USER OPINION LEARNING

by

Yuanyuan Gao

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Business Administration

David Eccles School of Business

The University of Utah

December 2017

Copyright © Yuanyuan Gao 2017

All Rights Reserved

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Yuanyuan Gao**

has been approved by the following supervisory committee members:

Paul J. Hu	, Chair	**6/16/2017** Date Approved
Olivia R. Liu Sheng	, Member	**6/16/2017** Date Approved
Vandana Ramachandran	, Member	**6/16/2017** Date Approved
Jeffrey M. Phillips	, Member	**6/16/2017** Date Approved
Xiao Fang	, Member	Date Approved

and by **William S. Hesterly**, Associate

Dean of **David Eccles School of Business**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

The theme of my dissertation is users' opinion learning. We propose three different studies to learn users' opinion using various approaches and to address several important research questions. Firstly, in order to discover the significant factors that induce the rating differences from user-generated reviews, we first extract possible specific influences from the review, known as aspects, and then we propose an unsupervised aspect-based sentiment learning system that assigns sentiment scores to potential aspects. Based on the sentiment scores, we adopt linear regression models to identify the aspects that lead to the rating differences. Food quality, service, dessert and drink quality, location, value, and general opinion toward the restaurants are recognized as the main influential factors that cause the Yelp rating differences among chain restaurants. Secondly, to understand the impact of time reminder designs such as counting down clock, progressing bar indicator, and remaining number of advertisements reminder embedded in specific long and short advertisement videos, we propose a 4 by 2 between-subject experimental study with follow-up survey questions to collect user's opinions toward different temporal designs in the video. Thirdly, our study analyzes the advertisement video designs from the content level. We design the advertisement video with high and low content relevance levels with the desired video. A 2 by 2 between-subject experimental study with follow-up survey questions is proposed. Results point out that advertisement videos with high content relevance levels can lead to shorter video duration perception and less negative attitudes toward the video, but can also diminish the effectiveness of the advertisement with users recalling fewer products and brands promoted in both longer and shorter advertisement videos.

I would like to dedicate my dissertation work to my family: my loving parents Quan Gao and Xiaoshi Li, who support and encourage me, providing every possible good opportunity to me; and my husband Haibo Ding, who inspires me and has been there for me throughout the entire doctorate program.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I am very fortunate to have performed my graduate work at the University of Utah. I would like to first thank my advisor, Dr. Paul J. Hu, for his guidance and opportunities afforded. He is incredibly productive and organized. Under his mentorship, I have learned how to effectively move my projects forward.

I would also like to thank my dissertation committee members, Dr. Olivia Sheng, Dr. Vandana Ramachandran, Dr. Xiao Fang, and Dr. Jeff Phillips, for agreeing to serve on my committee and providing valuable feedback for my revisions.

CHAPTER 1

INTRODUCTION

The theme of my dissertation is users' opinion learning. User's opinions can be altered when the surrounding environments change. Learning user's opinions can provide various beneficial effects. For example, by knowing users' preferences, companies can design customer-personalized services and merchandise so as to satisfy user's specific needs. In addition, learning users' feedback allows the marketing department to adjust the appropriate strategy for promoting their products. More importantly, understanding users' opinion can allow us to discover the potential causal relations among different factors. Since user's opinions can be expressed and extracted from different channels such as their generated reviews, answers to specific survey questions, and other observations, we can adopt various techniques to retrieve user's opinions for different purposes. Under the users' opinion learning theme, three different studies are proposed to answer different specific research questions in my dissertation.

As technology has advanced over the past decade, the number of restaurants and other business-related social networks has increased rapidly. People are more likely to read, communicate, and evaluate the performance of restaurants on different websites such as Yelp.com (Potamias, 2012). Yelp.com is a website that hosts users' reviews about local businesses and restaurants. Users can write a review and rate a business by giving stars from 1 to 5. In addition, Yelp responds to relevant search results by displaying the average ratings in half star format and the average number of reviews received (Potamias, 2012). The ratings reflect users' average evaluation of specific restaurants and businesses in regards to various related aspects, such as food, service, environment, and so on. Besides expressing opinions in terms of numeric ratings, users nowadays also like to state their ideas in words. They can discuss specifically the preferred and unfavorable facets of a particular restaurant or business in the review section on Yelp.com. Based on the existing literature, in general, we have recognized a few factors that influence restaurant evaluation ratings, such as food quality (Haghighi, Dorosti, Rahnama, & Hoseinpour, 2012; Lewis, 1981; Soriano, 2002), menu variety (Lewis, 1981), service quality (Kim & Han, 2008; Meng & Elliott, 2008; Namkung & Jang, 2008; Soriano, 2002), restaurant environment and atmosphere (Haghighi et al., 2012; Jang & Namkung, 2009; Lewis, 1981), cleanliness (Andaleeb & Caskey, 2007), location (Haghighi et al., 2012; Soriano 2002; Yüksel & Yüksel, 2003), and price fairness (Haghighi et al., 2012; Han & Ryu, 2009; Kim & Han, 2008). However, given standard common menu items and prices, food sources, and the same advertising, we have observed diverse Yelp ratings of chain restaurants owned by the same corporation given the similar locations. Therefore, in the first study, we aim at discovering the factors affecting the differences of chain restaurants' Yelp ratings from users' opinions expressed and extracted from their generated reviews. The research question we are trying to answer is, given similar menu items, food quality, price, standard decoration style, and environment settings, what are the significant factors that influence the Yelp rating differences amongst chain restaurants under the same brand. To answer this question, aspects are firstly extracted from user-generated reviews. The aspects are considered as the distinctive features users mentioned about restaurants that might possibly play important roles in determining the overall Yelp ratings. To better understand different aspects, an unsupervised sentiment analysis learning method is proposed for learning the sentiment score of each retrieved aspect. Significant aspects that have impacts on restaurants' rating difference are then further analyzed from learned users' preferences. In the end, rating prediction models are built based on the aspects' sentiment score for future Yelp rating predictions.

Compared to the prior related studies, the first study is different in several ways. Firstly, the main focal unit of our study is chain restaurants instead of general restaurants (Andaleeb & Caskey, 2007; Haghighi et al., 2012; Soriano, 2002). Secondly, in order to evaluate the importance of different factors extracted from user-generated reviews, we propose a sentiment analysis model extending the current label propagation method (Brody & Elhadad, 2010). Thirdly, after assigning the sentiment score to each specific aspect retrieved, the main purposes of our study are to recognize significant aspects that impact chain restaurants' rating differences and to measure the predictive power of the polarities of different aspects.

To allow companies and their marketing department to better promote their products and brands, we propose two other studies to learn the impacts of different advertisement video designs on users' perceptions, opinions, attitudes, and recalls so as to assist them in appropriate advertisement video selection and design.

According to the latest infographic form released by invespcro.com, estimated digital video advertisement spending in the United States reach $9.59 billion dollars in

2016 from $5.96 billion dollars in 2014, while the worldwide revenue from online video advertising is estimated to achieve $11.4 billion by 2016 from $8.3 billion in 2014. However, given online video advertising as a relatively new advertising format, the studies related to it are still sparse. A few studies focus on the effects from distinctive features of online advertising videos, such as the skip button and ad placement (Hegner, Kusse, & Pruyn, 2016; Krishnan & Sitaraman, 2013; Kusse, 2013). A large number of other characteristics, such as temporal reminder designs and video content relevance with the surrounding media, are overlooked by existing studies.

Online video advertising is normally presented to viewers with a few time reminder designs. During the online video advertising, several temporal reminder designs such as counting down timer, progressing bar, and remaining number of sponsors are time reminders notifying the remaining advertisement time and duration. In the second study, we apply time perception theories, reactance theories and selective attention theories to analyze the effects of different time reminder designs embedded in the advertisement video on users' perception of the duration of the video, their attitude, and final products/brands recall moderated by the actual length of the advertisements. Based on the theories, we address the following research questions in this study: How does (1) the counting down clock, (2) progressing bar, and (3) number of remaining sponsors affect online users' perception of advertisement duration, the attitude effects, and message recall? We further analyze (4) how does the length of the online video advertisement change the effect that temporal designs have on the perceptions of advertisement duration time, the attitudes, and message recall? One 4 by 2 between-subject experimental study with follow-up survey questions is proposed to collect users' opinions toward different temporal designs in the video.

From the content design aspect, online video advertisement content relevance refers to the extent to which an online video advertisement is similar to the video content that users are currently watching, in terms of context, topic, or execution styles (Hua & Li, 2009; Li & Lo, 2014; Mei, Hua, Yang, & Li, 2007; Moorman, Neijens, & Smit, 2002; Van Reijmersdal, Neijens, & Smit, 2005). Apart from online video advertising format, for traditional advertisements, such as banner advertisements and newspaper advertisements, prior studies have evaluated various impacts of the content relevance between the ad and its surrounding media on purchasing intentions (Goldfarb & Tucker, 2011), return intentions (McCoy, Everard, Galletta, & Polak, 2004), customer attitudes (Jeong & King, 2010; Kim & Sundar, 2012) and advertising effectiveness (Jeong & King, 2010; Moorman et al., 2002; Porta, Ravarelli, & Spaghi, 2013; Rodgers, 2003; Simola, Kivikangas, Kuisma, & Krause, 2013; Yaveroglu & Donthu, 2008; Zanjani, Diamond, & Chan, 2011). As for the online video advertising type, existing studies have limited the general purposes of video content relevance on evaluating the customer attitudes and intrusiveness of the video (Hegner et al., 2016; Kim, 2015; Kusse, 2013).

In the third study, we intend to understand how the content design of advertisement videos affects users' perception of the video duration, the attitude, and final products and brands recall. Users' opinion was gathered after they watched the advertisement video with matching or not matching contents with the video they were currently watching. We apply time perception theories, contextual priming effect, and schema incongruent effect to evaluate the influences of different content relevance levels of online video advertisements with the desired video on advertisement duration perception, perceived users' negative attitude toward both advertisement itself and brands it promoted, and recall of advertisement contents. Based on the theories, we address the following research questions in this study: How does (1) content relevance, affect online users' perception of advertisement duration, the attitude effects, and message recall? We further analyze: (2) how does the actual length of the online video advertisement change the effect that content relevance has on the perceptions of advertisement duration time, the attitudes, and content recall? A 2 by 2 between-subject control experiment with follow-up questions was designated for querying users' opinion toward the advertisement contents.

The three studies composing my dissertation are presented in the following chapters. "What makes chain restaurants' rating diverse?: An aspect-based sentiment analysis" is in Chapter 2. "Online video advertisements: How does temporal design affect user's opinions?" is in Chapter 3. "Online video advertisements: How does content relevance design affect user's opinions?" is in Chapter 4. Chapter 2 is constituted by an introduction, literature review, theoretical foundation, aspect-based sentiment analysis, important aspect analysis, Yelp rating prediction, discussion, and conclusion section. Chapter 3 and Chapter 4 are comprised of an introduction, literature review, theoretical foundation, hypotheses development, study design, result, discussion, and conclusion section.

.

CHAPTER 2

WHAT MAKES CHAIN RESTAURANTS' RATING DIVERSE?: AN ASPECT-BASED SENTIMENT ANALYSIS

2.1 Introduction

Chain restaurants refer to a series of restaurants with the same name in many different locations under shared corporate ownership or franchising agreements. According to Nation's Restaurant News, the majority of the top chain restaurants are fast food restaurants and midscale or upscale restaurants. The most distinguishable characteristics of chain restaurants, in comparison to the independent restaurants, include the same brand name (e.g., Applebee's or Red Lobsters), standard common menu items and prices, food sources, same atmosphere (e.g., similar decorations and settings), and advertising. Unlike independent restaurants that operate in specific locations, chain restaurants can be scattered in different places.

