who develop their videos with attention retention in mind can design the content with consistent video conversation pace and variety in video topic selection.

Finally, in practice, sustaining users' attention is important because the more time users pay attention to a particular online content, the more likely the user will notice the advertisement – which is the major source of income for many social media online content creators who provide free-access contents. It is especially important to pop-up ads in the video picture area without interrupting the video and ads that embedded in the video content which requires more sustained attention from the audience. In sum, content creators can adjust their content creation strategies basing on the results of our study to attract and sustain more attention, and achieve success.

7. Limitations and Future Research

This study is not without limitations. For this study, it is better to examine this question at an individual user level if we have video-users' individual characteristics and watching behavior information. However, due to practical data collection reasons, we examine the phenomenon using individual video over time as the unit of analysis and not individual user viewing patterns. As a result, we are unable to measure individual's optimal stimulation level, and rely on the law of averages to approximate that if a user is attracted by a video (by viewing it), the *initial* or *expected* stimulation level should be close to his or her optimal stimulation level. As the individual user level information is not available in the field (in a social media site), one possible way to mitigate this limitation in the future is to conduct experiments to track users' physiological response to videos with different conversation pace and videos with different topic coherence level.

The second limitation of this study is that we didn't consider the images during a long pause. It could be possible that a long pause may not necessarily be of low stimulation if there are other actions or emotional scenes going on. We are searching for effective methods and tools to analyze the visual content. One possible future research is to look at how the stimulation level raised by the linguistic content and the stimulation level raised by the visual content interactively impact people’s sustained attention.

In this study, we didn’t examine other aspect of delivering linguistic content. One possible future research is to use some machine learning techniques to detect the tone of the speech and the emotion reflected in the on-going images of videos. This would allow us to examine how the tone of speech influences sustained attention and further whether the emotion reflected in the video image and emotion of speech collectively play a part in sustaining audiences’ attention.

The presence of an ad that plays in the middle of a video by interrupting the video may affect the sustained attention as there is interruption in the content delivery pace and it provides irrelevant information. In this study, we didn’t consider the effect of advertisements because the insertion of advertisements in a YouTube video is not readily measured and can be dependent on various factors such as timing and browsing history of the individual. Nevertheless, assuming that the algorithmic process of inserting these advertisement are consistent within the time period of two months of which we collect the data, we should see limited impact on our results as they can be captured by the video level error term specified in our estimation model. However, to completely mitigate this issue, future research can consider conducting experiments to test the effects of advertisements in a video on sustained attention. Theoretically, the ads that play in the middle of a video interrupt the content delivery pace and would lead to a drop in stimulation level. But the situation is bit different for a drop of stimulation level due to advertisement or due to the linguistic content. Advertisement is interruption, however, audience would expect the advertisement end within a known time period, but whether the drop of stimulation level would recover soon is unknown. It is would be interesting empirically investigate the effects of position in a video and length of advertisements on sustained attention.

8. Conclusion

In this study, we took exploratory steps to examine the factors that sustain attention of individuals from a dynamic perspective in this information economy. We situated our study in an empirical context and hope to unpack the psychological underpinnings which lead to greater focus and attention. We used YouTube as the empirical setting to examine the factors that influence sustaining attention. The results of this study showed that consistency in informational content delivery pace and variety in content topic matter to sustain the attracted attention.

CHAPTER 2: Impact of Online Review on Consumer Post-purchase Attitude Change

1. Introduction

In the e-commerce age, it is challenging for consumers to choose a suitable product or service from thousands of providers on e-commerce platforms. Nevertheless, e-commerce platforms not only provide overwhelming choices, but also give us opportunity to learn purchasing experience from other consumers. Online consumer review has become an important information source for consumers to evaluate and judge product and service quality and to make purchase decisions (Babić Rosario et al. 2016; Baker et al. 2016; Cantallops and Salvi 2014; Chevalier and Mayzlin 2006; Lu et al. 2014; Schlosser 2011; Sparks and Browning 2011; Sun 2012; Tang et al. 2014; Zhu and Zhang 2010). However, sometimes purchasing and consumption are two separated stage of an entire transaction. For example, a customer may pre-order a product online and will only receive the product two weeks later, or a customer may book a hotel one month in advance and will complete the transaction only when the customer check in at the hotel one month later. Under these situations, purchasing behavior (e.g. pre-ordering a product, booking a hotel) does not guarantee the completion of the transaction. Consumers have the chance to regret by cancelling the order or booking before they consume the product or service. Although research evidence has shown that, on average, online review is positively correlated with sales, purchase behavior, and purchase intention (Chevalier and Mayzlin 2006;

Lu et al. 2014; Ye et al. 2009; Ye et al. 2011), little attention is paid to the effect of online review on consumer post-purchase attitude change. Specifically, how the online reviews appear after purchasing (post-purchase review) can shape the belief of consumers on the product or service and eventually change the purchase decision (cancelling the order or booking) is not clear. It is important to answer this question as many industries are facing the same situation, such as hotel industry, car rental industry, and many firms that provide pre-ordered products like software, video games, and iPhone.

Online review consists of two parts, the numeric rating and the textual comment. There are extensive examination on how the numeric rating affects consumer decisions from two dimensions, valence and variance of the ratings (Baker et al. 2016; Basuroy et al. 2003; Chen et al. 2011; Chevalier and Mayzlin 2006; De Langhe et al. 2015; Kwark et al. 2016; Lu et al. 2014; Ludwig et al. 2013). In addition, online review volume is another factor that is used to predict consumer decisions (Chevalier and Mayzlin 2006; De Langhe et al. 2015; Liu 2006). Nevertheless, limited research effort has been made to investigate how the textual information of online reviews influence consumer decision (Schuckert et al. 2015). With the increasing accurate rate of text mining technique, mining the textual information provides more fruitful insights on how consumers utilize online review to make decisions. Literature, in general, shows a positive relationship between pre-purchase online review rating and sales (Babić Rosario et al. 2016; Chevalier and Mayzlin 2006; Lu et al. 2014; Ye et al. 2009; Ye et al. 2011). However, we are lack of knowledge on whether consumers re-check the online reviews, both numeric rating and textual comment, after purchasing or not and how the post-purchase reviews lead to consumer attitude change (cancellation behavior).

In this study, we conduct a pre-test to survey if consumers do re-check online reviews after purchasing and apply empirical analysis to find out how the online reviews after consumers make purchase decision but before consumption can lead to consumer regret from two dimensions, numeric rating and textual comment. We choose hotel industry to conduct the empirical analysis. Majority of customers book hotels through online travel agencies, such as Booking.com, Expedia.com, and Agoda.com, in advance and will only consume it at a later time. A booking can be canceled before check-in date. The nature of hotel business allows us to observe consumers' booking and cancellation behavior. The observable interval between hotel booking date and check-in date provides possibility to track the online reviews published after a customer book a hotel.

From numeric rating dimension, we test the interactive effect of pre-purchase rating and post-purchase negative rating on the cancellation behavior. Literature points out the negative information has relatively larger impact on consumer decision making than positive information and this phenomenon is called negativity bias (Basuroy et al. 2003; Chen et al. 2011; Chevalier and Mayzlin 2006). We show that negativity bias also exists in the consumer post-purchase attitude change context. We find that negative post-purchase rating leads to more cancellation and it is strengthened by a higher pre-purchase rating. From textual comment dimension, we classify review text into hedonic comment and utilitarian comment. Then we test the different effect of hedonic comment and utilitarian comment on cancellation behavior. We find consumers' different trust levels on hedonic vs. utilitarian comments. The results show that consumers distrust highly rated hedonic comments, exhibiting more cancellation behavior.

This study have important theoretical and practical contributions. From an academic standpoint, the findings contribute to consumer online review literature and consumer regret literature. We find that consumers do re-check the online review and make decisions basing on the newly posted post-purchase review. We add knowledge to consumer online review literature that negativity bias exists in the process of consumer post-purchase attitude change and distrust is placed on hedonic comment. We also add knowledge to consumer regret literature by examining a nontraditional type of consumer regret (regret because of non-personal experienced information). From a practice standpoint, a good understanding about the mechanism that consumer online review works on influencing consumer decisions makes marketers better prepared to potential challenges. Although we use hotel industry as the empirical setting, the results of this study are generalizable to other context such as car rental and pre-ordered products (e.g., software, video games, and iPhone). For some context, the online reviews may not be well organized on the purchase platform but are spread out across different online forums and social media platforms. When a consumer has a specific information source (a preferred online forum or social media platform) to look for product reviews, the features of the reviews on this particular information source are expected to have similar impacts on the consumer's decisions (possible attitude change) as the results of this study.

We will review the literature on consumer online review and consumer regret in section 2 and set up the hypotheses in section 3. In section 4, we outline the research context and the methodology employed followed up by the results in section 5. We report robustness check results in section 6. Section 7 and 8 discuss the contributions and limitations, respectively. Section 9 concludes.

2. Literature

Consumer Online Review

The value of online review is reflected on its impact on consumer decisions. Comparing to offline word of mouth, online review has great accessibility and is much more impactful to more audience. Since the emergence of e-commerce, online review has become an attractive research topic for scholars. A typical online review consists of two parts, the numeric rating and the textual comment. In the past, due to the constraint of natural language processing, most scholars limit their research on the numeric rating aspect of online reviews. Two popular variables are valence (Baker et al. 2016; Basuroy et al. 2003; Chen et al. 2011; Chevalier and Mayzlin 2006; De Langhe et al. 2015; Kwark et al. 2016; Lu et al. 2014; Ludwig et al. 2013) and variance (Lu et al. 2014; Sun 2012) of online review rating. In addition, online review volume also attracted certain attention (Chevalier and Mayzlin 2006; De Langhe et al. 2015; Liu 2006).