On the other hand, as technology has advanced over the past decade, the number of restaurants and other business-related social networks have increased rapidly. People are more likely to read, communicate, and evaluate the performance of restaurants on different websites such as Yelp.com (Potamias, 2012). Yelp.com is a website that hosts users' review about local businesses and restaurants. Users can write a review and rate a business by giving stars from 1 to 5. Also, Yelp responds to relevant search results by displaying the average ratings in half star format and the average number of reviews received (Potamias, 2012). The Yelp ratings are important to business and restaurants. An extra half-star rating can cause restaurants to sell out 19% more frequently. Three-star restaurants receive 384% more calls than two-star restaurants, and four-star restaurants can receive 294% more directions and map views than three-star restaurants.

There are various factors that impact the restaurant evaluation and ratings such as food quality (Haghighi et al., 2012; Lewis, 1981; Soriano, 2002), menu variety (Lewis, 1981), service quality (Kim & Han, 2008; Meng & Elliott, 2008; Namkung & Jang, 2008; Soriano, 2002), restaurant environment and atmosphere (Haghighi et al., 2012; Jang & Namkung, 2009; Lewis, 1981), cleanliness (Andaleeb & Caskey, 2007), location (Haghighi et al., 2012; Soriano, 2002; Yüksel & Yüksel, 2003), and price fairness (Haghighi et al., 2012; Han & Ryu, 2009; Kim & Han, 2008). Surprisingly, given standard common menu items and prices, food sources, same advertising strategy, and similar locations (e.g., same city or metropolitan statistical area), we have observed different Yelp ratings of chain restaurants operated by the same company. As shown in Figure 2.1, three Applebees receive different Yelp ratings from 2, 2.5 to 3 stars, even though they are all located close to each other in Arizona.

Our study, therefore, is interested in discovering the potential features that influence the Yelp rating differences of chain restaurants. Prior studies have paid little attention to only the chain restaurants and focused their studies on customers' perceived value, satisfaction, and loyalty toward both chain and general independent restaurants (Haghighi et al., 2012; Han & Ryu, 2009; Kim & Han, 2008). According to restaurants' business report, $125 billion each year is spent on chain restaurants, accounting for over 50% of all restaurant spending in the United States. As of 2016, more than half of the restaurant industry's roughly $491 billion sales each year come from top chain restaurants (Luca, 2016). The giant growth rate of chain restaurants' sales and store counts is stronger than the restaurant industry overall. Given the great profit chain restaurants bring to the industry and the significant impact of Yelp ratings on customers' curiosity and restaurants' sales, in order to adjust the performance of chain restaurants, it is important to learn the criteria that affect the customers' evaluations and opinions for particular chain restaurants.

It is important but difficult to extract customers' opinion from general customers. Prior studies mostly use designated survey questionnaire querying customers' opinion regarding specific criteria affecting their evaluations of restaurants (Andaleeb & Caskey, 2007; Haghighi et al., 2012; Soriano, 2002). With the development and growing popularity of social networks such as Yelp.com, more and more people praise and criticize a variety of aspects of the target restaurant in the reviews, such as the waiting time of the restaurants and the portion of the food in the plate. As a result, an increasing number of studies start to extract the criteria and aspects customers mentioned in their generated reviews from different domains using content analysis (Pantelidis, 2010), pages crawler and regression model (Zhang, Ye, Law, & Li, 2010), and Latent Dirichlet Allocation model (Brody & Elhadad, 2010; Jo & Oh, 2011).

To extract the criteria that influence the customer ratings of chain restaurants, besides only retrieving the aspects of restaurants from reviews, the evaluations of each aspect for each restaurant are equally important for us to analyze further which features are significant factors. Therefore, in this study, we extract restaurant-related aspects in customer-generated reviews for chain restaurants, propose a sentiment analysis method to assign a sentiment score to each retrieved feature for factor comparison and evaluation, adopt econometric linear regression models for important factors recognition, and evaluate the predictive power of the important aspects discovered using predictive regression models.

Our study is different from the prior restaurant criteria evaluation-related studies in several ways. First, in comparison with analyzing criteria for general restaurants (Andaleeb & Caskey, 2007; Haghighi et al., 2012; Soriano, 2002), our study aims at discovering the evaluation features for only chain restaurants. Second, in order to assess the importance of different factors extracted from user-generated reviews, our study proposes a sentiment analysis model extending the existing label propagation method (Brody & Elhadad, 2010). Third, after assigning the sentiment scores to each particular aspect retrieved, the primary purposes of our study are to recognize significant facets that impact chain restaurants' rating differences and to evaluate the predictive power of the polarities of different aspects.

The remainder of the study is organized as follows. We first review related studies and highlight key differences between our study and previous representative research. We then formally propose an unsupervised aspect-based sentiment analysis method for aspects extraction and evaluation. Furthermore, we empirically use three regression models for important aspects analysis. Moreover, the recognized aspects and their polarity sentiment scores are used for building predictive models for future Yelp rating predictions. Restaurant attributes that can be easily extracted from Yelp.com serve as additional predictors. The models constructed only using these predictors are treated as benchmarks. We conclude the chapter by discussing important contributions, implications, and limitations.

2.2 Literature Review

Several streams of research are relevant to our study, including restaurant performance criteria evaluation, aspect-based opinion mining and sentiment analysis, and business rating prediction. In this section, we review representative research in each stream and highlight the gaps that motivate our research.

2.2.1 Restaurant Performance Criteria Evaluation

Prior studies evaluate various criteria that impact the customers' evaluation, satisfaction, and loyalty in the restaurant industry. The most common variables they evaluate are food, service, price, restaurant atmosphere, and location (Andaleeb & Caskey, 2007; Ha & Jang, 2010; Haghighi et al., 2012; Han & Ryu, 2009; Namkung & Jang, 2008; Soriano, 2002). For instance, Soriano (2002) studies the factors including food quality, service quality, cost of each meal, and location of the restaurants that affect customer' decisions about revisiting the place. Andaleeb and Caskey (2007) analyze restaurant cleanliness, atmosphere, space, food quality, price, responsiveness, and staff behavior. Besides the standard variables, customer orientation and relationship interests are evaluated for loyalty relationship quality in luxury restaurants (Meng & Elliott, 2008). Food neophobia (fearing new foods) is additionally assessed to investigate the relationships between customer personality, satisfaction, and loyalty (Kim & Sundar, 2012). Responsiveness, staff behavior, and restaurant working hours are further evaluated for affecting satisfaction of college students with food services (Andaleeb & Caskey, 2007). The food healthiness, restaurants' smoking environment, and visibility of food preparation area are studied for influencing tourists' satisfaction with restaurant services (Yüksel & Yüksel, 2003).

2.2.2 Aspect-Based Opinion Mining and Sentiment Analysis

To extract aspects from expressions and sentences, studies have focused on four main approaches: extracting aspects based on the frequent nouns and noun phrases (Blair-Goldensohn, Hannan, McDonald, Neylon, Reis, & Reynar, 2008; Hu & Liu, 2004; Popescu & Etzioni, 2005), extracting aspects by exploiting the opinion and target relationships (Blair-Goldensohn et al., 2008; Kobayashi, Iida, Inui, & Matsumoto, 2006; Somasundaran & Wiebe, 2009; Zhuang, Jing, & Zhu, 2006), extracting aspects using supervised learning methods (Jakob & Gurevych, 2010; Jin, Ho, & Srihari, 2009; Kobayashi, Inui, & Matsumoto, 2007; Lafferty, McCallum, & Pereira, 2001; Rabiner, 1989), and extracting aspects using topic modeling (Blei, Ng, & Jordan, 2003; Griffiths & Steyvers, 2003; Hofmann, 1999; Steyvers & Griffiths, 2007; Titov & McDonald, 2008).

As we mentioned above, Hu and Liu (2004) identify nouns and noun phrases based on the part-of-speech (POS) tagging. The nouns and noun phrases are extracted based on their frequencies. The infrequent nouns and noun phrases can be filtered out. Furthermore, the algorithm has been improved by Popescu and Etzioni (2005). They adopt a pointwise mutual information score between phrases to remove the noun phrases that are not associated with aspects. In addition, Blair-Goldensohn et al. (2008) refine the nouns and noun phrases extraction approaches by only identifying the aspects that co-occur with sentiment words.

The second method category assumes the sentiment opinions have target words and the target words are aspects. The relationships between the sentiment views and target words can be retrieved as sentence dependencies. By first identifying the sentiment words and then the nouns or noun phrases that modified by the sentiment words, the aspects can be extracted (Blair-Goldensohn et al., 2008). Since important aspects are often changed by sentiment opinions, they can also filter out unlikely nouns and noun phrases without the modification. Dependency parsers are used to identify the modification relationship (Kobayashi et al., 2006; Zhuang et al., 2006). Different from standard dependency parsers that define the dependency of individual words only, Wu, Zhang, Huang, and Wu (2009) use a phrase dependency parser, which recognizes dependency relationships from phrases, that is more suitable for aspect extraction.

Supervised learning methods have also been applied to information and aspect mining. The most widely used supervised learning models include Hidden Markov Models (HMM) and Conditional Random Fields (CRF). For example, Jin et al. (2009) apply a lexicalized HMM to learn patterns and to extract aspects and opinion expressions, while Jakob and Gurevych (2010) use trained CRF on review sentences from different domains with additional features such as tokens, POS tags, and word distances. On the other hand, other supervised methods such as tree classification models and Support Vector Machines (SVM) models can be adopted for aspects extraction. For example, Kobayashi et al. (2007) first discover candidate aspects and sentiment words using a dependency tree, then select a tree classification model to classify the candidate pairs as an actual aspect or not. A one-class SVM (Manevitz & Yousef, 2002) classification model is used to extract aspects. The candidate words are labeled as positive and negative. The SVM model is then applied to identify the positive candidates.

The last category of methods is topic modeling. Two basic models are probabilistic Latent Semantic Analysis (pLSA) and LDA models (Blei et al., 2003; Griffiths & Steyvers, 2003; Hofmann, 1999; Steyvers & Griffiths, 2007). There are a few drawbacks of basic topic modeling such as global topics, which include only the product names and brand names when the model is applied on customer reviews. Titov and McDonald (2008) therefore adopt a multigrain topic model with a local model discovering aspects from a few sentences as a document. Li and He (2009) do not explicitly separate the aspect words and the sentiment words that modify them and use a join topic-sentiment model to extend the LDA model. By considering only the adjectives as sentiment words, Brody and Elhadad (2010) extract aspects using local LDA models to identify both aspects and their sentiment opinions.

Regarding further sentiment classification and aspect sentiment analysis, most studies are usually formulated as a two-class classification problem with positive and negative polarities as two classes. Both supervised and unsupervised learning models are widely employed. The existing supervised learning methods can be applied to the text classification problem. The work by Pang, Lee, and Vaithyanathan (2002) is the first paper to take supervised learning models such as naïve Bayes and SVM to classify the polarity of movie reviews into two classes: positive and negative. In addition, there are a large number of papers use nonstandard machine learning methods for sentiment classification. For example, Dave, Lawrence, and Pennock (2003) propose a scoring function based on the words in positive and negative reviews. The minimum cut algorithm working on a graph is adopted in Pang and Lee (2004). A genetic algorithm-based feature selection is employed for sentiment classification in different languages (Abbasi, Chen, & Salem, 2008). There are two important drawbacks of using supervised learning methods. Firstly, accurate label annotations are required for training supervised learning methods, which is time-consuming. Secondly, the annotations used for model building are usually domain-specific, which can be only utilized in certain content for a particular domain. In terms of unsupervised learning models for aspect-based sentiment analysis, the most common methods are lexicon-based approaches (Ding, Liu, & Yu, 2008; Hu & Liu, 2004). They use a sentiment lexicon, which is composed of a list of sentiment words, phrases, and idioms, and then employ the sentence parse tree to determine the polarity orientation of each aspect from an individual sentence. For example, Hu and Liu (2004) sum up the sentiment scores of all opinion words in a sentence. Multiplication of sentiment scores from words has been adopted in Kim and Hovy (2004). Besides lexicon-based models, rules of opinions (Ikeda, Takamura, Ratinov, & Okumura, 2008; Jia, Yu, & Meng, 2009; Morante, Schrauwen, & Daelemans, 2011), and composition rules (Neviarouskaya, Prendinger, & Ishizuka, 2010) are necessary to identify sentiment words for aspects. Six types of composition rules are introduced: sentiment reversal, aggregation, propagation, domination, neutralization, and intensification, in Neviarouskaya et al. (2010).

2.2.3 Business Rating Prediction

Besides sentiment analysis, researchers also develop various models for review rating scores prediction. Amongst all studies, regression models and machine learning algorithms are mostly used for the prediction problem. In the study proposed by Pang and Lee (2005), they experiment with SVM regression, SVM multiclass classification using the one-versus-all strategy, and a meta-learning method called metric labeling. By using both labeled and unlabeled reviews, Goldberg and Zhu (2006) use a Graph-based semi-supervised learning model to improve the performance from the study Pang and Lee (2005) proposed. Instead of using the bag of words as predictive features, Qu, Ifrim, and Weikum (2010) introduce a bag of opinions to represent documents and to capture the strength of ideas. Besides the features that extract from the text, Li, Wang, Zhou, and Lee (2011) additionally incorporate reviewer and product information into a three-dimension tensor for review rating prediction. Apart from predicting the rating with a review as one unit, Snyder and Barzilay (2007) predict the rating for each aspect from the review. This study proposes two models using unigram and bigram from reviews as features. One model is the aspect model that targeted individual aspects, while the other model is the agreement model, which integrates the rating agreement among all aspects. Jiang, Yu, Zhou, Liu, and Zhao (2011) employ a similar Bayesian Network approach to the study proposed by Pang and Lee (2005), but they also aim at predicting aspect ratings instead of review ratings. The aspects are predicted only from the subset of the reviews instead of the whole review sets. The selected reviews are evaluated as the most representative of specific aspects.

2.2.4 Research Gap Summary

As we discussed in the introduction, understanding and recognizing the significant aspects that impact chain restaurants' rating differences are essential for restaurants' performance improvement. In comparison to the current studies (Andaleeb & Caskey, 2007; Ha et al., 2010; Haghighi et al., 2012; Han & Ryu, 2009; Kim & Sundar, 2012; Namkung & Jang, 2008; Soriano, 2002) that evaluate the criteria focusing on general restaurants, we explore a new realm of study that pays attention to only the chain restaurants.

Our proposed aspect-based sentiment analysis method adopts the topic modeling LDA model (Blei et al., 2003) for aspect extraction from restaurant reviews. Instead of using the standard LDA model that might extract global topics such as only the general product and brand names (Blei et al., 2003; Griffiths & Steyvers, 2003; Hofmann, 1999; Steyvers & Griffiths, 2007), we adjust and employ the local LDA model that treats each sentence as a document (Brody & Elhadad, 2010). Label propagation approach is further employed for sentiment evaluations of each aspect. Different from original label propagation methods (Brody & Elhadad, 2010; Zhu & Ghahramani, 2002;) that minimize the sentiment differences between two words connecting with each other in the word network, our objective function contains additional criteria that also reduce the differences between new polarity distribution of seed words and their original sentiment probability values, and between assigned sentiment score and default sentiment score value for each word.

Based on the sentiment score differences explored, our study aims at recognizing significant factors from extracted aspects that determine the rating differences. In comparison to rating predictions for selected reviews or specific aspects from reviews (Goldberg & Zhu, 2006; Pang & Lee, 2005; Snyder & Barzilay, 2007), we employ the retrieved and evaluated critical facets for overall restaurant rating predictions. Besides sentiment and text features, additional restaurant-related features such as price range, delivery, and available credit cards service are included as well for future restaurant rating prediction.

2.3 Theoretical Foundation

The essential characteristics of chain restaurants are their standardized menu and price, food resources, similar decoration and atmosphere, and promotion strategy. Based on the studies in environmental psychology, people's behavior is strongly associated with the surrounding physical environment (Mehrabian & Russel, 1974; Russel & Pratt, 1980). According to the theory "environmental stimuli are linked to behavioral responses by the primary emotional responses of arousal, pleasure, and dominance" (p. 31) that is proposed by Mehrabian and Russel (1974), individuals react to the environment with approaching if they have positive responses to the environment of the place and respond to the environment with avoidance if they have negative responses to the environment. Given the similar decoration, setting, and atmosphere of chain restaurants operated by the same corporation, the customer's responses should be similar. However, we have observed the different Yelp ratings given from customers as shown in Figure 2.1. Based on disconfirmation theory (Barsky, 1992), customers compare a new service experience with their developed service standard baseline. The rating difference between chain restaurants indicates a "gap" in service quality, food, environment, or location when the perception of these aspects received is less than the expectation (Zeithaml, Parasuraman, & Berry, 1990).

The motivation of this study is, therefore, identifying the important aspects that exist in the gaps between different chain restaurants under the same brand. Our proposed aspect-based sentiment analysis method is based on the label propagation model, which generates a network with similar nodes connecting with each other. The method assumes the opinion between two connecting nodes should be comparable. The effectiveness of the method can be explained by social influence theory (Deutsch & Gerard, 1955). One of the broad varieties of social influence is referred to as internalization (Kelman, 1958). According to internalization, individuals accept a set of values or norms (established by others) through socialization (Kelman, 1958). The label propagation generates a network that can be considered as a small social network with each node as one individual and each edge as a social connection. By socialization interactions, the node without sentiment values can receive the values from its nearest neighbors. That is exactly how the label propagation works.

2.4 Aspect-Based Sentiment Analysis

In order to analyze the aspects that significantly impact customers when they give different ratings to chain restaurants owned by the same parent company in various locations, we try to extract customers' opinions regarding a number of aspects for each restaurant from Yelp reviews. Customers' opinions regarding different facets such as food quality, service, and location are retrieved in the format of sentiment scores. We precede aspect-based sentiment analysis by extracting topics/aspects, determining optimal aspect number, categorizing aspects, selecting representative words for each topic, and generating and analyzing sentiment scores. The detailed methods are described fully in the following subsections.

2.4.1 Topic/Aspect Extraction

To extract topics from customer reviews, we use the local version of the Latent Dirichlet Allocation (LDA) model (Brody & Elhadad, 2010), which operates on sentences, treats each sentence as a document instead of the whole review, and employs a small topic number that can directly refer to different aspects. Directly adopting the LDA model only extracts global topics from the data, especially from online reviews, which is not suitable for our topic extraction (Blei et al., 2003; Griffiths & Steyvers, 2003; Hofmann, 1999; Steyvers & Griffiths, 2007). By adjusting the global LDA model to the local LDA model and treating each sentence as one document, we can discover specific aspects pertinent to the review. We adopt a standard implementation of the LDA model embedded in R package “mallet”. Integrated Yelp reviews are first processed by the Stanford NLP tool, and each word in the sentences is given a part-of-speech tag. To better recognize the construction of aspects, words from sentences are filtered, and only noun words are used for LDA model building. The output of the model is the distribution over inferred topics for each sentence in the data. We use the default value for the magnitude of the Dirichlet prior over the topic distribution of each sentence, which is 5.0. The per-word weight of the Dirichlet prior over the topic-word distributions is set as the number of words in the vocabulary. We run the algorithm with the number of aspects ranging from 2 to 20. The optimal number of different aspects is determined by cluster validation scheme.

2.4.2 Optimal Aspect Number Selection

To extract topics from customer reviews, we use the local version of Latent Dirichlet Allocation (LDA), which operates on sentences, treats each sentence as a document instead of the whole review, and employs a small topic number that can directly refer to different aspects.

To determine the optimal number of topics occurred in the review, we adjust the cluster validation procedure (Lange, Roth, Braun, & Buhmann, 2004; Levine & Domany, 2001; Niu, Ji, & Tan, 2007). By adapting such a procedure, a different number of topics can be compared, and the topic number with the highest consistency can be selected for our aspect extraction.

We treat each aspect/topic as a dependent cluster. For each sentence, we first pick the aspect with the highest probability as its representative aspect. Therefore, each sentence can belong to one particular cluster. We then define a connectivity matrix. Given all the sentences in our data, if two sentences belong to the same cluster, the value of that certain cell in the connectivity matrix is 1. Otherwise, the value can be assigned as 0. Furthermore, a random subsample from all sentences is selected from the data. One additional connectivity matrix is constructed based on the randomly selected subsample. Then we define the below equation (Levine & Domany, 2001) for cluster consistency validation:

$$V(C^{\mu}, C) = \frac{\sum_{i,j} 1\{C_{i,j}^{\mu} = C_{i,j} = 1, d_i \in D^{\mu}, d_j \in D^{\mu}\}}{\sum_{i,j} 1\{C_{i,j} = 1, d_i \in D^{\mu}, d_j \in D^{\mu}\}} \tag{2.1}$$

As we mentioned, D^{μ} is a random subset with size $\alpha\ |D|$ sampled from the full data set

D. Matrices C and C^{μ} are connectivity matrices computed based on D and D^{μ}, respectively, $0 \leq \alpha \leq 1$. We set α to 0.5 in this study.

The detailed validation procedure is demonstrated in the following: (1) Run the local LDA model with topic number = k on data D for aspect probabilities of each sentence (2<=k<=20); (2) Obtain connectivity matrix C_k; (3) Randomly assign probabilities to each sentence regarding the aspects it should belong to and construct one random comparison connectivity matrix R base on uniformly drawn random assignment of the sentences in data D; (4) Sample a random subset D^{μ} with size $\alpha\,|D|$ from the full data set D with $\alpha = 0.5$; (5) Run the local LDA model on the subset D^{μ} and construct the connectivity matrix C_k^{μ} based on the aspect probabilities computed; (6) Randomly assign probabilities to each sentence regarding the aspects it should belong to and construct one subrandom comparison connectivity matrix R_k^{μ} base on uniformly drawn random assignment of the sentences in data D^{μ}; (7) Calculate the validation score of $V(C_k^{\mu}, C_k)$, $V(R_k^{\mu}, R_k)$ and score(k) $= V(C_k^{\mu}, C_k) - V(R_k^{\mu}, R_k)$ given in the Equation 2.1; (8) Repeat the steps from (4) to (7) 10 times for topic number k; (9) Return the average $score(k)$ over 10 iterations.

By using a similar sized random assignment of aspects as comparison, this validation procedure can evaluate the consistency of our aspect solution. This validation procedure can select the aspect number k with the highest $score(k)$. We experiment with values of k from 2 to 20. For our Yelp reviews, we receive the highest score when k equals to 14.

2.4.3 Words Selection and Aspects Categorization

We select aspect number k = 14 for our Yelp reviews and operate the local LDA model on our processed sentences with only noun words. The representative words for each topic are presented with highest probabilities. The probabilities are calculated using the following equation (2.2):

$$p(word|aspect) = \frac{count[aspect,word]}{\sum_{w}^{words} count[aspect,w]} \tag{2.2}$$

We then select, for each aspect, the top 100 words for each of the 14 aspects. These words are used further for aspects categorization.