Online review valence is found to be an important factor that explains the variation in consumer decisions. Research in different contexts shows that, in general, positive reviews help to increase sales. In an early study on the effect of online review on book sales, Chevalier and Mayzlin (2006) found that improvement in book review leads to increase in book sales. In hotel industry, a series of studies showed that the average rating of online review also have positive significant impact on hotel sales (Lu et al. 2014; Ye et al. 2009; Ye et al. 2011). In addition to the findings that positive online review leads to more purchase intention and sales, and negative online review leads to less purchase intention and sales, previous research also suggest that the effect of positive online review and negative online review are asymmetric (Baker et al. 2016). The asymmetric effects are reflected on research findings that negative online review has relatively larger effect than positive online review on sales and is called negativity bias (Basuroy et al. 2003; Chen et al. 2011; Chevalier and Mayzlin 2006). The phenomenon of negativity bias is found in film industry (Basuroy et al. 2003) as well as online retailer (Chen et al. 2011; Chevalier and Mayzlin 2006), through natural experiment (Chen et al. 2011) and empirical analysis (Basuroy et al. 2003; Chevalier and Mayzlin 2006).

Another stream of studies looks at the explanation power of textural content of online reviews. Textural content contains richer information than numeric rating and is expected to explain a large part of variation in consumer decisions (Archak et al. 2011; Schellekens et al. 2010; Schlosser 2011; Tang et al. 2014; Yin et al. 2017). Numeric rating only indicates an overall judgement of reviewers, which is a one-dimensional indicator of product/service quality, while textural content provide concrete information from different aspects of product/service as well as reviewers' emotion. For a given product/service, consumers evaluate different aspects differently (Archak et al. 2011). The research on online review textural content is mainly from three perspectives, reviewer affect, product/service feature, and textual content valence. From reviewer affect perspective, Ludwig et al. (2013) found that positive affective content and negative affective content showed different effect on conversion rate, such that positive affective content has a diminishing marginal effect while negative affective content doesn't have such effect. Another study by Yin et al. (2017) also found a similar diminishing return pattern of the effect of expressed emotional arousal in online review on reader perceptions of its helpfulness. From textual content valence perspective, Schlosser (2011) found that comparing to one-sided argument (positive or negative argument only), two-sided argument reduces online review credibility perceptions, helpfulness, and persuasiveness. However, when considering neutrally rated textual content only, mixed-neutral textual content, which contains an equal amount of positive and negative claims, amplifies the effects of positive and negative online review. From product/service feature perspective, Archak et al. (2011) did find different product features have different impacts on product sales.

Plenty of research has examined the effect of online review on various things, such as sales, purchase intention, brand evaluation, and perceived helpfulness of online review (Baker et al. 2016; Chevalier and Mayzlin 2006; Schlosser 2011;

Tang et al. 2014). However, to some types of product or service, purchase behavior is not an end. Consumers can choose to return the product or cancel a booked product or service. To the best of our knowledge, little attention has been paid to the effect of online review on consumer post purchase attitude change.

Consumer Regret

Consumer regret normally refers to consumers' feeling of regret after making purchasing decisions. Studies have examined the antecedents and consequences of consumer regret (Abendroth and Diehl 2006; Bui et al. 2011; Inman and Zeelenberg 2002; Lee and Cotte 2009; Tsiros and Mittal 2000). Consumer regret is mainly caused by disconfirmation between expectancy and actual user experience and the discovery of better alternatives. The feeling of regret affect consumer choice to repurchase or to switch brand. Traditional research on consumer regret is conducted under the condition that consumers has consumed the product/service and feel regret. The information they get is directly from the consumption of the product/service, which is very trustworthy as it's from their personal experience. However, when consumers have the channel to acquire updated product/service quality information from others before they personally consume the product/service, how consumer attitude and decision would be affected remains unclear. This study will answer this question.

3. Theory and Hypotheses Development

It is not questionable that many consumers read online reviews before making purchase decision. Plenty of research on consumer online review has shown that reviews do affect consumer decisions. In this study, we argue that consumers usually re-check product/service review after purchasing and before actual consumption. Loss aversion may account for consumers' tendency to go back to check online reviews. There is asymmetry effect of loss and gain in consumer choice. The impact of loss is greater than that of gain (Tversky and Kahneman 1991). Consumers naturally fear to loss. When there is a time gap between making purchase decision and receiving the product or consuming the service, consumers have chance to reconsider the purchase decision. For consumers, bad experience with a product/service is regarded as loss because they pay for a product/service that is not satisfying. To decrease the probability of loss, consumers are very likely to seek for more information to confirm if they made the right decision. The assumption that consumer usually re-check online review after purchasing is fundamental to the hypotheses development in the rest of this section.

The Negativity Bias

The Negativity bias is the phenomenon that negative information have relatively larger effect than positive information on consumer decisions (Basuroy et al. 2003; Chen et al. 2011; Chevalier and Mayzlin 2006). In the online review context, frequency-as-information might account for the negativity bias (Chen and Lurie 2013). That is positive information is dominant in online review (Chevalier and Mayzlin 2006), making rare negative information more informative (Fiske 2018; Peeters and Czapinski 1990). The frequency-as-information account is supported by research findings that positivity bias exists in a context where negative information is dominant (Rozin and Royzman 2001).

We argue that negativity bias also exist in the effect of post-purchase online review on consumer attitude change. We have hypothesized consumers seek for information to confirm that they made the right purchase decision. The relatively rare negative post-purchase review has greater impact than positive post-purchase review. That is to say, despite positive pre-purchase and post-purchase reviews, the newly posted negative post-purchase reviews trigger consumers to re-consider the purchase decision due to negativity bias. The more number of negative post-purchase review consumers see, the more likely they would change purchase decision. Therefore, we hypothesize a positive relationship between number of negative post-purchase review and cancellation probability. It is notable that even though we hypothesize the impact of more number of negative post-purchase review, negative review is rare in general comparing to positive rating.

Hypothesis 1 a (H1a): Number of negative post-purchase review is positively related to the likelihood of cancellation.

Cognitive Dissonance

The effect of negative post-purchase ratings has boundary conditions. It depends on the pre-purchase rating according to the theory of cognitive dissonance. Cognitive dissonance is the psychological discomfort comes from conflicting attitudes, beliefs, or behaviors (Akerlof and Dickens 1982; Festinger 1962).

Consumer decisions are affected by multiple factors in different stages. We believe the rating information before making purchase decision has impact on post-purchase behavior. When consumers see a high rating before purchasing but negative reviews after purchasing, cognitive dissonance is raised by the disagreement between expectation and possible reality. According to the theory of cognitive dissonance, consumers would feel very uncomfortable and try to reduce cognitive dissonance by taking proper actions (Festinger 1962). As a result, consumers are very likely to eliminate cognitive dissonance by cancelling the order/booking and looking for alternatives. When there is negative post-purchase review, the higher pre-purchase rating they see, the greater cognitive dissonance they have. The greater cognitive dissonance they have, the higher desire to reduce it. Therefore we hypothesize the interact effect of pre-purchase rating and post-purchase negative ratings that is higher pre-purchase rating would strengthen the effect of negative post-purchase review on cancellation possibility.

Hypothesis 1 b (H1b): This positive relationship between negative post-purchase review and likelihood of cancellation is strengthened when pre-purchase rating is high.

Hedonic vs. Utilitarian Comment

The nature of online review is consumer attitudes toward product/service. Consumer attitudes is a two-dimensional concept, hedonic dimension and utilitarian dimension. We differentiate the comments into utilitarian and hedonic perspectives to examine their different effects on consumer decision making as there are fundamental difference between utilitarian and hedonic perspectives of product/service comments. Hedonic dimension is affects and sensations from the experience of using products, and utilitarian dimension is functional features of products (Batra and Ahtola 1991; Voss et al. 2003). In textual content of online review, hedonic/utilitarian comments are the review comment from hedonic/utilitarian dimension.

Hedonic comments reflect reviewers' personal feelings, opinions and affects (Hirschman and Holbrook 1982). For example, "The staff are very friendly" and "Nice service and nice stay". Hedonic comments are subjective (comparing to utilitarian comments) and not verifiable as different individuals can have different expectations and judgment reference level, which result in different affective opinions (Yang and Lee 2010). Something is good to individual A, but probably is just fare or even unsatisfied to individual B. Subjective reviews are less credible and trustworthy than objective reviews and consumers are more willing to accept objective comment without strong subjective emotion (Filieri 2016; Luo et al. 2015; Park and Lee 2008). Furthermore, extreme positive ratings are more likely to be viewed as untrustworthy as they are more likely to be perceived as promotional (Filieri 2016). Therefore, when a high post-purchase rating is due to more hedonic comments, consumers start to concern the trustworthiness of the hedonic comments because they are not verifiable and the very positive hedonic comments seem too good to be true. Therefore, we argue that consumers are more likely to cancel the order/booking when there are more post-purchase hedonic comments come with higher ratings.

Hypothesis 2 (H2): In post-purchase review, higher ratings due to more hedonic comments lead to higher possibility of cancellation.

Utilitarian comments are statements of functional perspective of products without affective judgement (Ryu et al. 2010). For instance, "The hotel has free shuttle bus to airport" and "The hotel has swimming pool and gym". As a high rating indicates high quality, when a high rating is due to utilitarian comments, consumer perceive high quality of the mentioned utilitarian perspectives of the product. In addition, the perception of high quality is trustworthy because utilitarian comments are objective and do not change with personal preference. Therefore, more utilitarian comments with higher rating suggest a reliable confirmation of product/service quality and consumers are less likely to cancel the order/booking.

Hypothesis 3 (H3): In post-purchase review, higher ratings due to more utilitarian comments lead to lower possibility of cancellation.

4. Methodology

Pre-test

There is a critical assumption in this study that consumers usually re-check online reviews after purchasing and may change purchase decision because of post-purchase reviews. This assumption is fundamental to the hypotheses development and discussion in this study. In order to verify this assumption, we conducted a simple pre-test on Amazon Mechanical Turk (MTurk). MTurk operates a marketplace for work that requires human intelligence. We published survey on MTurk to ask people if they do check the updated hotel reviews after booking a hotel room. Valid respondents are those who have hotel booking experience with OTAs. In the survey, we asked following questions, "recall your last booking, did you check the hotel reviews, after you book a room and before the check-in date, on either the platform you book the room or third-party review websites like TripAdvisor.com" and "If your answer is yes, did the hotel reviews after you book a room and before the check-in date cause you to cancel the booking". In addition, this survey also included demographic questions about respondents' gender, age, and education level. We conducted several pilot tests to make sure the questions are clear and without ambiguity. In the pre-test, we sent out 100 questionnaires and got 97 valid responses. The pre-test results will be discussed in section 5.