The 14 selected aspects are represented by the top 100 words with highest probabilities. Three annotators manually inspect the noun words for each aspect and determine the general aspect names. These 14 specific aspects are further categorized into nine general categories. The inferred aspects are presented in Table 2.1.

The model defines a finer granularity of aspects solely based on the review text. By adapting the local LDA model that treats each sentence as a document, in comparison to standard LDA that prefers global topics, the local LDA model distinguishes topics between reviews and recognizes local aspects between reviews, such as food and service.

2.4.4 Sentiment Analysis

To assess each facet of restaurants in order to discover the significant factors that impact the rating differences for the chain restaurants, we calculate the sentiment score for each aspect based on the reviews customers provided. The sentiment scores are computed for each aspect based on the polarity of the verbs and adjectives. These verbs and adjectives modify the nouns that represent the aspect. A word graph network is then constructed with all the extracted verbs and adjectives based on the similarity of the word embedding calculated from the words. We use existing words in the dictionary with polarities and their sentiment scores as seeds in the network and adjust the label propagation method for the estimation of sentiment score for each word in the network. The detailed process is described in the following subsections.

2.4.4.1 Relative Adjective and Verb Extraction

The business records and review data we used are downloaded from a Yelp data set challenge. The available Yelp English reviews are first processed by the Stanford NLP tool for tokenization and part-of-speech tagging. Then we adopt the Stanford parser tool for standard and advanced dependency relations extraction. Based on the dependency relations, we retrieve all adjective words if they are modifying a noun word. If there is an additional negation that also modifies the noun, we obtain the negation as well. For example, from the sentence "The waitress is friendly, but the food was not warm," we can extract the following pairs {friendly waitress} and {not-warm food}. Furthermore, besides adjectives that customers used to express their attitude toward different aspects, verbs are also words with polarity indications. For example, from the sentence "we love the service here" and "the food here disappoints us," we can extract the verbs "love" and "disappoints" that represent customer's sentiment opinions toward aspects such as service and food as well. The word "love," in most cases, shows an affectionate and positive polarity attitude. Generally speaking, the word "disappoint" speaks for an unfavorable and negative opinion direction. Moreover, any negations that modify the verbs can also be retrieved. For example, from the sentence "I do not adore this restaurant," we can extract the pair {not-adore restaurant}. Apart from individual words, the frequencies of adjectives and verbs that appear in the review are also calculated.

2.4.4.2 Sentiment Label Propagation

To estimate the sentiment score for different aspects from Yelp reviews, we aim at determining the sentiment score for the adjectives and verbs that modify the nouns and represent the aspects. The calculation of sentiment scores for adjectives and verbs is stemming from label propagation.

2.4.4.2.1 Word Network Building

We build a word network that is composed of different word nodes and edges connecting them. The words we use for creating the network are the adjectives and verbs modifying nouns extracted from reviews. The nodes denoted as n_i and n_j represent i_{th} and j_{th} words in the network, $i, j \in W$ and W is the collection of all adjectives and verbs we extracted from reviews. To construct the word network that used for further label propagation analysis, we aim at using edges to connect similar words. Similar words are different words with higher similarity scores.

To determine which two nodes are connected, we need to estimate the similarities between various words. We first compute the embedding of the words using the Global Vectors (GloVe) for word representation. GloVe is an unsupervised learning algorithm for obtaining vector representations for words (Pennington, Socher, & Manning, 2014). The model is trained on the non-zero entries of the global word-word co-occurrence matrix, which tabulated how frequently words co-occur with one another in a given corpus. The model has been trained on 6 billion tokens and builds a vocabulary of 400,000 frequent words (Pennington et al., 2014). We select word vectors for our adjectives and nouns from pretrained vectors. Cosine similarity between two word vectors can provide a practical measurement for evaluating the linguistic or semantic similarity between the corresponding words. For example, woman and man, sir and madam, queen and king are similar word pairs based on the cosine similarity using word embedding vectors.

To construct the word network, for each word (adjective or verb) we extracted, we calculate the cosine similarity values for this focal word with all available words based on their embedding word vectors. The cosine similarity is computed using the following equation (2.3):

$$similarity\left(n_i, n_j\right) = cos(\theta) = \frac{n_i \cdot n_j}{\|n_i\|_2 \|n_j\|_2} = \frac{\sum_{i=1}^{n} V_i V_j}{\sqrt{\sum_{i=1}^{n} V_i^2}\sqrt{\sum_{i=1}^{n} V_j^2}} \tag{2.3}$$

where V_i and V_j are components of vectors for words n_i and n_j, respectively. Based on the cosine similarity, for each focal word in the collection, we select k (k=5) nearest neighbors with highest similarity and draw an edge between each nearest neighbor with the focal word. Specifically, each focal word can be assigned with k nearest neighbors, besides the edges established when the focal word is selected as nearest neighbors by other words. Therefore, each node in the word graph network is connected to at least k other nodes by more than k edges. The weight of the undirected edge connecting n_i and n_j is the cosine similarity between the word vectors embedding. The node and edge of the graph can be demonstrated in Figure 2.2.

2.4.4.2.2 Iterative Sentiment Label Propagation

Intuitively, we want words that are close to each other to have the similar polarity labels. Given a word graph network we defined in the prior section, each node represents an individual adjective or verb that is modifying a noun. The undirected edge connects two different words with high cosine similarity. The purpose of our iterative label propagation is to define an objective function and assign a sentiment score to each node in the graph iteratively until we reach the minimum value of the objective function.

2.4.4.2.2.1 Original label propagation method. Our iterative label propagation method is based on the original label propagation algorithm that is proposed by Zhu and Ghahramani (2002) and Brody and Elhadad (2010). The aim of the method is using propagation from edge to edge to define the labels for all the nodes in the network with the assumption that similar nodes should have the same label. Firstly, there is a group of nodes that contain predefined labels in the network, known as seeds. The rest of the nodes are given random or consistent labels as initial values before the propagation process. For example, predefined positive sentiment words are given label value 1, predefined negative words are given label 0, while the rest of the undefined words are assigned label 0.5 (Brody & Elhadad, 2010).

Then a propagation step is repeated. During the iteration *t*, for each node that is not the seed nodes, the following update rule is adapted shown in the equation (2.4):

$$L(i)^t = \frac{\sum_{j \in N(i)} w(j,i) \cdot L(j)^{t-1}}{\sum_{j \in N(i)} w(j,i)} \tag{2.4}$$

where $L(i)^t$ is the label of node i at step t; $N(i)$ is the set of all nodes connecting focal node i; $w(j,i)$ is the edge weight between node i and neighbor node j; and $L(j)^{t-1}$ is the label of node j at step t-1, where j is one of the neighbors of focal node i. The propagation step is repeated until the values of the label converge.

2.4.4.2.2.2 Word dictionary selected for seed set. To start the propagation process, a set of words with a predefined label is necessary. In our study, we use MPQA subjectivity lexicon (Wilson, Wiebe, & Hoffmann, 2005) to match the words in our network and assign initial seed labels to them. The total number of words with polarity positive or negative in the lexicon is around 8,222. Out of these 8,222 lexicons, we match 2,253 words out of 10,002 total words in our network and assign them predefined sentiment scores. Brody and Elhadad (2010) assign a sentiment score 1.0 to the word with positive label, a sentiment score 0.0 to the word with negative label, and neutral scores 0.5 to the rest of the nodes. Unlike that, our study assigns a sentiment score vector to each node in the network. If the node is identified as positive in the lexicon, the sentiment vector distributed to this focal node can be denoted as Q and Q = <1, 0>. The first element of the vector indicates the probability of the node having positive polarity and the second value indicates the likelihood of the node having negative polarity. Similarly, if the node is identified as negative in the lexicon, the sentiment vector describing this node is <0, 1>. For those unidentified nodes, the vector can be <0.5, 0.5>. As a result, each node in the network is represented as a score vector with probability distribution <positive, negative>. Some of the seed words with polarities are demonstrated in Table 2.2.

2.4.4.2.2.3 Kullback–Leibler Divergence. Since each node/word in our network is denoted as a vector with sentiment probability distributions, to measure the differences and inconsistencies between the vectors, we adopt Kullback–Leibler Divergence (KL-Divergence) (Kullback & Leibler, 1951). The KL-Divergence is a measure of how one probability distribution diverges from a second expected probability distribution. Assuming we have two node vectors denoted as C_i and C_j representing the sentiment probability distributions of nodes n_i and n_j, the differences between these nodes can be calculated by KL-Divergence in equation (2.5):

$$D_{KL}(C_i \parallel C_j) = \sum_m^M C_i(m) \log(\frac{C_i(m)}{C_j(m)}) \quad (2.5)$$

where M is the total number of probability class. In our case, M is 2, since we have two classes for the probability distribution {positive, negative}. $D_{KL}(C_i \parallel C_j)$ in the context of machine learning can also be referred to as information gain achieved if C_i is used instead of C_j.

2.4.4.2.2.4 Edge weight normalization. As described in the prior section, for each word as one node in the network, we connect it with its k most similar words using edges with their cosine similarity as weights. For nodes n_i and n_j, the edge weight connecting these two is denoted as E_{ij}^{sim}. E^{sim} is a $N \times N$ matrix with N equal to the total number of nodes in the network. If nodes n_i and n_j are not connected with each other, we define $E_{ij}^{sim} = 0$. Given the estimated sentiment probability distribution on two similar nodes n_i and n_j, the difference between these two nodes is measured by KL-Divergence $D_{KL}(C_i \parallel C_j)$.

One issue of the current edge weight is that high-frequency words can become more highly connected in the network than low frequency words. To normalize the edge weight based on the node frequency so as to achieve better results, we introduce a diagonal matrix I in the below equation (2.6), which is also used in spectral clustering (Ng, Jordan, & Weiss, 2001).

$$I_{ii} = \sum_{j=1}^{N} E_{ij}^{sim} \tag{2.6}$$

where N is the total number of matrix columns. The values in the diagonal matrix I are therefore the sum of values in each row from edge weight matrix E^{sim}.

We then define the normalized edge weight matrix $\tilde{E}^{sim}$ using the below equation (2.7):

$$\tilde{E}^{sim} = I^{-\frac{1}{2}} E^{sim} I^{-\frac{1}{2}} \tag{2.7}$$

As a result, the normalized edge weight between nodes n_i and n_j can be calculated as following in the equation (2.8):

$$\tilde{E}_{ij}^{sim} = \frac{E_{ij}^{sim}}{\sqrt{I_{ii} I_{jj}}} \tag{2.8}$$

2.4.4.2.2.5 Objective function. The objective of using the label propagation method in our study is to estimate the sentiment probability distribution for adjectives and verbs in the network based on the probability distribution of seed words and the graph network construction. The theory of the label propagation method as mentioned in the prior section is that similar nodes should have comparable sentiment labels. Therefore, in our study, the first objective of our method is to minimize the probability distribution differences between two conjunct nodes. In addition, for those nodes that have already been identified in the lexicon, we can also include them in the propagation step. The second objective is to minimize the probability distribution differences between the updated seed word probability values during propagation and their original probability distribution initially assigned by matching lexicon. If a node's probability distribution cannot be estimated by either seed words or its nearest neighbors, we keep the probabilities close to 0.5. The third objective, therefore, is to minimize the probability differences between the estimated probability distribution and distribution *P*, where *P* = <0.5, 0.5>.

We define the objective function as shown in *Object (C)* demonstrated in the following equation (2.9):

$$C = \sum_{i=0}^{N_l} D_{KL}\left(\widehat{C_i} \parallel C_i\right) + \sum_{(i,j)} \tilde{E}_{ij}^{sim} D_{KL}\left(\widehat{C_i} \parallel \widehat{C_j}\right) + \alpha \sum_{i=0}^{N} D_{KL}\left(\widehat{C_i} \parallel P\right) \quad (2.9)$$

The first term penalizes the probability distribution of seed words when they are far away from their initial distribution found in the lexicon. We only contain this term in the objective function when n_i is one of the seed words. The second term summarizes the differences between two conjunct words n_i and n_j with $\tilde{E}_{ij}^{sim}$ as the normalized edge weight between the two nodes. The third term keeps the consistency of probability distribution between n_i and *P* with the adjusted parameter α, when no estimation can be made during propagation with the first term and second term.

2.4.4.2.2.6 Iterative sentiment probability propagation method. Since each term in the objective function $Object\ (C)$ is convex, the sum of the terms is convex as well, when the parameters are non-negative. Derived from objective function (2.9), to determine the sentiment probability distribution for each node in the network, we can use an iterative probability update algorithm to minimize the objective function. The iterative function is displayed in the following equation (2.10):

$$\widehat{C_i}^{\,t} \propto \exp \frac{\beta(i) \log(C_i) + \sum_j \tilde{E}_{ij}^{sim} \widehat{C_j}^{\,t-1} + \alpha \log(P)}{\beta(i) + \sum_j \tilde{E}_{ij}^{sim} + \alpha} \quad (2.10)$$

where $\beta(i)$ is 1 when node n_i is contained in the lexicon and used as seed word, otherwise $\beta(i)$ is 0. In our study, we adjust the parameter of α to be 0.1. $\widehat{C_i}^{\,t}$ is the probability distribution of node n_i at step t and $\widehat{C_j}^{\,t-1}$ is the probability distribution of node n_j at step $t-1$. The propagation step is repeated until the values of distribution converge or the iteration goes on for more than 100 times.

The complete iterative process can be summarized in the following steps: (1) given $\tilde{E}_{ij}^{sim}$ and P, use iterative function equation (2.10) to update $\hat{C}$; (2) when the value of $\hat{C}$ has not converged or the iteration time is less than 100, go back to step 1; (3) when the value of $\hat{C}$ has converged or the iteration time has reached 100, return the value of $\hat{C}$ to users.

2.4.4.2.2.7 Sentiment score computation. The above iterative propagation step can generate the sentiment probability distribution for each node in the network. To further estimate the sentiment score for each restaurant aspect, we can further transform the probability distribution to a single sentiment score for each adjective and verb. Given the $\widehat{C_\iota}$ as the final sentiment probability distribution estimation with $\widehat{C_\iota} = < \hat{p}_i^p, \hat{p}_i^n >$, where $\hat{p}_i^p$ is the probability that node n_i has a positive polarity label and $\hat{p}_i^n$ is the probability that node n_i has a negative polarity label, we introduce the following equation (2.11) for calculating one single sentiment score for the node n_i in the network.

$$Score_i = \frac{\hat{p}_i^p - \hat{p}_i^n + 1}{2} \tag{2.11}$$

If $\hat{p}_i^p = \hat{p}_i^n = 0.5$, $Score_i$ is 0.5 as well, which represent that node n_i has a neutral polarity. $Score_i$ can be 1 when $\hat{p}_i^p = 1$ and $\hat{p}_i^n = 0$, which indicates node n_i has a strong positive polarity. On the other hand, n_i has a strong negative polarity when $\hat{p}_i^p = 0$ and $\hat{p}_i^n = 1$, and therefore $Score_i$ is 0.

2.4.5 Aspect-Based Sentiment Analysis

By adopting iterative sentiment probability propagation, each node receives a probability distribution estimating the likelihood of the focal node to be contained in positive or negative polarities. Through further sentiment score computation, the sentiment probability distribution can be transformed to a single sentiment score. Because the purpose of our study is to evaluate the sentiment score of different aspects discovered in chain restaurant reviews and further analyze various aspects, we shall assign sentiment evaluations to restaurant aspects on the basis of the polarities defined by adjectives and verbs that modify nouns.

2.4.5.1 Chain Restaurant Yelp Review Extraction

As mentioned in section 2.4.4.1, we download Yelp challenge data sets incorporating various detailed information such as their business locations, users, tips, reviews, and check-ins. Because our study is interested in restaurants and chain restaurant, in particular, we filter the available Yelp data and only keep the information regarding chain restaurants located in the United States. Based on the top 100 chain restaurant list in the United States available at Nation's Restaurant News, we match the restaurant names with business names from Yelp data sets and successfully obtain information for chain restaurants under 81 brands such as McDonald's, Burger King, Applebee's, Dairy Queen, and Red Lobster. The particular chain restaurants are located in six U.S. major cities including Madison, Las Vegas, Phoenix, Champaign, Pittsburgh, and Charlotte. In total, we retrieve 2,309 individual chain restaurants and obtain 48,784 reviews specifically for these restaurants. Other related information such as business id, business name, location, the number of reviews of the restaurants is also extracted.

2.4.5.2 Sentiment Score for Aspects

In section 2.4.4.1, adjectives and verbs are extracted when they are modifying a noun based on the dependencies of review sentences. If there are any negations modifying the nouns, adjectives, or verbs, they are also retrieved. The adjectives and verbs are further included in the word graph network and assigned with the sentiment score using our iterative sentiment propagation method.

To approximate and evaluate the sentiment score for each aspect describing the chain restaurants, we adopt the following steps.

Firstly, we select the top 20 representative words with highest probabilities for nine categories of general aspects summarized in section 2.2.4.3. Secondly, the nouns with adjectives and verbs as modifiers are used to match the representative words. If any adjectives or verbs are modifying the representative words, their sentiment scores are further assigned to the particular aspect the words correspond to. For example, assuming adjective "excellent" has a polarity distribution <0.8, 0.2> with computed sentiment score 0.8, for the dependency pair "excellent ice cream," we can give the score 0.8 to the aspect "ice cream" stands for, which is "drink and dessert." If there are negations incorporated in the dependency relations, for example, for the pair "not-excellent ice cream," instead of using equation (2.11) for "excellent" sentiment score computation, we can apply the following equation (2.12) to calculate the score for "not-excellent":

$$Score_i^{neg} = \frac{\hat{p}_i^n - \hat{p}_i^p + 1}{2} \tag{2.12}$$

Therefore, the sentiment score for "not-fantastic" and for the aspect "drink and dessert" can be transformed to 0.2. Thirdly, after assigning the sentiment scores to representative words and their corresponding aspects from all reviews available for chain restaurants, we integrate the sentiment scores of each aspect for each specific restaurant. Specifically, all reviews with the same business id are considered belonging to the same individual restaurant. The sentiment scores retrieved from its reviews for different aspects are averaged and assigned to the certain restaurant. For example, in total, three words "fries," "cheese," and "shrimp" representing aspect "food" are retrieved from reviews of one of the Red Lobster restaurants. The sentiment scores for adjectives modifying these three words are 0.2, 0.4, and 0.8. As a result, we can receive the food aspect sentiment score as the average score 0.47 for the particular Red Lobster restaurant. Last but not least, if there are any missing aspects of the nine general categories in the reviews for one individual restaurant, we assign the default sentiment value 0.5 to the missing aspects.

2.5 Important Aspects Analysis

After aspect-based sentiment analysis, we gain the sentiment scores for nine general aspects "food," "drink and dessert," "value," "price," "location," "service," "environment," "anecdote," and "general" for each chain restaurant located in different places. To analyze the aspects that play a significant role in determining chain restaurant rating differences, three different Ordinary Least Square (OLS) linear regression models are introduced, and their results are discussed in the following.

2.5.1 Linear Regression Model

We first construct a linear regression model without controlling any other effects to analyze the general influences of different aspects.

2.5.1.1 Data Preparation

To study the important aspects impacting the rating difference of chain restaurants, we adopt rating differences as our dependent variables. Specifically, the rating difference can be calculated using the following equation (2.13):

$$rateDiff_i = rate_i - \frac{\sum_i^R rate_i}{R} \tag{2.13}$$

where $rate_i$ is the average rating score given by customers on Yelp for one focal restaurant i. R is the total number of restaurants owned by the same corporation and use the same brand name. $rateDiff_i$ is the rating difference between $rate_i$ and average rating of all restaurants with the same brand name. For example, we retrieve three different Olive Garden restaurants from Yelp data sets with average rating scores 2.0, 3.75, and 3.25. The rating differences can be calculated as -1, 0.75, and 0.25.

Instead of using sentiment scores of aspects as our independent variables, we transform the aspect sentiment scores to aspect sentiment score differences. In a similar manner, we calculate the sentiment score differences for each aspect of the specific restaurant by subtracting the sentiment score by the average of certain aspects' sentiment score of all restaurants under the same brand. For example, the sentiment scores for the "service" aspect of three different Olive Garden restaurants are 0.2, 0.6, and 0.7. One of the independent variables, sentiment score differences, can be computed as -0.3, 0.1, and 0.2.

2.5.1.2 General Linear Regression Model Design

We build a linear regression model with rating difference as the dependent variable and aspects' sentiment score differences as independent variables shown below in model (2.14):

$$rateDiff_i = \beta_0 + \beta_1 foodDiff_i + \beta_2 serviceDiff_i + \beta_4 valueDiff_i \quad (2.14)$$
$$+ \beta_3 drinkdessertDiff_i + \beta_5 enviornmentDiff_i + \beta_6 anecdoteDiff_i$$
$$+\beta_7 priceDiff_i + \beta_8 locationDiff_i + \beta_9 generalDiff_i + \varepsilon_i$$
$$i = 1, \ldots, 2{,}309$$

where i represents restaurant index, which is from 1 to 2,309. $foodDiff_i$, $serviceDiff_i$, and $generalDiff_i$ are the sentiment score differences for different aspects of the focal restaurant i. β_1 to β_9 are the coefficients of interest.

2.5.2 Time Fixed Effect Model

It is possible that there is a common time trend that affects the rating differences, such as remodeling the restaurant or training the employees. Therefore, we include time fixed effects to account for time trends.

2.5.2.1 Data Preparation

The average rating difference and sentiment score difference data used for the building model (2.16) are time invariant, which are the results from all variable reviews from the year 2005 to the year 2015. To build a time fixed effect model, we further construct a panel data with rating differences for each restaurant per year as dependent variables and sentiment score differences for aspects per year as independent variables.

Specifically, since users also post numeric rating scores corresponding to their unstructured text reviews on Yelp, the average rating for each restaurant for a particular year is calculated by obtaining the average rating of the given reviews during that year. For example, three reviews were given to the focal Olive Garden during the year 2006 and the average rating of these three reviews is 3.25. The average Olive Garden rating of the year 2006 based on all reviews during that time can be calculated as 3.0. As a result, the difference between 3.25 and 3.0, which is 0.25, is used as the dependent variable for the focal Olive Garden. The rating difference equation can be updated in the following equation (2.15), where parameter t is the year from 2005 to 2015.

$$rateDiff_i^t = rate_i^t - \frac{\sum_i^R rate_i^t}{R} \tag{2.15}$$

Similarly, the sentiment scores for different aspects are also retrieved from the three reviews only during the year 2006 for the focal Olive Garden. The sentiment score differences in the year 2006 for specific aspects are further calculated with the differences between the average score of all Olive Gardens and the score of the focal Olive Garden during the year 2006.