Data

We conduct an empirical study in the hotel industry. A hotel (represented by "focal hotel" in the rest of the paper) in Singapore provided us their customers' online booking records for one year period, from 1st October 2015 to 31st September 2016. The focal hotel is a popular four-star hotel that locates at a convenient place in Singapore. The majority of online bookings of this hotel came from three online travel agencies (OTAs), Agoda.com, Booking.com and Expedia.com. In total, about 18% of all hotel bookings (including online and offline) come from the three platforms. The hotel is rated as very good on

Agoda.com, Booking.com and Expedia.com. The average booking price is 180 Singapore Dollar and the average occupancy rate of the hotel is 86% during the one year period, from 1st October 2015 to 31st September 2016.

In the booking dataset, for each booking record, we have booking platform, insert date (the date a consumer place the order), check-in date (planed check-in date), and other information such as length of stay and price. Importantly, for each booking record, we have booking status which indicates if this booking was eventually canceled by the customer. In addition, we got price and occupancy information of a set of comparable competitors, from 1st October 2015 to 31st September 2016, from a marketing research company.

To investigate the effect of online review on consumer post purchase attitude change, specifically in this study, cancellation behavior, we collected online reviews from the three OTAs (Agoda.com, Booking.com, and Expedia.com) and TripAdvisor.com. Initially, TripAdvisor.com is a third-party review platform for hotel and restaurants. Although TripAdvisor.com launched Instant Booking in June 2014, which allows consumers to directly book hotels on TripAdvisor.com, we didn’t see booking records of the focal hotel from TripAdvisor.com within the one year scope. We included online reviews from TripAdvisor.com because it is a major travel website that travelers are likely to check hotel information and reviews to help with decision making. We collected all the online reviews, including numeric rating and review text, of focal hotel and a director competitor from the four websites (Agoda.com, Booking.com, Expedia.com, and

TripAdvisor.com) posted on or before Sep 26th 2017. The direct competitor has a same star rating as the focal hotel and just locates adjacent to the focal hotel.

Topic Modeling and Coding

In order to investigate how the hedonic and utilitarian comments affect consumer attitude change, we classified review topics in two steps. In the first step, we used topic modeling to identify topics of collected online review contents. Topic modeling is a text mining technique that applies statistical methods to discover the topics that appear in a collection of documents. We utilized SAS® Enterprise Miner to generate topics for review content from the four websites, respectively. We didn't combine reviews from the four websites when conduct topic modeling because of the potential platform differences. We identified twenty topics for each website. Each discovered topic is represented by the most frequent words. In topic modeling, the number of topic k is predetermined. We decided to let k = 20 after checking topic modeling results with k = 5, 10, 15, 20, and 25. The selection criteria are that 1) the identified topics are understandable, 2) there are no obvious duplicate topics. After excluding topics with non-English words, finally we got twenty topics for Agoda.com, twenty topics for Booking.com, eighteen topics for Expedia.com, and sixteen topics for TripAdvisor.com. Then we used SAS® Enterprise Miner to conduct topic modeling again, with predetermined topics that are the identified final topics in the previous procedure, to get loadings of each piece of review for each topic. A loading is the probability of the review content belongs to the corresponding topic, and this is a similar score which measures the fit of the content with a particular topic.

In the second step, we hired two PhD students as coders to independently code each topic as hedonic comments or utilitarian comments. Hedonic comments are based upon personal opinions, interpretations, emotions and judgment. Objective comments are statements of functional feature without subjective judgement. As the results of topic modeling only show key words of each topic, coders need to classify each topic into hedonic or utilitarian exclusively, basing on the key words. Then we compared the coding results of the two coders and calculated Cohen Kappa, which equals to 0.92. Disagreement was discussed and solved.

Variables

The dependent variable, *cancellation*, is if a booking is canceled by the customer. *Cancellation* equals to 1 if a booking is canceled by the customer, and 0 if a booking is not canceled. In the dataset, the percentage of cancellation among all bookings is about 24%.

Independent variables are review information from OTAs and TripAdvisor.com. In our dataset, each booking is from one of the three OTAs, Agoda.com, Booking.com and Expedia.com.

There are two sets of independent variables in this study. The first set is numeric rating related variables. We operationalized the variables as follows. Pre-purchase rating (*pre_purchase_rating*) is the average review ratings on the first page of OTAs (*pre_purchase_rating_P*) or TripAdvisor.com (*pre_purchase_rating_T*) that a consumer saw before he/she booked a room/rooms. We didn't use the historical average rating that is normally shown next to the hotel name on OTAs and TripAdvisor.com because the historical average rating didn't have significant variance within a one-year time period. Negative post-purchase rating (*post_purchase_neg*) is the number of negative ratings on the first page OTAs (*post_purchase_neg_P*) or TripAdvisor.com (*post_purchase_neg_T*) after a customer booked a room/rooms and before the check-in date. Negative post-purchase rating is used to test H1a. We interacted negative post-purchase rating and pre-purchase rating on OTAs and TripAdvisor.com to test H1b.

Post_purchase_rating is the average review ratings on the first page of website after a customer book a room/rooms and before the check-in date. For focal hotel, we have two variables to represent post-purchase rating, post-purchase rating on TripAdvisor (*post_purchase_rating_T*) and post-purchase rating on booking platform (*post_purchase_rating_P*).

The second set is review content related variables. Hedonic comment is the average number of hedonic topics in each piece of review on the first page of website, after a customer book a room/rooms and before the check-in date. For focal hotel, we have two variables to represent hedonic comment after booking and before check-in date, number of hedonic comment on TripAdvisor (*#hedonic_T*) and number of hedonic comment on booking platform

(*#hedonic_P*) after booking and before check-in date. Utilitarian comment is the average number of utilitarian topics in each piece of review on the first page of website, after a customer book a room/rooms and before the check-in date. For focal hotel, we have two variables to represent utilitarian comment after booking and before check-in date, number of utilitarian comment on TripAdvisor (*#utilitarian_T*) and number of utilitarian comment on booking platform (*#utilitarian_P*) after booking and before check-in date. In the topic modeling and coding procedure, we got loadings of each piece of review for each topic and classified each topic into hedonic or utilitarian topic. If review "A" has a loading larger than zero on topic 1 and topic 1 is classified as hedonic (utilitarian) comment, then we count 1 hedonic (utilitarian) topic for review "A". We repeated this procedure to count the total number of hedonic and utilitarian topics for review "A" across all N topics (N is 20 for Agoda.com, 20 for Booking.com, 18 for Expedia.com, and 16 for TripAdvisor.com). In the analysis, we interact *post_purchase_rating* and *#hedonic*/*#utilitarian* to test for H2 and H3.

All the independent variables are from the first page of all reviews. This is because consumers can see the reviews on the first page without additional operation (click).

In addition to online reviews, there are some other factors that may lead to cancellation behavior, for example, booking platform policy differences, consumer plan change, and room price change of focal hotel and competitor. In the analysis, we include proper control variables as follow. We use three dummy variables to control for booking platform differences. The three dummies are agoda.com, expedia.com, and booking.com. From booking record perspective, we control for focal hotel's room price, length of stay, and interval[2] between booking date and check-in date. Room price of focal hotel is the price a customer actually paid for his/her room booking. Length of stay is the planned number of nights at focal hotel. Length of stay and time interval between booking date and check-in date are used to control for consumer plan change possibility. Plan change is due to unforeseen circumstances which are random factors. The longer interval, the higher possibility in encounter unexpected situations that result in changing plans. Length of stay may affect cancellation in different ways. On the one hand, a longer stay indicates a well planned trip, which is less likely to be affected by unimportant unforeseen circumstances. On the other hand, a longer stay means high involvement with the hotel and any unsatisfied thing can lead to a greater impact on consumer experience, comparing to a shorter stay. To control for seasonal effect, we control the occupancy rate of focal hotel (De Cantis et al. 2011; Koenig and Bischoff 2004). For competitors, we control for competitor price and occupancy rate as well. Competitor price is the average price of a set of competitors on the same check-in date. Competitor occupancy rate is the average occupancy rate of a set of competitors on the same check-in date. The set of competitors are selected by the focal hotel as main competitors. For online reviews, in terms of pre-purchase rating, we control for the variance of rating of

[2] The variable is *interval* is recorded by month.

focal hotel on the first page of booking platform and TripAdvisor.com, respectively. We also control for the same set of variables for the director competitor. For post-purchase rating, we control for the variance of rating on the first page for both focal hotel and director competitor on booking platform and TripAdvisor.com, respectively. We also control for the number of hedonic and utilitarian comments as well as average rating on the first page of director competitor on booking platform and TripAdvisor.com, respectively.

In addition, there are also possible selection effects that might be influencing the results. Most bookings that allow for cancellations are more expensive than bookings that do not allow for cancellations. So if consumers decide to make an online booking that allows for cancellation, it suggests that they are catering for the possibility that they might want to cancel – either because they are unsure of their travel plans, or because they are uncertain about the hotel they have booked, and want to give themselves the options of cancelling. Hence, the group cancelling likely value flexibility over price. To address the selection effects, I include one more control variable price difference, the difference between each booking/reservation price and average price on the same check-in date. A positive price difference means the booking/reservation price is higher than average. A high positive price difference indicates a high probability of being a free cancellation booking.

Analysis Model

As the dependent variable, cancellation, is binary, we applied logit regression model to test the hypotheses. For each booking record *i*, we have the cancellation label $cancellation_i$ to indicate if the order is cancelled by customer. We tested the hypotheses on booking platform (represented by P in variable name) and TripAadvisor.com (represented by T in variable name) in one model.

$$Cancellation_i = \beta_0 + \beta_1 Pre\_purchase\_rating_i\_T + \beta_2 Post\_purchase\_neg_i\_T + \beta_3 Pre\_purchase\_rating_i\_T * Post\_purchase\_neg_i\_T + \beta_4 Pre\_purchase\_rating_i\_P + \beta_5 Post\_purchase\_neg_i\_P + \beta_6 Pre\_purchase\_rating_i\_P * Post\_purchase\_neg_i\_P + \beta_7 \#Hedonic\_T_i + \beta_8 \#Utilitarian\_T_i + \beta_9 Post\_purchase\_rating\_T_i + \beta_{10} \#Hedonic\_T_i * Post\_purchase\_rating\_T_i + \beta_{11} \#Utilitarian\_T_i * Post\_purchase\_rating\_T_i + \beta_{12} \#Hedonic\_P_i + \beta_{13} \#Utilitarian\_P_i + \beta_{14} Post\_purchase\_rating\_P_i + \beta_{15} \#Hedonic\_P_i * Post\_purchase\_rating\_P_i + \beta_{16} \#Utilitarian\_P_i * Post\_purchase\_rating\_P_i + \beta_{17} U_i + \varepsilon_i \quad (2.1)$$

where β represents parameters of the model, ε represents stochasticity across booking record, U_i is the vector of control variables.

5. Results and Discussion

Pre-test Results

We delivered 100 survey and got 97 valid responses. 75 of the 97 respondents have booked a hotel room from at least one of the three platforms discussed in this study. In the pre-test, we asked them to recall their last booking experience on OTAs and found 67 of the 97 respondents did check online review after they booked the room and before checking in. In addition, 22 of the 67 respondents cancelled the booking after checking the post-purchase review. The cancellation rate of the 97 respondents is 22.7%, which is very close to the cancellation rate, 24%, of our dataset in the main analysis. We also collected respondents' demographic information. 52 of the 97 respondents are female. The 97 respondents have a wide range of age, from 18 to above 55 and their education level distributes from high school to doctorate. Details of the pre-test results and details of demographics are shown in table 2.1 and table 2.2, respectively. In sum, 69.07% of valid respondents checked post-purchase online review in the last hotel booking. Among the 67 respondents who checked post-purchase review in the last booking, 32.8% (22/67) canceled the hotel booking thereafter. The results show that consumers do usually re-check reviews after purchasing and may change purchase decision because of the post-purchase reviews.

Table 2.1: Pre-test Results

	Number of respondents	**Percentage among all valid responses**
Valid responses	97	100.00%
Checked post-purchase review in last booking	67	69.07%
Canceled booking after checking post-purchase review	22	22.68%

Table 2.2: Demographics of Respondents

Gender		
Female	52	53.61%
Male	45	46.39%
Age		

18-24	16	16.49%
25-34	47	48.45%
35-44	23	23.71%
45-54	5	5.15%
above 55	6	6.19%
Education level		
High school degree or equivalent (e.g. GED)	8	8.25%
Some college, no degree	22	22.68%
Associate degree (e.g. AA, AS)	8	8.25%
Bachelor's degree (e.g. BA, BS)	45	46.39%
Master's degree (e.g. MA, MS, MEd)	12	12.37%
Professional degree (e.g. MD, DDS, DVM)	1	1.03%
Doctorate (e.g. PhD, EdD)	1	1.03%

Logistic Regression

We used a logistic regression to test H1 to H3. The regression results are shown in Table 2.3. We report the main effect model and full model in the table. As hypothesized in H1a, we find positive and significant relationship between negative post-purchase rating (*Post_purchase_neg_T*) and log-odds of *cancellation* (coeff: 0.343; p-value < 0.001) on TripAdvisor.com. When the number of negative post-purchase review increase, we see an increase in cancellation possibility, supporting hypothesis 1a. However, we didn't find a significant effect on booking platform (coeff: 0.363; p-value > 0.05). Therefore, H1a is partially supported. The interaction between pre-purchase rating and negative post-purchase rating (*Pre_purchase_rating_T* × *Post_purchase_neg_T*) is positively significant (coeff: 0.433; p-value < 0.05) on TripAdvisor.com. High pre-purchase rating strengthened the relationship between negative post-purchase rating and cancellation possibility. Figure 2.1 shows that when pre-purchase rating is high, more negative post-purchase rating has greater impact on cancellation possibility than that when pre-purchase rating is low. The result is in line with our argument that high pre-purchase rating and more negative post-purchase rating cause greater cognitive dissonance, which leads to higher possibility to cancel. However, such effect is not found on booking platform (coeff: -0.029; p-value > 0.05). Hypothesis 1b is also partially supported.

To test H2 and H3, we interact number of post-purchase hedonic/utilitarian comment and post-purchase rating on booking platform and TripAdvisor.com. We didn't hypothesize the main effect, number of post-purchase hedonic/utilitarian comment, because it's meaningless to look at post-purchase hedonic/utilitarian comment without considering the valence of the review containing these comments. The interaction between number of post-purchase hedonic comment and post-purchase rating on booking platform (*Post_purchase_rating_P* × *#Hedonic_P*) is positive and significant (coeff: 0.214; p-value < 0.001). Figure 2.2 shows this interaction effect. Post-purchase rating has a much stronger effect on cancellation when there is a large number of hedonic comment than that when number of hedonic comment is low. More post-purchase hedonic comment with high rating results in high cancellation possibility, supporting hypothesis 2. The interaction between number of post-purchase hedonic comment and post-purchase rating on TripAdvisor.com (*Post_purchase_rating_T* × *#Hedonic_T*) is also positive and significant (coeff:

0.040; p-value < 0.05). Therefore, hypothesis 2 is supported on both booking platform and TripAdvisor.com.

Hypothesis 3 is partially supported. The interaction between number of post-purchase utilitarian comment and post-purchase rating on booking platform (*Post_purchase_rating_P* × *#Utilitarian_P*) is negative and significant (coeff: -0.074; p-value < 0.05). Figure 2.3 shows this interaction effect. Post-purchase rating has a much stronger negative effect on cancellation when there is a large number of utilitarian comment than that when number of utilitarian comment is low. More post-purchase utilitarian comment with high rating results in low cancellation possibility, supporting hypothesis 3. The interaction between number of post-purchase utilitarian comment and post-purchase rating on TripAdvisor.com (*Post_purchase_rating_T* × *#Utilitarian_T*) is negative but not significant (coeff: -0.052; p-value > 0.05). Hypothesis 3 is not supported on TripAdvisor.com.

Table 2.3: Logistic Regression Result

DV: Cancellation		
	Main effect model	**Full model**
Variables	**Coefficients (Std. Err.)**	**Coefficients (Std. Err.)**
Pre_purchase_rating_P	0.186 (0.124)	0.197 (0.135)
Post_purchase_neg_P	0.099 (0.056)	**0.363 (0.526)**
Pre_purchase_rating_P × Post_purchase_neg_P		**-0.029 (0.063)**
Pre_purchase_rating_T	0.044 (0.051)	0.054 (0.052)
Post_purchase_neg_T	0.383*** (0.062)	**0.343*** (0.064)**
Pre_purchase_rating_T × Post_purchase_neg_T		**0.433* (0.193)**
Post_purchase_rating_P	-0.034 (0.069)	-0.116 (0.092)

#Hedonic_P	0.126 (0.069)	-1.621*** (0.469)
Post_purchase_rating_P × #Hedonic_P		**0.214*** (0.056)**
#Utilitarian_P	-0.020 (0.047)	0.589 (0.319)
Post_purchase_rating_P × #Utilitarian_P		**-0.074* (0.036)**
Post_purchase_rating_T	0.203*** (0.057)	0.020 (0.099)
#Hedonic_T	0.033 (0.028)	-0.285* (0.126)
Post_purchase_rating_T × #Hedonic_T		**0.040* (0.015)**
#Utilitarian_T	-0.164 (0.101)	0.278 (0.493)
Post_purchase_rating_T × #Utilitarian_T		**-0.052 (0.060)**
agoda	-0.156 (0.190)	0.142 (0.206)
booking	0.294 (0.225)	0.695** (0.256)
Price difference	0.002 (0.002)	0.002 (0.002)
Length of stay	0.125*** (0.016)	0.125*** (0.016)
Interval	0.002*** (0.000)	0.002*** (0.000)
Price	0.004 (0.002)	0.005 (0.002)
Price_competitor	0.003 (0.003)	0.003 (0.003)
Occupancy rate	0.003 (0.005)	0.004 (0.005)
Occupancy rate_competitor	0.016 (0.008)	0.012 (0.009)
Pre_purchase_rating_1st_page_P_competitor	-0.232* (0.102)	-0.198 (0.104)
Pre_purchase_rating_1st_page_T_competitor	-0.014 (0.057)	-0.053 (0.058)
Pre_purchase_var_1st_page_P	0.019 (0.050)	-0.008 (0.050)
Pre_purchase_var_1st_page_P_competitor	-0.234*** (0.054)	-0.207*** (0.054)
Pre_purchase_ var _1st_page_T	0.123*** (0.034)	0.128*** (0.035)
Pre_purchase_ var _1st_page_T_competitor	-0.003 (0.018)	-0.009 (0.018)
Post_purchase_var _P	0.030 (0.042)	-0.001 (0.045)
Post_purchase_rating_P_competitor	0.091 (0.091)	-0.092 (0.121)
#Hedonic_P_competitor	-0.027 (0.071)	-0.746 (0.491)
Post_purchase_rating_P_competitor × #Hedonic_P_competitor		0.090 (0.056)
#Utilitarian_P_competitor	0.067 (0.044)	-0.208 (0.298)

Post_purchase_rating_P_competitor × #Utilitarian_P_competitor		0.030 (0.033)
Post_purchase_var_P_competitor	0.130** (0.046)	0.220*** (0.050)
Post_purchase_neg_P_competitor	-0.189** (0.064)	-0.243*** (0.066)
Post_purchase_var _T	-0.113** (0.034)	-0.087* (0.036)
Post_purchase_rating_T_competitor	-0.004 (0.061)	0.050 (0.151)
#Hedonic_T_competitor	-0.021 (0.032)	0.286 (0.170)
Post_purchase_rating_T_competitor × #Hedonic_T_competitor		-0.036 (0.020)
#Utilitarian_T_competitor	-0.227 (0.119)	-1.602* (0.773)
Post_purchase_rating_T_competitor × #Utilitarian_T_competitor		0.156 (0.090)
Post_purchase_var_T_competitor	-0.062** (0.022)	-0.041 (0.024)
Post_purchase_neg_T_competitor	0.162* (0.070)	0.132 (0.074)
Constant	-6.548** (2.060)	-3.739 (2.681)
Note: * *p*<0.05; ** *p*<0.01; *** *p*<0.001		