2.5.2.2 Time Fixed Effect Model Design

Our time fixed effect model using the above described panel data is specified as model (2.16):

$$\begin{gathered} rateDiff_{it} = \beta_0 + \beta_1 foodDiff_{it} + \beta_2 serviceDiff_{it} + \beta_3 alueDiff_{it} \qquad (2.16) \\ + \beta_4 drinkdessertDiff_{it} + \beta_5 enviornmentDiff_{it} + \beta_6 anecdoteDiff_{it} \\ + \beta_7 priceDiff_{it} + \beta_8 locationDiff_{it} + \beta_9 generalDiff_{it} + \gamma_t + \varepsilon_{it} \\ i = 1, \ldots, 2{,}309 \end{gathered}$$

where t represents year index and γ_t represents time fixed effects.

2.5.3 Two-Way Fixed Effects Model

Although we control for restaurant location characteristics based on sentiment score differences, there may still be unobserved location characteristics, which are not accounted for by the control variables. To account for the location heterogeneity, we include city fixed effects in addition to time fixed effects. Hence, our two-way fixed effects model is specified as model (2.17):

$$rateDiff_{it} = \beta_0 + \beta_1 foodDiff_{itc} + \beta_2 serviceDiff_{itc} + \beta_3 valueDiff_{itc} \quad (2.17)$$

$$+ \beta_4 drinkdessertDiff_{itc} + \beta_5 enviornmentDiff_{itc} + \beta_6 anecdoteDiff_{itc}$$

$$+\beta_7 priceDiff_{itc} + \beta_8 generalDiff_{itc} + \gamma_t + \delta_i + \varepsilon_{itc}$$

$$i = 1, \ldots, 2{,}309$$

$$t = 2005, \ldots, 2015$$

$$c = 1, \ldots, 6$$

where c represents city index and δ_i represents location fixed effects.

2.5.4 Estimation Results

We now report our estimation results for the models described in the previous section. Model (2.14) in Table 2.3 represents the results of the linear regression model. As shown in model (2.14), the coefficients of overall sentiment score differences of food, drink and dessert, service, and restaurant brand general are positive and significant (*p*-value <0.01), which indicate the differences of customers' attitude toward the aspects including food, drink/dessert, service, and restaurant general among restaurants under the same brand, significantly impact the rating differences they gave. To be precise, for example, a 0.1 overall increase in the sentiment score difference in food aspect results in around 0.05 rating difference increase for the focal restaurant. In the overall linear regression model, the aspect price difference is positive and not significant. This result may be surprising for ordinary independent restaurants. However, for chain restaurants under the same brand, since the menu and price are standard, customers' attitude differences toward price may not significantly influence their ratings and rating differences.

Next, we explain the results of time fixed effect model (2.16) and two-way fixed effects model (2.17). The joint F test of all time effects has a value of 0.455 (p-value >0.001), which does not confirm the existence of a time trend. On the other hand, the joint F test of all location effects has a value of 9.070 (p-value <0.001), which confirms the existence of a location difference.

In addition to the four aspects that have a significant impact on the rating difference shown in the linear regression model, the aspects "value" and "location" play a major role as well in model (2.16) and model (2.17). The aspect "value" represents the portion, selection, and size of food by the specific restaurant. Even though the price and menu might be similar from all restaurants under the same brand, different restaurants serve various portions and amounts of food to their customers. As a result, the sentiment differences toward the "value" also influence the rating differences. Furthermore, it is also reasonable that customers consider the location of the restaurant as one of the important evaluation elements when they give ratings. Customers might find different convenience levels between a restaurant located in the center of the downtown mall and a restaurant located in the suburbs.

Similar to the aspect “price,” the sentiment score difference of aspect “environment” is not significant in all three models. The environment of the restaurant refers to the overall atmosphere of the place, which can include whether they play music, have TV, or locations of their tables and booth. Given the units of analysis in this study are various chain restaurants under the same brand name, the restaurant styles and facilities should also be similar. It is not surprising that the different polarity score of “environment” aspect does not significantly impact the overall rating differences.

2.6 Yelp Rating Prediction

After the important aspect analysis, we gain six significant factors “food,” “drink and dessert,” “value,” “location,” “service,” and “general” out of the nine general aspects that affect the difference of Yelp ratings amongst chain restaurants located in different places. We further evaluate the predictive power of the sentiment score of these extracted aspects from customer-generated reviews in predicting the future Yelp rating of chain restaurants. To assess the predictive power of aspects for chain restaurant rating prediction, six different predictive linear regression models are introduced with different feature sets using two panel data sets, respectively.

2.6.1 Data Preparation for Rating Prediction

To validate the predictive power of the sentiment scores from different aspects for restaurant Yelp rating forecast, instead of using the Yelp rating differences as the dependent variables, we use the actual Yelp ratings as the target variables this time.

Similarly, sentiment scores from the extracted aspects are obtained as distinct predictors.

Two different panel data sets are constructed for building the predictive models. For the first data set, sentiment scores of various aspects from reviews of each particular year for each restaurant are retrieved individually. It is used for predicting the Yelp rating of the focal restaurant for the following year. For example, in total, 10 reviews were collected for one of the Applebee's restaurant during the year 2006. The average sentiment scores for each aspect of the year 2006 can be extracted and computed from these 10 reviews. These scores are used for predicting the Yelp rating of the focal Applebee's restaurant during the year 2007, which is the average of ratings given by customers who left reviews specifically during the year 2007. In terms of the second data set, we retrieve and use the integrated sentiment scores and Yelp ratings. In this case, sentiment scores of the distinct aspects from reviews of past years for each restaurant are retrieved together. It is used for predicting the combined Yelp rating of the focal restaurant for the following year. For instance, besides the 10 reviews written for the focal Applebee's restaurant in 2006, there are five reviews drafted in the year 2005, and three reviews were written as early as the year 2004. The combined average sentiment scores for each aspect of the year 2006 are extracted and computed from these eighteen reviews from the year 2004 to the year 2006. These scores are used for predicting the united Yelp rating of the focal Applebee's restaurant for the year 2007, which is the average of ratings given by customers who also left reviews specifically between the year 2004 and year 2007. If there are any sentiment scores of certain aspects missing from the reviews, we assign the default sentiment score 0.5 to the missing aspects. The differences between the two constructed data sets are indicated in Table 2.4.

In total, we have collected 6,493 restaurant records with sentiment scores for nine aspects and their corresponding Yelp ratings from the year 2005 to the year 2015 for both data set one and two. Besides sentiment scores, additional features of each restaurant are extracted from Yelp as well, so as to use as benchmark predictors and to evaluate the predictive power of sentiment scores. The additional features include the price range, delivery, credit card, city, kids, group, and review number of each focal restaurant. Only the attribute review number is numeric among all the additional features. In order to match the yearly average and integrated yearly average calculation method of sentiment scores and ratings, the review number for each restaurant is extracted and computed for both individual year and integrated past years as well. The detailed description of the additional features is presented in Table 2.5.

2.6.2 Yelp Rating Predictive Models Design

Six predictive linear regression models are built using data one and two, respectively, with different predictor combinations. The target variable is the Yelp rating of the following year. In the first model (2.18), we only include predictors City and ReviewNum, which are the two most easily extracted attributes of the restaurant from Yelp. We further contain the rest five attributes as additional predictors in model (2.19). Except for the attribute ReviewNum, the other six attributes in Table 2.5 are time-invariant. In model (2.20), sentiment scores of eight general aspects "food," "drink and dessert," "value," "price," "service," "environment," "anecdote," "general" as well as city are used for predicting the Yelp ratings. Then only the significant aspects "food," "drink and dessert," "value," "service," "general" evaluated in the previous section, as well as city, are adopted for building the predictive model (2.21). A combination of all predictors from model (2.19) and (2.20) are used for constructing model (2.22), and all predictors from model (2.19) and (2.21) are utilized for model (2.23).

$$rating_t = \alpha_{t-1} + \beta_1 City + \beta_2 ReviewNum_{t-1} + \varepsilon_t \tag{2.18}$$

$$rating_t = \alpha_{t-1} + \beta_1 City + \beta_2 ReviewNum_{t-1} + \beta_3 PriceRange \tag{2.19}$$
$$+\beta_4 Delivery + \beta_5 CreditCard + \beta_6 Kids + \beta_7 Group + \varepsilon_t$$

$$rating_t = \alpha_{t-1} + \beta_1 City + \beta_2 foodscore_{t-1} + \beta_3 drinkdessertscore_{t-1} \tag{2.20}$$
$$+\beta_4 valuescore_{t-1} + \beta_5 servicescore_{t-1} + \beta_6 environmentscore_{t-1}$$
$$+\beta_7 anecdotescore_{t-1} + \beta_8 pricescore_{t-1} + \beta_9 generalscore_{t-1} + \varepsilon_t$$

$$rating_t = \alpha_{t-1} + \beta_1 City + \beta_2 foodscore_{t-1} + \beta_3 drinkdessertscore_{t-1} \tag{2.21}$$
$$+\beta_4 valuescore_{t-1} + \beta_5 servicescore_{t-1} + \beta_6 generalscore_{t-1} + \varepsilon_t$$

$$rating_t = \alpha_{t-1} + \beta_1 City + \beta_2 ReviewNum_{t-1} + \beta_3 PriceRange \tag{2.22}$$
$$+\beta_4 Delivery + \beta_5 CreditCard + \beta_6 Kids + \beta_7 Group + \beta_8 foodscore_{t-1}$$
$$+\beta_9 drinkdessertscore_{t-1} + \beta_{10} valuescore_{t-1} + \beta_{11} servicescore_{t-1}$$
$$+\beta_{12} environmentscore_{t-1} + \beta_{13} anecdotescore_{t-1} + \beta_{14} pricescore_{t-1}$$
$$+\beta_{15} generalscore_{t-1} + \varepsilon_t$$

$$rating_t = \alpha_{t-1} + \beta_1 City + \beta_2 ReviewNum_{t-1} + \beta_3 PriceRange \tag{2.23}$$
$$\beta_4 Delivery + \beta_5 CreditCard + \beta_6 Kids + \beta_7 Group + \beta_8 foodscore_{t-1}$$
$$+\beta_9 drinkdessertscore_{t-1} + \beta_{10} valuescore_{t-1}$$
$$+\beta_{11} servicescore_{t-1} + \beta_{12} generalscore_{t-1} + \varepsilon_t$$

We present six predictor combinations in Table 2.6.

2.6.3 Model Evaluation and Discussion

The predictive model performances are evaluated and compared using correlation coefficient, mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE). The results for model (2.18) to model (2.23) using data set one and two, respectively, are presented in Table 2.7 and 2.8.

Table 2.7 demonstrates the model performance results using yearly average data one. In comparison with only using easily extracted attributes review number and city, using additional time-invariant features such as price range, delivery, credit card, and so on decreased the error measures MAE by 0.02, RRSE and RAE by around 3%, and increased the correlation coefficient between the prediction and actual rating from 0.0855 to 0.2036. When only use the sentiment score of all different aspects or just significant aspects from customer-generated reviews, the errors generated by the four error measures are comparable with the errors using all restaurant-related attributes. When combining the sentiment scores from all aspects with other attributes in model (2.22), the correlation coefficient increased to 0.2389, and the errors evaluated by four error measures decreased by 2-3%, in comparison to the four previous models. By associating only significant aspects' sentiment score with additional extracted features in model (2.23), the correlation coefficient further increased, and prediction errors decreased from model (2.22). This indicates that the significant aspects that impact the rating differences of chain restaurants play a major role in rating prediction as well.