Figure 2.1: The Interaction Effect of Pre-purchase Rating and Negative Post-purchase Rating

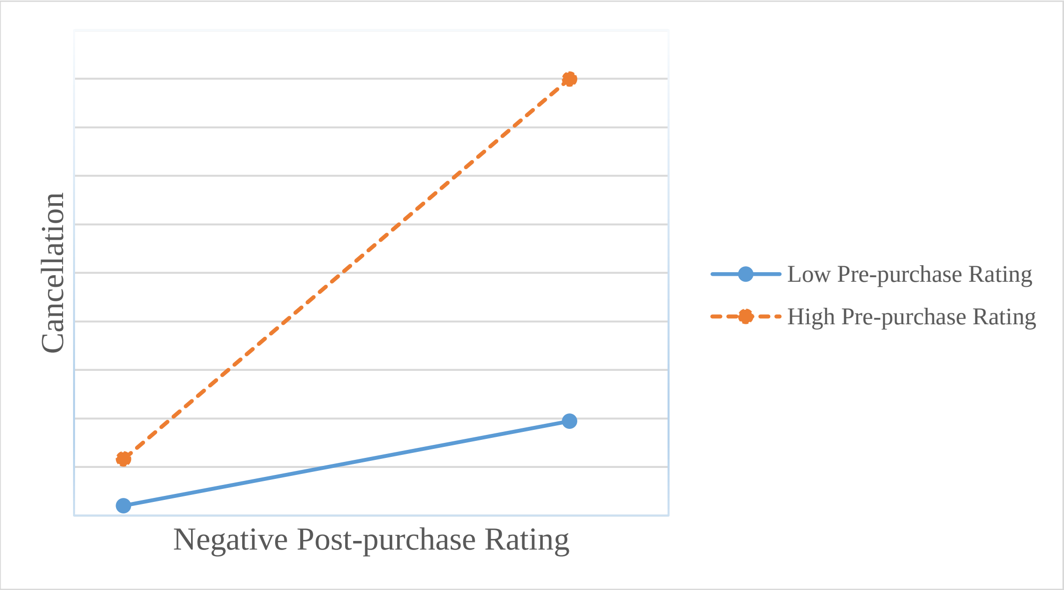
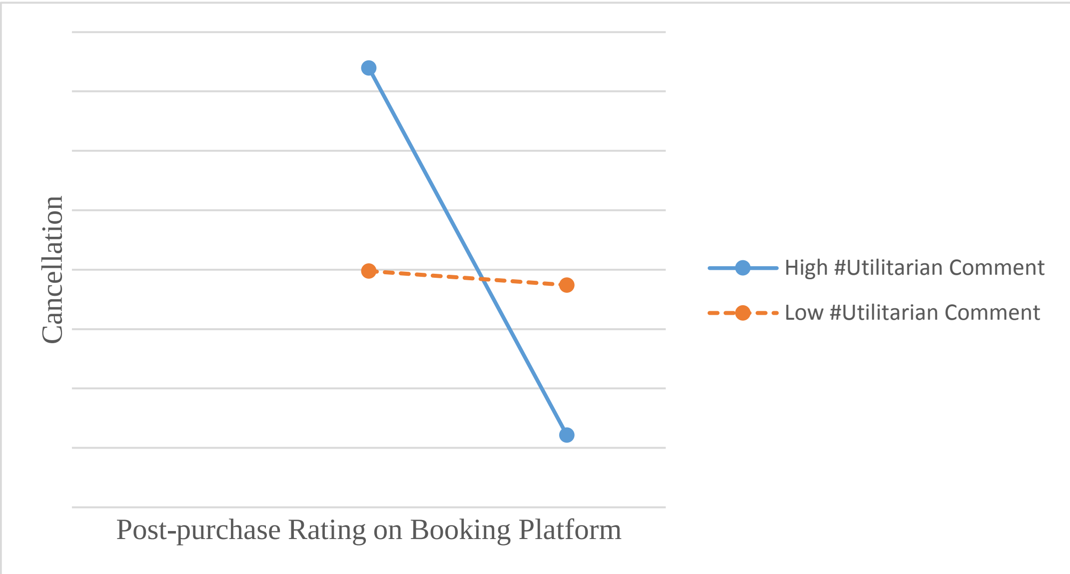


Figure 2.2: The Interaction Effect of Post-purchase Rating and Number of Hedonic Comment

Figure 2.3: The Interaction Effect of Post-purchase Rating and Number of Utilitarian Comment



We can get some insights from the results. First of all, the pre-test results show that consumers do look for updated information to confirm their purchase decision. So we should not overlook the power of post-purchase review.

Secondly, there is difference in trust consumers place on online review. Consumers are more trustful towards utilitarian comment compared to hedonic comment. Very good hedonic comment seems not trustworthy and leads to more cancellation, while very good utilitarian comment results in less cancellation.

6. Robustness check

In case of situations that consumers only check hotel reviews on either booking platform or TripAdvisor.com after booking a room, we test our hypotheses again separately on booking platform and TripAdvisor.com. Equation 2.2 tests hypotheses on booking platform and equation 2.3 tests hypotheses on TripAdvisor.com.

$$Cancellation_i = \beta_0 + \beta_1 Pre\_purchase\_rating\_P_i + \beta_2 Post\_purchase\_neg\_P_i + \beta_3 Pre\_purchase\_rating\_P_i * Post\_purchase\_neg\_P_i + \beta_4 \#Hedonic\_P_i + \beta_5 \#Utilitarian\_P_i + \beta_6 Post\_purchase\_rating\_P_i + \beta_7 \#Hedonic\_P_i * Post\_purchase\_rating\_P_i + \beta_8 \#Utilitarian\_P_i * Post\_purchase\_rating\_P_i + \beta_9 V_i + \varepsilon_i \quad (2.2)$$

where β represents parameters of the model, ε represents stochasticity across booking record, V_i is the vector of control variables.

$$Cancellation_i = \beta_0 + \beta_1 Pre\_purchase\_rating\_T_i + \beta_2 Post\_purchase\_neg\_T_i + \beta_3 Pre\_purchase\_rating\_T_i * Post\_purchase\_neg\_T_i + \beta_4 \#Hedonic\_T_i + \beta_5 \#Utilitarian\_T_i +$$

$\beta_6 Post\_purchase\_rating\_T_i + \beta_7 \#Hedonic\_T_i *$
$Post\_purchase\_rating\_T_i + \beta_8 \#Utilitarian\_T_i *$
$Post\_purchase\_rating\_T_i + \beta_9 W_i + \varepsilon_i$ (2.3)

where β represents parameters of the model, ε represents stochasticity across booking record, W_i is the vector of control variables.

The results of robustness check are shown in table 2.4 and table 2.5. Table 2.4 is the results for booking platform and table 2.5 is for TripAdvisor.com. The results are consistent with the results of main analysis. On booking platform, the relationship between negative post-purchase rating and cancellation is positive but not significant (coeff: 0.688; p-value > 0.05). The interaction of negative post-purchase rating and pre-purchase rating is not significant, neither (coeff: -0.060; p-value > 0.05). Therefore, H1a and H1b are not supported on booking platform. H2 and H3 are supported on booking platform. The interaction of number of hedonic comment and post-purchase rating is positively significant (coeff: 0.231; p-value < 0.001), while the interaction of number of utilitarian comment and post-purchase rating is negatively significant (coeff: -0.088; p-value < 0.01). In the TripAdvisor.com only model, the relationship between negative post-purchase rating and cancellation is positively significant (coeff: 0.364; p-value < 0.001), supporting H1a. The interaction of negative post-purchase rating and pre-purchase rating also positive and significant (coeff: 0.433; p-value < 0.05), supporting H1b. The interaction of number of hedonic comment and post-purchase rating on TripAdvisor.com is positively significant (coeff:

0.030; p-value < 0.05), supporting H2. However, the interaction of number of utilitarian comment and post-purchase rating is not significant (coeff: -0.058; p-value > 0.05). H3 is not supported on TripAdvisor.com.

In sum, the results of booking platform only model and TripAdvisor.com only model are in line with the main results with both platforms in one model. H1a and H1b are partially supported on TripAdvisor.com. H2 is supported on both platform and H3 is partially supported on booking platform.

Table 2.4: Results of Booking Platform Only Model

DV: Cancellation		
	Main effect model	**Full model**
Variables	**Coefficients (Std. Err.)**	**Coefficients (Std. Err.)**
Pre_purchase_rating_P	0.275* (0.119)	0.298* (0.128)
Post_purchase_neg_P	0.165** (0.052)	**0.688 (0.486)**
Pre_purchase_rating_P × Post_purchase_neg_P		**-0.060 (0.058)**
Post_purchase_rating_P	0.078 (0.062)	-0.005 (0.080)
#Hedonic_P	0.074 (0.062)	-1.816*** (0.436)
Post_purchase_rating_P × #Hedonic_P		**0.231*** (0.052)**
#Utilitarian_P	0.005 (0.042)	0.745** (0.285)
Post_purchase_rating_P × #Utilitarian_P		**-0.088** (0.032)**
agoda	-0.110 (0.182)	0.144 (0.193)
booking	0.041 (0.207)	0.400 (0.228)
Price difference	0.001 (0.002)	0.001 (0.002)
Length of stay	0.139*** (0.016)	0.138*** (0.016)
Interval	0.002*** (0.000)	0.002*** (0.000)
Price	0.005* (0.002)	0.006* (0.002)
Price_competitor	0.003 (0.003)	0.003 (0.003)
Occupancy rate	0.001 (0.005)	0.002 (0.005)
Occupancy rate_competitor	0.022** (0.008)	0.020** (0.008)

Pre_purchase_rating_1st_page_P_competitor	-0.217* (0.096)	-0.193* (0.097)
Pre_purchase_var_1st_page_P	0.015 (0.048)	-0.004 (0.048)
Pre_purchase_var_1st_page_P_competitor	-0.192*** (0.050)	-0.178*** (0.050)
Post_purchase_var _P	0.041 (0.038)	0.005 (0.041)
Post_purchase_rating_P_competitor	0.188* (0.080)	-0.034 (0.106)
#Hedonic_P_competitor	-0.071 (0.063)	-0.914* (0.444)
Post_purchase_rating_P_competitor × #Hedonic_P_competitor		0.100* (0.050)
#Utilitarian_P_competitor	0.046 (0.037)	-0.123 (0.248)
Post_purchase_rating_P_competitor × #Utilitarian_P_competitor		0.020 (0.028)
Post_purchase_var_P_competitor	0.126*** (0.038)	0.191*** (0.041)
Post_purchase_neg_P_competitor	-0.092 (0.057)	-0.120* (0.058)
Constant	-7.974*** (1.875)	-6.152** (2.016)
Note: * *p*<0.05; ** *p*<0.01; *** *p*<0.001		

DV: Cancellation		
	Main effect model	**Full model**
Variables	**Coefficients (Std. Err.)**	**Coefficients (Std. Err.)**
Pre_purchase_rating_P	0.275* (0.119)	0.298* (0.128)
Post_purchase_neg_P	0.165** (0.052)	**0.688 (0.486)**
Pre_purchase_rating_P × Post_purchase_neg_P		**-0.060 (0.058)**
Post_purchase_rating_P	0.078 (0.062)	-0.005 (0.080)
#Hedonic_P	0.074 (0.062)	-1.816*** (0.436)
Post_purchase_rating_P × #Hedonic_P		**0.231*** (0.052)**
#Utilitarian_P	0.005 (0.042)	0.745** (0.285)
Post_purchase_rating_P × #Utilitarian_P		**-0.088** (0.032)**
agoda	-0.110 (0.182)	0.144 (0.193)
booking	0.041 (0.207)	0.400 (0.228)
Price difference	0.001 (0.002)	0.001 (0.002)
Length of stay	0.139*** (0.016)	0.138*** (0.016)
Interval	0.002*** (0.000)	0.002*** (0.000)

Price	0.005* (0.002)	0.006* (0.002)
Price_competitor	0.003 (0.003)	0.003 (0.003)
Occupancy rate	0.001 (0.005)	0.002 (0.005)
Occupancy rate_competitor	0.022** (0.008)	0.020** (0.008)
Pre_purchase_rating_1st_page_P_competitor	-0.217* (0.096)	-0.193* (0.097)
Pre_purchase_var_1st_page_P	0.015 (0.048)	-0.004 (0.048)
Pre_purchase_var_1st_page_P_competitor	-0.192*** (0.050)	-0.178*** (0.050)
Post_purchase_var _P	0.041 (0.038)	0.005 (0.041)
Post_purchase_rating_P_competitor	0.188* (0.080)	-0.034 (0.106)
#Hedonic_P_competitor	-0.071 (0.063)	-0.914* (0.444)
Post_purchase_rating_P_competitor × #Hedonic_P_competitor		0.100* (0.050)
#Utilitarian_P_competitor	0.046 (0.037)	-0.123 (0.248)
Post_purchase_rating_P_competitor × #Utilitarian_P_competitor		0.020 (0.028)
Post_purchase_var_P_competitor	0.126*** (0.038)	0.191*** (0.041)
Post_purchase_neg_P_competitor	-0.092 (0.057)	-0.120* (0.058)
Constant	-7.974*** (1.875)	-6.152** (2.016)
Note: * *p*<0.05; ** *p*<0.01; *** *p*<0.001		

Table 2.5: Results of TripAdvisor.com Only Model

DV: Cancellation		
	Main effect model	**Full model**
Variables	**Coefficients (Std. Err.)**	**Coefficients (Std. Err.)**
Pre_purchase_rating_T	0.039 (0.049)	0.039 (0.050)
Post_purchase_neg_T	0.393*** (0.057)	**0.364*** (0.058)**
Pre_purchase_rating_T × Post_purchase_neg_T		**0.433* (0.183)**
Post_purchase_rating_T	0.264*** (0.050)	0.133 (0.090)
#Hedonic_T	0.041 (0.025)	-0.201 (0.115)
Post_purchase_rating_T × #Hedonic_T		**0.030* (0.014)**
#Utilitarian_T	-0.144 (0.092)	0.363 (0.450)
Post_purchase_rating_T		**-0.058 (0.054)**

× #Utilitarian_T		
agoda	-0.011 (0.103)	-0.016 (0.104)
booking	0.484*** (0.096)	0.494*** (0.096)
Price difference	0.000 (0.002)	-0.000 (0.002)
Length of stay	0.131*** (0.016)	0.132*** (0.016)
Interval	0.002*** (0.000)	0.002*** (0.000)
Price	0.006** (0.002)	0.006** (0.002)
Price_competitor	0.001 (0.003)	0.001 (0.003)
Occupancy rate	-0.000 (0.005)	0.001 (0.005)
Occupancy rate_competitor	0.021** (0.008)	0.018* (0.008)
Pre_purchase_rating_1st_page_T_competitor	-0.069 (0.054)	-0.085 (0.054)
Pre_purchase_ var _1st_page_T	0.126*** (0.032)	0.132*** (0.032)
Pre_purchase_ var_1st_page_T_competitor	-0.022 (0.017)	-0.024 (0.017)
Post_purchase_var _T	-0.057 (0.031)	-0.039 (0.032)
Post_purchase_rating_T_competitor	-0.022 (0.056)	0.000 (0.136)
#Hedonic_T_competitor	-0.032 (0.030)	0.320* (0.157)
Post_purchase_rating_T_competitor × #Hedonic_T_competitor		-0.041* (0.018)
#Utilitarian_T_competitor	-0.146 (0.107)	-1.803* (0.714)
Post_purchase_rating_T_competitor × #Utilitarian_T_competitor		0.189* (0.082)
Post_purchase_var_T_competitor	-0.065** (0.021)	-0.048* (0.022)
Post_purchase_neg_T_competitor	0.092 (0.064)	0.058 (0.068)
Constant	-6.702*** (0.952)	-5.527*** (1.633)
Note: * *p*<0.05; ** *p*<0.01; *** *p*<0.001		

7. Contribution

This study adds knowledge to consumer online review literature. Although plenty of research has examined the effect of online review on consumer decisions from various perspectives, such as sales, purchase intention, brand evaluation, and perceived helpfulness of online review (Baker et al. 2016; Chevalier and Mayzlin 2006; Schlosser 2011; Tang et al. 2014), it is not clear how consumers response to post-purchase online reviews. For some business formats, purchase stage and getting product on hands are separated. For example, a customer may pre-order a product online and will only receive the product two weeks later, or a customer may book a hotel one month in advance and will complete the transaction only when the customer check in at the hotel one month later. Consumers can choose to cancel an ordered product or cancel a booked service. Therefore, we can't overlook the effect of post-purchase online review on consumer attitude change. This study contribute to consumer online review literature by investigating how post-purchase online review affect consumer cancellation behavior. This study shows that consumers do look for information to confirm that they make the right decision. In addition, we find negativity bias in the effect of post-purchase online review on cancellation behavior. The effect of negative information is stronger than positive information and consumer are very likely to feel regret and change purchase decision when the pre-purchase rating is high and they see negative information after purchasing. We also classify the review text into hedonic and utilitarian perspectives of product/service. Online review consists of numeric information and textual information. However, study on textual information of online review is limited, not to say examining the effect of textual information on consumer post-purchase attitude change. In this study, we find the textual information and numeric rating interactively affect cancellation behavior. Consumers show different trust levels on hedonic and utilitarian comment. They distrust very good feeling based hedonic comment but trust functional based utilitarian comment.

This study also contribute to consumer regret literature. Literature on consumer regret typically examined the antecedents and consequence of consumer regret and the examined consumer regret happens after they personally used the product (Abendroth and Diehl 2006; Bui et al. 2011; Inman and Zeelenberg 2002; Lee and Cotte 2009; Tsiros and Mittal 2000). However, another type of consumer regret is that consumer can cancel a product order or service booking even before they use it. Online review provides consumers opportunity to easily judge product/service quality before they use the product or consume the service. Nevertheless, there is lack of literature to examine this type of consumer regret. This study fills this gap by studying how the post-purchase online review can lead to consumer regret.

This study also have important practical implication. We show that consumers do look for updated online review after they make purchase decision and the post-purchase review has significant impact on consumer attitude change. Marketers should pay attention to the strong power of online review. Especially for those whose business models give consumer opportunity to change the purchase decision before they use the product or consume the service. Negative post-purchase review hurts a lot. This study is not telling marketers to manipulate online reviews, but telling them to be careful with negative reviews. Improving the product/service quality is always a good measure to eliminate negative comments. Marketers should also be careful with fake negative reviews. In regard to the different effect of hedonic and utilitarian comment on consumer cancellations, marketers may encourage experienced customers to provide more positive reviews on utilitarian features of their product/service. A good understanding about the mechanism that consumer online review works on influencing consumer decisions makes marketers better prepared to potential problems.

8. Limitation and Future Research

The first limitation of this study is that we only have one-year hotel booking records. Across only one year, the variance of average hotel ratings that is normally shown next to a hotel name is not significant or even unobservable. We are not able to see the effect of the hotel average rating on consumer post-purchase behavior. In this study, we only use the average ratings on the first page of online review before purchasing as the pre-purchase rating because firstly the variance is larger and observable, secondly most of the consumers would see reviews on the first page, and thirdly most recent reviews reflect the hotel's latest quality and have more effect on consumer decision-making.