In terms of using aggregated yearly average data two for building predictive models and predicting cumulative chain restaurant Yelp ratings, as demonstrated in Table

2.8, the performances of six models are overall better than the models using data one. The correlation coefficients were over 0.1, and MAE reduced to at most 0.6182. Similarly, by incorporating attributes such as price range, delivery, credit card, and kids in addition to existing review number and city variables in model (2.19), the model increased the correlation coefficient to 0.2857 and decreased the errors by around 4%. At this time, when only use the sentiment score of all various aspects or just significant aspects in model (2.20) and model (2.21), the correlation coefficient between aggregated rating prediction increased from 0.2857 to 0.3245 and to 0.3125, in comparison to model (2.19). The errors generated by the four error measures decreased by at least 1%, in comparison to model (2.19). When combining the sentiment scores from all aspects with extracted attributes in model (2.22), the correlation coefficient increased to 0.3898, and the errors decreased by an additional 3-8%, in comparison to the four previous models. By associating only significant aspects' sentiment score with extracted features in model (2.23), the correlation coefficient and prediction errors were comparable with model (2.22).

Overall, the performance of Yelp rating prediction can be improved by incorporating time-invariant attributes of restaurants such as review number, city, and price range. By only including sentiment scores from available aspects and only the significant aspects, the prediction performances of the models were comparable to prior models in data one and were better in aggregated data two. Instead of adopting all aspects' sentiment scores, only sentiment scores from significant aspects including "food," "service," "drink and dessert," "value," "location" and "general" as predictors generated effectively good and even better prediction results.

2.7 Conclusion and Discussion

Understanding the factors that influence the rating differences from chain restaurants operated by the same company is important to improve chain restaurants' performance and enhance the sales of the whole restaurant industry (Luca, 2016). By integrating both text mining and network analysis, our study identifies the important aspects that significantly influence the rating differences, which has not been analyzed before. The significant aspects together with the extracted sentiment scores are further utilized for future Yelp rating prediction. To recognize the significant features affecting Yelp ratings, our study first aims at discovering the representative topics/aspects from unstructured user-generated reviews. We adopt the local LDA model that treats individual sentences as documents in order to avoid extracting general topics such as brand names and product names. To determine the optimal aspect number, we employ the cluster validation scheme to select the number of aspects with consistent performance. Our study then tries to discover the adjectives and verbs that can modify nouns and contain sentiment polarities. After constructing a word graph network connecting words with similar word embedding, and by proposing an objective function on the basis of label propagation method, we assign each extracted adjective and verb with a sentiment polarity probability distribution. The sentiment scores for individual aspects are further integrated from representative nouns modified by adjectives and verbs with sentiment probability distributions. Food quality, service, location, value, drink and dessert, and general opinion are identified as the significant impactors of the rating difference by using three different linear regression models. Furthermore, we employ predictive regressions for future Yelp rating prediction. The predictive models incorporating the sentiment scores from significant aspects as predictors achieve comparable lower error measures than including all possible aspects' sentiment score. By further extracting restaurant-related attributes from Yelp such as price range, credit cards, kids, and group as additional features for constructing predicative models, we can obtain the best performing models for rating prediction.

Besides the academic value shown here, the sentiment score of each aspect for restaurants also provides important practical implications for the chain restaurants. For example, knowing the sentiment score of drink and dessert aspect from specific Applebee's restaurant is lower than the average sentiment score of this certain aspect from all Applebee's restaurants, restaurant managers can hire new bakers with better performance so as to improve the performance of this particular aspect. Training programs can also be designated for employees if the restaurant knows it is the service that pulls the rating back.

Our study also suffers from a few limitations. We highlight several important ones and potential future work in the following. First of all, we only evaluate the top 81 chain restaurants' reviews and ratings from Yelp.com. Limited reviews are extracted from certain restaurants. The data sparsity problem might affect our model performance and further empirical analysis. In addition, Yelp.com is the only review website from which we extract ratings and reviews. The rating differences of chain restaurants might result from other factors if we use alternative business review websites such as TripAdvisor, Yahoo Local, and Angie's List. It is therefore important to extend the variety of data sources. Furthermore, even though the aspect sentiment analysis is an unsupervised learning approach, our study does not require any supervised data at the model-building phase. Labeled data of aspects and sentiment scores for model evaluation is still necessary. There are possibilities that the performance of our proposed aspect sentiment model is not good. Therefore, the model performance can be improved if the method can be evaluated by labeled data.

Figure 2.1: Rating Differences of Chain Restaurants Under the Same Brand

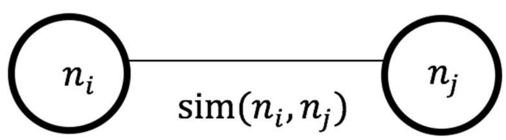

Figure 2.2: Node and Edge Example From Network

Table 2.1: Inferred Aspects With Representation Words

Manually Aspects	Representative Words
Food	sandwich, salad, bread, bagel, soup, cheese, egg, chicken, bacon, lettuce, tomato, turkey, burger, fries, onion, sauce, dog, shrimp;
Service	order, staff, customer, employee, manager, waitress, time, minute, waiter, wait, delivery, complaint, attitude, smile, drive-thru, online;
Drink and dessert	coffee, ice, cream, tea, chocolate, water, soda, cake, juice, cookie, smoothie, milk, cheesecake, coke, vanilla, latte, lemonade, fruit;
Value	option, portion, choice, deal, size, selection, ingredient, variety, calorie, amount, combo, special, gluten, addition, bit, picture, side;
Environment	area, room, bar, table, bathroom, floor, seating, hooters, music, booth, chair, pool, wall, door, patio, casino, hotel, seat, heat, screen, light;
Price	money, card, coupon, tip, gift, dollar, credit, receipt, change, charge, cent, discount, cash, bill, reward, deal, refund, free, fee, total, offer;
Location	vegas, lot, parking, strip, hotel, street, area, shop, mall, airport, center, corner, station, park, town, spot, south, valley, coast, downtown;
Anecdotes	lunch, night, friend, dinner, family, Friday, birthday, wife, husband, Sunday, Saturday, fun, group, weekend, mom, daughter, boyfriend;
General	experience, review, visit, subway, Starbucks, john, Wendy, Panera, outback, jimmy, restaurant, star, chain, quality, people, denny, joint;

Table 2.2: Example of Word Seeds and Representative Sentiment Vectors

Words	Polarity	Sentiment vector
respect, vigilant, qualified, supportive, easygoing, comfortable	positive	<1, 0>
hazardous, afraid, argumentative, slug, disturbed, bore, phony	negative	<0, 1>
classical, tasted, absorbent, honorary, optimize, recycled	unknown	<0.5, 0.5>

Table 2.3: Results of Linear Regression Models

DV: rateDiff	Model (2.14):	Model (2.16):	Model (2.17):
Food Diff	0.4339*** (0.0936)	0.4587*** (0.0646)	0.4588*** (0.0645)
Service Diff	1.197*** (0.1155)	0.9512*** (0.0571)	0.9588*** (0.0569)
Drink dessert Diff	0.2383** (0.0896)	0.2121** (0.0745)	0.2386** (0.0738)
Value Diff	0.1337 (0.0828)	0.1628** (0.0615)	0.9463** (0.0614)
Environment Diff	-0.099 (0.0841)	0.038 (0.0669)	0.0429 (0.0664)
Anecdote Diff	0.0201 (0.0994)	-0.0817 (0.0614)	-0.0832 (0.0613)
Price Diff	0.0444 (0.0947)	0.1072 (0.0976)	0.1209 (0.0974)
Location Diff	0.1295 (0.0852)	0.1657* (0.0699)	
General Diff	0.9659*** (0.1179)	0.9288*** (0.0589)	0.9463*** (0.0584)
Time FE	No	Yes	Yes
Location FE	No	No	Yes
R-squared	0.1511	0.0996	0.1032
N	2,308	9,958	9,958

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.4: Examples of Two-Panel Data Sets

Data Set	Target Variable	Retrieved Predictors
Data 1	Yelp rating of year 2006	Sentiment scores from reviews during 2005
	Yelp rating of year 2007	Sentiment scores from reviews during 2006
	…	…
Data 2	Integrated average Yelp rating of year 2006	Integrated average sentiment scores from reviews during 2005
	Integrated average Yelp rating of year 2007	Integrated average sentiment scores from reviews during 2005 and 2006
	…	…

Table 2.5: Description and Categories of Additional Features

Feature Name	Description	Categories
PriceRange	The price range of the food in the restaurant	Three class levels: {Low, Med, High}
Delivery	Whether the restaurant has delivery service	Two class levels: {Yes, No}
CreditCard	Whether the restaurant allows credit card	Two class levels: {Yes, No}
City	The city the restaurant located	Six class levels: {Madison, Las Vegas, Phoenix, Champaign, Pittsburgh, Charlotte}
Kids	Whether the restaurant is good for kids	Two class levels: {Yes, No}
Group	Whether the restaurant is good for group	Two class levels: {Yes, No}
ReviewNum	The restaurant's average yearly and integrated yearly review numbers	Yearly average range: [1,190] Integrated Yearly average range: [1, 710]

Table 2.6: Predictor Sets of Six Prediction Models

Model	Predictors
2.18	City, ReviewNum
2.19	City, ReviewNum, PriceRange, Delivery, CreditCard, Kids, Group,
2.20	City, foodscore, drinkdessertscore, valuescore, servicescore, enviornmentscore, anecdotescore, pricescore, generalscore
2.21	City, foodscore, drinkdessertscore, valuescore, servicescore, generalscore
2.22	City, ReviewNum, PriceRange, Delivery, CreditCard, Kids, Group, foodscore, drinkdessertscore, valuescore, servicescore, enviornmentscore, anecdotescore, pricescore, generalscore
2.23	City, ReviewNum, PriceRange, Delivery, CreditCard, Kids, Group, foodscore, drinkdessertscore, valuescore, servicescore, generalscore

Table 2.7: Model Performance Results With Yearly Average Data One

Model	Correlation Coefficient	MAE	RMSE	RAE	RRSE
2.18	0.0855	0.8521	1.0502	99.9298%	99.6207%
2.19	0.2036	0.836	1.0321	98.0366%	97.9051%
2.20	0.1727	0.8435	1.0383	98.9133%	98.4886%
2.21	0.1507	0.8468	1.042	99.3101%	98.8462%
2.22	0.2389	0.8281	1.0237	97.1159%	97.1088%
2.23	0.2729	0.8199	1.0142	96.1491%	96.2051%

Table 2.8: Model Performance Results With Yearly Aggregated Data Two

Model	Correlation Coefficient	MAE	RMSE	RAE	RRSE
2.18	0.1071	0.6182	0.7726	98.8416%	99.4164%
2.19	0.2857	0.5931	0.7447	94.8345%	95.8273%
2.20	0.3245	0.5828	0.735	93.1913%	94.5739%
2.21	0.3125	0.5853	0.7381	93.5917%	94.9772%
2.22	0.3898	0.5654	0.7156	90.3965%	92.0743%
2.23	0.3798	0.5678	0.7188	90.7729%	92.4887%

CHAPTER 3

ONLINE VIDEO ADVERTISEMENTS: HOW DOES TEMPORAL REMINDER DESIGN AFFECT USER'S OPINIONS?

3.1 Introduction

An online advertisement video is a video clip about a commodity or a brand, which aims at delivering promotional marketing messages to consumers. It is placed in different forms such as pre-roll (before the actual desired video content), mid-roll (between the actual desired video content), or post-roll (after the actual desired video content) in terms of different appearances and time durations (Mei et al., 2007). The length of the majority of video advertisements is from 15 seconds up to a few minutes. The online video currently is the most rapidly growing advertising format. It is estimated that 74% of all Internet traffic is video based by 2017 and video ads account for 35.6 % of total videos viewed.

Given the increasing trend of marketing in online video advertising format, advertisement sponsors intend to receive the maximum profit out of the advertisement. Aiming at further advocating their brands and promoting the commodities through video advertising, sponsors want viewers to not only generate positive attitudes toward their advertisements, but also recall the advertisement contents after viewing the video.