The second limitation is that we only have booking records data from one hotel. Therefore, the features of this hotel may limit the generalization of the results. The focal hotel in this study is a decent 4-star hotel in a developed country. Comparing to budget hotel, customers of this hotel is not very price sensitive and emphasize more on hotel facility and service. Therefore, we can see significant effect of post-purchase online review on consumer attitude change. The results of this study can be generalized to products or service whose customers are not purely price driven.

The third limitation is that we are not able to fully control for price change and better choices in terms of price. We use the average price of the competitors on the same check-in date to control for the alternatives. We also use the price difference between each booking price and the average room price on the same check-in date as a control variable. The price difference can serve as an indicator of price premium. We understand the controls are not the best, however, price change (increase or decrease) information of either the focal hotel or the competitors is not available.

This research examined the effect of online reviews on booking platform and TripAdvisor.com on consumer cancellation behavior. Future research can extend this study by exploring the effect of post-purchase online review on off-line consumer decisions. Consumers who order product or booking service off-line may still check online reviews. It would be interesting to see how the online review can affect off-line consumer decision and if the pattern is different from the effect of online review on online purchasing.

9. Conclusion

In this study we examine the effect of post-purchase online review on consumer post-purchase attitude change. Our pre-test results show that consumers do look for online reviews after purchasing. Furthermore, we explore how the numeric rating and textual information interactively affect consumer cancellation behavior. Using a proprietary hotel booking dataset and public consumer online review, we find that number of negative post-purchase review is positively related to cancellation possibility and this relationship is strengthened by high pre-purchase rating. In addition, number of hedonic/utilitarian comment and post-purchase rating interactively affect cancellation behavior. We find consumers distrust very good, feeling based hedonic comment, but trust very good, functional based utilitarian comment. This study contributes to consumer online review literature and consumer regret literature.

References

Abendroth, L.J., and Diehl, K. 2006. "Now or Never: Effects of Limited Purchase Opportunities on Patterns of Regret over Time," *Journal of Consumer Research* (33:3), pp 342-351.

Akerlof, G.A., and Dickens, W.T. 1982. "The Economic Consequences of Cognitive Dissonance," *The American economic review* (72:3), pp 307-319.

Archak, N., Ghose, A., and Ipeirotis, P.G. 2011. "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," *Management science* (57:8), pp 1485-1509.

Babić Rosario, A., Sotgiu, F., De Valck, K., and Bijmolt, T.H. 2016. "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors," *Journal of Marketing Research* (53:3), pp 297-318.

Baker, A.M., Donthu, N., and Kumar, V. 2016. "Investigating How Word-of-Mouth Conversations About Brands Influence Purchase and Retransmission Intentions," *Journal of Marketing Research* (53:2), pp 225-239.

Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. 2012. "The Role of Social Networks in Information Diffusion," *Proceedings of the 21st international conference on World Wide Web*: ACM, pp. 519-528.

Basuroy, S., Chatterjee, S., and Ravid, S.A. 2003. "How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets," *Journal of marketing* (67:4), pp 103-117.

Batra, R., and Ahtola, O.T. 1991. "Measuring the Hedonic and Utilitarian Sources of Consumer Attitudes," *Marketing letters* (2:2), pp 159-170.

Bauer, S., Noulas, A., Séaghdha, D.O., Clark, S., and Mascolo, C. 2012. "Talking Places: Modelling and Analysing Linguistic Content in Foursquare," *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*: IEEE, pp. 348-357.

Baumgartner, H., and Steenkamp, J.-B.E. 1996. "Exploratory Consumer Buying Behavior: Conceptualization and Measurement," *International Journal of Research in Marketing* (13:2), pp 121-137.

Blei, D.M., Ng, A.Y., and Jordan, M.I. 2003. "Latent Dirichlet Allocation," *Journal of machine Learning research* (3:Jan), pp 993-1022.

Broadbent, D. 1958. *Perception and Communication*.

Brodersen, A., Scellato, S., and Wattenhofer, M. 2012. "Youtube around the World: Geographic Popularity of Videos," *Proceedings of the 21st international conference on World Wide Web*: ACM, pp. 241-250.

Brown, J.S. 2000. "Growing Up: Digital: How the Web Changes Work, Education, and the Ways People Learn," *Change: The Magazine of Higher Learning* (32:2), pp 11-20.

Bui, M., Krishen, A.S., and Bates, K. 2011. "Modeling Regret Effects on Consumer Post-Purchase Decisions," *European Journal of Marketing* (45:7/8), pp 1068-1090.

Cantallops, A.S., and Salvi, F. 2014. "New Consumer Behavior: A Review of Research on Ewom and Hotels," *International Journal of Hospitality Management* (36), pp 41-51.

Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, P.K. 2010. "Measuring User Influence in Twitter: The Million Follower Fallacy," *ICWSM* (10:10-17), p 30.

Chaffey, D. 2018. "Global Social Media Research Summary 2018." Retrieved 5th June, 2018, from https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

Chen, Y., Wang, Q., and Xie, J. 2011. "Online Social Interactions: A Natural Experiment on Word of Mouth Versus Observational Learning," *Journal of marketing research* (48:2), pp 238-254.

Chen, Z., and Lurie, N.H. 2013. "Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth," *Journal of Marketing Research* (50:4), pp 463-476.

Cheng, X., Dale, C., and Liu, J. 2008. "Statistics and Social Network of Youtube Videos," *Quality of Service, 2008. IWQoS 2008. 16th International Workshop on*, Enschede, Netherlands: IEEE, pp. 229-238.

Chevalier, J.A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of marketing research* (43:3), pp 345-354.

Daugherty, T., and Hoffman, E. 2014. "Ewom and the Importance of Capturing Consumer Attention within Social Media," *Journal of Marketing Communications* (20:1-2), pp 82-102.

Davenport, T.H., and Beck, J.C. 2001. *The Attention Economy: Understanding the New Currency of Business*. Harvard Business Press.

De Cantis, S., Ferrante, M., and Vaccina, F. 2011. "Seasonal Pattern and Amplitude–a Logical Framework to Analyse Seasonality in Tourism: An Application to Bed Occupancy in Sicilian Hotels," *Tourism Economics* (17:3), pp 655-675.

De Langhe, B., Fernbach, P.M., and Lichtenstein, D.R. 2015. "Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings," *Journal of Consumer Research* (42:6), pp 817-833.

De Vries, L., Gensler, S., and Leeflang, P.S. 2012. "Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing," *Journal of Interactive Marketing* (26:2), pp 83-91.

Deutsch, J.A., and Deutsch, D. 1963. "Attention: Some Theoretical Considerations," *Psychological review* (70:1), p 80.

Falkinger, J. 2008. "Limited Attention as a Scarce Resource in Information-Rich Economies," *The Economic Journal* (118:532), pp 1596-1620.

Festinger, L. 1962. *A Theory of Cognitive Dissonance*. Stanford university press.

Figueiredo, F., Benevenuto, F., and Almeida, J.M. 2011. "The Tube over Time: Characterizing Popularity Growth of Youtube Videos," *Proceedings of the fourth ACM international conference on Web search and data mining*: ACM, pp. 745-754.

Filieri, R. 2016. "What Makes an Online Consumer Review Trustworthy?," *Annals of Tourism Research* (58), pp 46-64.

Fiske, S.T. 2018. "Attention and Weight in Person Perception: The Impact of Negative and Extreme Behavior15," in: *Social Cognition*. Routledge, pp. 14-40.

Garcia, C., Ponsoda, V., and Estebaranz, H. 2000. "Scanning Ads: Effects of Involvement and of Position of the Illustration in Printed Advertisements," *ACR North American Advances*).

Goldhaber, M.H. 1997. "The Attention Economy and the Net," *First Monday* (2:4).

Grobe, R.P., Pettibone, T.J., and Martin, D.W. 1973. "Effects of Lecturer Pace on Noise Level in a University Classroom1," *The Journal of Educational Research* (67:2), pp 73-75.

Hirschman, E.C., and Holbrook, M.B. 1982. "Hedonic Consumption: Emerging Concepts, Methods and Propositions," *Journal of marketing* (46:3), pp 92-101.

Hoffman, E., and Daugherty, T. 2013. "Is a Picture Always Worth a Thousand Words? Attention to Structural Elements of Ewom for Consumer Brands within Social Media," *ACR North American Advances*).

Hong, L., and Davison, B.D. 2010. "Empirical Study of Topic Modeling in Twitter," *Proceedings of the first workshop on social media analytics*: ACM, pp. 80-88.

Hornik, I.F.a.K. 2015. "Tm: Text Mining Package."

Hornik, K., and Grün, B. 2011. "Topicmodels: An R Package for Fitting Topic Models," *Journal of Statistical Software* (40:13), pp 1-30.

Hua, X.-s., Lu, L., Zhang, H.-j., and District, H. 2005. "A Generic Framework of User Attention Model and Its Application in Video Summarization," *IEEE Transaction on multimedia* (7:5), pp 907-919.

Inman, J.J., and Zeelenberg, M. 2002. "Regret in Repeat Purchase Versus Switching Decisions: The Attenuating Role of Decision Justifiability," *Journal of consumer research* (29:1), pp 116-128.

InternetLiveStats. 2017. Retrieved May 3rd, 2017, from http://www.internetlivestats.com/one-second

Ishii, K., Reyes, J.A., and Kitayama, S. 2003. "Spontaneous Attention to Word Content Versus Emotional Tone: Differences among Three Cultures," *Psychological science* (14:1), pp 39-46.

Joachimsthaler, E.A., and Lastovicka, J.L. 1984. "Optimal Stimulation Level—Exploratory Behavior Models," *Journal of Consumer Research* (11:3), pp 830-835.

Jones, D.M., and Macken, W.J. 1993. "Irrelevant Tones Produce an Irrelevant Speech Effect: Implications for Phonological Coding in Working Memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition* (19:2), p 369.

Jones, D.M., Miles, C., and Page, J. 1990. "Disruption of Proofreading by Irrelevant Speech: Effects of Attention, Arousal or Memory?," *Applied Cognitive Psychology* (4:2), pp 89-108.

Keller, J.M. 1987a. "Development and Use of the Arcs Model of Instructional Design," *Journal of instructional development* (10:3), p 2.

Keller, J.M. 1987b. "Strategies for Stimulating the Motivation to Learn," *Performance+ Instruction* (26:8), pp 1-7.

Koenig, N., and Bischoff, E.E. 2004. "Analyzing Seasonality in Welsh Room Occupancy Data," *Annals of Tourism Research* (31:2), pp 374-392.

Kormos, J., and Dénes, M. 2004. "Exploring Measures and Perceptions of Fluency in the Speech of Second Language Learners," *System* (32:2), pp 145-164.

Kwark, Y., Lee, G.M., Pavlou, P., and Qiu, L. 2016. "The Spillover Effects of User-Generated Online Product Reviews on Purchases: Evidence from Clickstream Data,").

Lee, S.H., and Cotte, J. 2009. "Post-Purchase Consumer Regret: Conceptualization and Development of the Ppcr Scale," *ACR North American Advances*).

Leuba, C. 1955. "Toward Some Integration of Learning Theories: The Concept of Optimal Stimulation," *Psychological Reports*).

Liu, Y. 2006. "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of marketing* (70:3), pp 74-89.

Liu, Z. 2005. "Reading Behavior in the Digital Environment: Changes in Reading Behavior over the Past Ten Years," *Journal of documentation* (61:6), pp 700-712.

Lu, Q., Ye, Q., and Law, R. 2014. "Moderating Effects of Product Heterogeneity between Online Word-of-Mouth and Hotel Sales," *Journal of Electronic Commerce Research* (15:1), p 1.

Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E.C., Wetzels, M., and Pfann, G. 2013. "More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates," *Journal of Marketing* (77:1), pp 87-103.

Luo, C., Luo, X.R., Xu, Y., Warkentin, M., and Sia, C.L. 2015. "Examining the Moderating Role of Sense of Membership in Online Review Evaluations," *Information & Management* (52:3), pp 305-316.

Mattys, S.L., Davis, M.H., Bradlow, A.R., and Scott, S.K. 2012. "Speech Recognition in Adverse Conditions: A Review," *Language and Cognitive Processes* (27:7-8), pp 953-978.

Menon, R.V., Sigurdsson, V., Larsen, N.M., Fagerstrøm, A., and Foxall, G.R. 2016. "Consumer Attention to Price in Social Commerce: Eye Tracking Patterns in Retail Clothing," *Journal of Business Research* (69:11), pp 5008-5013.

Mou, J., and Shin, D. 2018. "Effects of Social Popularity and Time Scarcity on Online Consumer Behaviour Regarding Smart Healthcare Products: An Eye-Tracking Approach," *Computers in Human Behavior* (78), pp 74-89.

Muñoz-Leiva, F., Hernández-Méndez, J., and Gómez-Carmona, D. 2018. "Measuring Advertising Effectiveness in Travel 2.0 Websites through Eye-Tracking Technology," *Physiology & behavior*).

Navalpakkam, V., Kumar, R., Li, L., and Sivakumar, D. 2012. "Attention and Selection in Online Choice Tasks," *International conference on user modeling, adaptation, and personalization*: Springer, pp. 200-211.

Norman, D.A. 1968. "Toward a Theory of Memory and Attention," *Psychological review* (75:6), p 522.

Park, D.-H., and Lee, J. 2008. "Ewom Overload and Its Effect on Consumer Behavioral Intention Depending on Consumer Involvement," *Electronic Commerce Research and Applications* (7:4), pp 386-398.

Pedrycz, W., and Chen, S.-M. 2013. *Social Networks: A Framework of Computational Intelligence*. Springer International Publishing.

Peeters, G., and Czapinski, J. 1990. "Positive-Negative Asymmetry in Evaluations: The Distinction between Affective and Informational Negativity Effects," *European review of social psychology* (1:1), pp 33-60.

Pieters, R., and Wedel, M. 2004. "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects," *Journal of Marketing* (68:2), pp 36-50.

Pinto, H., Almeida, J.M., and Gonçalves, M.A. 2013. "Using Early View Patterns to Predict the Popularity of Youtube Videos," *Proceedings of the sixth ACM international conference on Web search and data mining*: ACM, pp. 365-374.

Raju, P.S. 1980. "Optimum Stimulation Level: Its Relationship to Personality, Demographics, and Exploratory Behavior," *Journal of consumer research* (7:3), pp 272-282.

Rhoades, S.A. 1993. "The Herfindahl-Hirschman Index," *Fed. Res. Bull.* (79), p 188.

Rozin, P., and Royzman, E.B. 2001. "Negativity Bias, Negativity Dominance, and Contagion," *Personality and social psychology review* (5:4), pp 296-320.

RStudio. 2015. "Rstudio: Integrated Development for R." Boston, MA: RStudio, Inc.

Ryu, K., Han, H., and Jang, S. 2010. "Relationships among Hedonic and Utilitarian Values, Satisfaction and Behavioral Intentions in the Fast-Casual Restaurant Industry," *International Journal of Contemporary Hospitality Management* (22:3), pp 416-432.

Schellekens, G.A., Verlegh, P.W., and Smidts, A. 2010. "Language Abstraction in Word of Mouth," *Journal of Consumer Research* (37:2), pp 207-223.

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), pp 226-239.

Schmiedel, T., Müller, O., and vom Brocke, J. 2018. "Topic Modeling as a Strategy of Inquiry in Organizational Research: A Tutorial with an Application Example on Organizational Culture," *Organizational Research Methods*), p 1094428118773858.

Schuckert, M., Liu, X., and Law, R. 2015. "Hospitality and Tourism Online Reviews: Recent Trends and Future Directions," *Journal of Travel & Tourism Marketing* (32:5), pp 608-621.

Sevilla, J., Lu, J., and Kahn, B.E. 2019. "Variety Seeking, Satiation, and Maximizing Enjoyment over Time," *Journal of Consumer Psychology* (29:1), pp 89-103.

Simon, H.A. 1971. "Designing Organizations for an Information-Rich World," in: *Computers, Communication, and the Public Interest, Baltimore,* M. Greenberger (ed.). The Johns Hopkins Press, pp. 40-41.

Sparks, B.A., and Browning, V. 2011. "The Impact of Online Reviews on Hotel Booking Intentions and Perception of Trust," *Tourism management* (32:6), pp 1310-1323.

Steenkamp, J.-B.E., and Baumgartner, H. 1992. "The Role of Optimum Stimulation Level in Exploratory Consumer Behavior," *Journal of consumer research* (19:3), pp 434-448.

Sun, M. 2012. "How Does the Variance of Product Ratings Matter?," *Management Science* (58:4), pp 696-707.

Susarla, A., Oh, J.-H., and Tan, Y. 2012. "Social Networks and the Diffusion of User-Generated Content: Evidence from Youtube," *Information Systems Research* (23:1), pp 23-41.

Szabo, G., and Huberman, B.A. 2010. "Predicting the Popularity of Online Content," *Communications of the ACM* (53:8), pp 80-88.

Tang, T., Fang, E., and Wang, F. 2014. "Is Neutral Really Neutral? The Effects of Neutral User-Generated Content on Product Sales," *Journal of Marketing* (78:4), pp 41-58.

Treisman, A.M. 1960. "Contextual Cues in Selective Listening," *Quarterly Journal of Experimental Psychology* (12:4), pp 242-248.

Treisman, A.M. 1964. "Selective Attention in Man," *British medical bulletin* (20:1), pp 12-&.

Tsiros, M., and Mittal, V. 2000. "Regret: A Model of Its Antecedents and Consequences in Consumer Decision Making," *Journal of Consumer Research* (26:4), pp 401-417.

Tversky, A., and Kahneman, D. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The quarterly journal of economics* (106:4), pp 1039-1061.

Voss, K.E., Spangenberg, E.R., and Grohmann, B. 2003. "Measuring the Hedonic and Utilitarian Dimensions of Consumer Attitude," *Journal of marketing research* (40:3), pp 310-320.

Wei, X., and Croft, W.B. 2006. "Lda-Based Document Models for Ad-Hoc Retrieval," *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*: ACM, pp. 178-185.

Weinreich, H., Obendorf, H., Herder, E., and Mayer, M. 2008. "Not Quite the Average: An Empirical Study of Web Use," *ACM Transactions on the Web (TWEB)* (2:1), p 5.

Weng, J., Lim, E.-P., Jiang, J., and He, Q. 2010. "Twitterrank: Finding Topic-Sensitive Influential Twitterers," *Proceedings of the third ACM international conference on Web search and data mining*: ACM, pp. 261-270.

Xu, C., and Zhang, Q. 2018. "The Dominant Factor of Social Tags for Users' Decision Behavior on E-Commerce Websites: Color or Text," *Journal of the Association for Information Science and Technology*).

Yang, K., and Lee, H.-J. 2010. "Gender Differences in Using Mobile Data Services: Utilitarian and Hedonic Value Approaches," *Journal of Research in Interactive Marketing* (4:2), pp 142-156.

Ye, Q., Law, R., and Gu, B. 2009. "The Impact of Online User Reviews on Hotel Room Sales," *International Journal of Hospitality Management* (28:1), pp 180-182.

Ye, Q., Law, R., Gu, B., and Chen, W. 2011. "The Influence of User-Generated Content on Traveler Behavior: An Empirical Investigation on the Effects of E-Word-of-Mouth to Hotel Online Bookings," *Computers in Human Behavior* (27:2), pp 634-639.

Yin, D., Bond, S.D., and Zhang, H. 2017. "Keep Your Cool or Let It Out: Nonlinear Effects of Expressed Arousal on Perceptions of Consumer Reviews," *Journal of Marketing Research* (54:3), pp 447-463.

YouTube. 2016. "Youtube Statistics." from https://www.youtube.com/yt/press/statistics.html

Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., and Li, X. 2011. "Comparing Twitter and Traditional Media Using Topic Models," *European Conference on Information Retrieval*: Springer, pp. 338-349.

Zhou, R., Khemmarat, S., and Gao, L. 2010. "The Impact of Youtube Recommendation System on Video Views," *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*: ACM, pp. 404-410.

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of marketing* (74:2), pp 133-148.