However, because viewers are goal-oriented and advertisements can interrupt their goal, the attitudes toward the online advertisements and social effects of advertising are generally negative and aggravating (Krugman, 1983; Previte & Forrest, 1998). Advertisement sponsors, therefore, lose the opportunity to communicate their message to customers (Teixeira, 2014).

Due to the ongoing trend of advertisement avoidance and opposition, viewers treat the duration of video advertisements as a negative waiting process before they access their desired video contents. Waiting on the Internet has been recognized as a commonly experienced problem and has been studied in a large number of studies (Ceaparu, Lazar, Bessiere, Robinson, & Shneiderman, 2004; Rose, Meuter, & Curran, 2005). Some studies suggest providing indicator bars, a clock, or visual contents can change people's perception of waiting duration (Dellaert & Kahn, 1999; Groth & Gilliland, 2006; Hui & Tse, 1996; Whiting & Donthu, 2006).

Online video advertising typically is presented to viewers with a few designs. During the online video advertising, several temporal reminder designs such as counting down clock, progressing bar, and remaining number of sponsors are standard time reminders notifying the time and duration. They can appear in various formats as demonstrated in Figures 3.1, 3.2, and 3.3. Besides focusing on the contents, advertisement video designers can also determine whether to embed a few add-on features so as to reduce the perception of online advertisement video duration, decrease the adverse attitudes toward the advertisement video, and increase the video content recall.

In this study, we apply time perception theories, reactance theories, and selective attention theories to examine how online users' perception of advertisement duration, negative attitudes, and contents recall from shorter and longer online video advertisements are influenced by providing different time reminder designs including counting down clock, progressing bar, and the remaining number of sponsors. Temporal reminders such as counting down clock and progress bar have been discussed before for the impact on traditional wait perception such as waiting in line physically and waiting on the phone or online waiting (Dellaert & Kahn, 1999; Groth & Gilliland, 2006; Hui & Tse, 1996; Whiting & Donthu, 2006). Regarding the attitudes toward advertisements, most prior studies analyze online advertisement attitudes from the content aspects of the advertisement. The controversy about the length of the video affecting negative attitude toward online advertisement videos has not been fully understood yet (Goodrich, Schiller, & Galletta, 2015; Kusse, 2013). Besides, temporal reminders such as counting down clock and progress bar affecting the perception and attitude toward online ad videos have not been analyzed yet in the prior research. Moreover, previous researchers examine the features of advertisements that promote the effectiveness of content perception (Goldfarb & Tucker, 2011) from various aspects including objective and subjective features in general. However, the features that hinder the effective recall of advertisement have not been widely discussed yet.

At this moment, video advertising is a relatively new advertising format (Kusse, 2013). Even though it shows similarities to other online ad formats like pop-ups, banners, and traditional media ads such as television commercials, there are a large number of unique designs incorporated in video advertising that have not been touched yet. For example, the additional temporal reminder designs of the online video advertisement that impact the recall and effectiveness of the advertisements have not been studied yet. The perceived advertisement designs that influence the duration of the perception, attitude, and information recall have not been examined in the prior research.

Based on the theories, we address the following research questions in this study: How does (1) counting down clock, (2) progressing bar, and (3) number of remaining sponsors affect online users' perception of advertisement duration, their attitude, and message recall? We further analyze: (4) how does the length of the online video advertisement change the impact of temporal designs on the perception of ad duration time, the attitude, and message recall? One experimental study is conducted to answer the above questions.

The remainder of the study is organized as follows. We first review related studies and highlight key differences between our study and previous representative research. We then formally propose the theoretical foundations including time perception theories, reactance theories, and selective attention theories of this research. Furthermore, we hypothesize a few hypotheses based on the existing literature and theories. We describe one experimental study with research design, data collection and analysis, and discussion of the results. Finally, we conclude the chapter with contributions and suggestions for future studies.

3.2 Literature Review

Several streams of research are relevant to our study, including advertisement and wait duration perception, attitudes toward advertisement, and advertisement content recall. In this section, we review representative studies in each stream and highlight the gaps that motivate our research.

3.2.1 Advertisement and Wait Duration Perception

Due to the ongoing trend of advertisement avoidance and opposition, viewers can treat the duration of video ad as a negative waiting process before they actually access their desired video contents. The perceived quickness and length of the wait is the cognitive aspect of the waiting experience that captures customers' loss of a very valuable time asset (Hong, Hess, & Hardin, 2012).

There are two types of designs provided during the waiting under both traditional waiting circumstances and the online waiting environment. In the traditional waiting environment, such as waiting physically in line or waiting on the phone, by providing temporal information, such as waiting duration information (Chebat, Salem, Poirier, & Gélinas-Chebat, 2010; Groth & Gilliland, 2006) and queuing information (Antonides, Verhoef, & Van Aalst, 2002; Whiting & Donthu, 2006), users perceive the estimation of the remaining duration of the wait. In an online waiting environment, a moving progressing bar is provided to serve the same purpose as the temporal information. On the other hand, nontemporal information, such as music, TV, and electronic news board (Cameron, Baker, Peterson, & Braunsberger, 2003; Katz, Larson, & Larson, 1991; Pruyn & Smidts, 1998) is used for distracting users' attention from the waiting duration.

The impacts of the temporal and nontemporal information on the perceived wait duration are not consistent though. Temporal information makes users perceive longer waiting (Chebat et al. 2010), shorter waiting (Dellaert & Kahn, 1999; Lee, Chen, & Ilie, 2012), and both shorter and longer waiting (Antonides et al., 2002; Hui & Tse, 1996;

Katz et al., 1991). The alternative reason for inconsistent results is not distinguishing the waiting length during the wait (Antonides et al., 2002; Hui & Tse, 1996; Katz et al., 1991). Hong et al. (2013) distinguish the length of waiting and evaluate the perception of wait given both temporal and nontemporal information. They discover that nontemporal information can reduce the waiting perception during the longer wait and increase the perception during the shorter wait, which verifies the paradigm shift of time perception theories.

The perceived duration of advertisements has been studied much less than waiting duration. Similar to providing nontemporal information, Mailov (2012) compares the impact of music and music tempo during advertisement videos. The result demonstrates that the nontemporal information music reduces the recall of advertisement content. Fast tempo music decreases the perceived ad video duration.

3.2.2 Attitudes Toward Advertisement

There are mixed findings regarding the intrusiveness of advertisements with different length. Goodrich et al. (2015) observe fewer intrusiveness responses from longer ads. Controversially, the perceived intrusiveness is marginally significantly lower toward shorter video advertisements than longer ones (Kusse, 2013). Kusse (2013) compares a 15-second video advertisement with the 60-second ad and concludes the 15-second advertisement is felt less intrusive than 60-second. Goodrich et al. (2015) first assume the longer ad is more intrusive than the shorter advertisement. Then they compare ads with lengths from 3, 8, 15, to 30 seconds. They discover the intrusiveness decreases as the length increases from 8 seconds to 30 seconds. The 15-second video that is considered as a short advertisement (Kusse, 2013) is considered as a medium-long advertisement in the research Goodrich et al. (2015) proposed.

Besides length, studies also focus on the impact of the ad content on the attitudes. For example, the factors such as entertainment, informativeness, irritation, credibility, and relevant demography (Azeem & ul Haq, 2012; Brackett & Carr, 2001), type of advertisement, personalized level (e.g., privacy invasive), control design (e.g., skip the video ad option) (Hegner et al., 2016; Kusse, 2013), emotional appeal (Hegner et al., 2016), interactivity, economy, and value (Azeem & ul Haq, 2012) are evaluated for the perceived attitudes effect on advertisements. Specifically, the personalized level (e.g., privacy invasive), control design (e.g., skip the video ad option) (Hegner et al., 2016; Kusse, 2013), and emotional appeal (Hegner et al., 2016) are particularly evaluated for online video advertisements. Whether video advertisements are skippable is significantly related to the attitudes toward the ads (Hegner et al., 2016; Kusse, 2013). The emotional appeal refers to the positive emotions when their personal interests are concerned (Teixeira, Wedel, & Pieters, 2012). When the advertisement video is perceived as valuable, users elicit less irritation and avoidance (Hegner et al., 2016).

3.2.3 Advertisement Effectiveness

How do users perceive and process online advertisement contents? Advertisement information processing is based on the function of the Internet and the structure of the Internet advertisement. The structure of the advertisements is manipulated by advertisement designers. The factors such as advertisement type, format, objective features, and subjective features are primary components that direct the perception process. Objective ad features include color, size, animation, length (Blair & Kuse, 2004; McDonald, 1997), the format such as in-line or pop-up ads (McCoy et al., 2004), and product type and so on. Subjective ad features include perceived excitement, users' interest, attitude toward the advertisement (Azeem & ul Haq, 2012; Mehta, 2000), interactivity, informativeness (Ducoffe, 1995), entertainment, credibility, economy, value, and content matching (Ducoffe, 1995; Goldfarb & Tucker, 2011).

Specifically, present studies measure several unique features for online advertisement videos. Teixeira et al. (2012) assess joy and surprise of ad videos and their impact on user's concentration and attention. Different lengths and placement of online advertisements affect completion rate. A 15-second ad is 2.9% more likely to be completed than a 20-second advertisement, while a 20-second advertisement is 3.9% more likely to be finished than a 30-second advertisement. A mid-roll ad is 18.1% more likely to be completed than a pre-roll, and a pre-roll is 14.3% more likely to be finished than a post-roll (Krishnan & Sitaraman, 2013).

3.2.4 Research Gap Summary

As we discussed in the introduction, video advertising is a relatively new advertising format (Kusse 2013). There are various unique designs incorporated in video-advertising that have not been touched yet. For example, the impact of temporal and nontemporal information is considered and measured in waiting duration perception studies (Cameron et al., 2003; Hui, Dube, & Chebat, 1997; Katz et al., 1991; Pruyn & Smidts, 1998). In comparison to the existing advertisement-related studies, little attention has been paid on the duration perception of the advertisements, especially for online video ads. Apart from the content level design of the video (Maliov, 2012), motivated by waiting duration perception studies, the available temporal reminder designs such as counting down clock, progressing bar, and remaining number of advertisements are important features that influence the received video duration. The reason for inconsistent results of temporal information effects is the different waiting lengths (Antonides et al., 2002; Hui & Tse, 1996; Katz et al., 1991). Our study, therefore, examines the effects of different temporal reminder designs on perceived advertisement video duration, moderated by the actual length of the video.

In terms of the attitudes and effectiveness of advertisements, prior studies analyze the factors at the video content design level as well, such as interactivity, informativeness, entertainment, credibility, economy, value, and content matching. Particularly for the evaluation of intrusiveness and effectiveness of online advertisement videos, the control design skip ad button (Hegner et al., 2016; Kusse, 2013) and video placement (Krishnan & Sitaraman, 2013) are the only features designated and measured.

3.3 Theory Foundation

3.3.1 Time Perception Theories

The two conflict models regarding time perception are memory-based models and attention-based models. Memory-based models propose that during a time interval, the more information cues that users remember, the longer they estimate time duration (Ornstein, 1969). On the opposite side, attention-based models suggest there is a cognitive timer in mind that people use to track the passage of time. When additional information that is not related to a temporal reminder requires people's attention, people spend less attention on the cognitive timer and passage of time. Therefore, they estimate shorter time duration (McClain, 1983; Zakay, 1993). Time is linearly but negatively associated with the experimental complexity with which given intervals are filled (Priestly, 1968).

One reconciliation between the two models is the resource-allocation model (RAM; Kahneman, 1973; Zakay, 1989). The RAM assumes people's central attention has limited capacity and the attention allocates to different tasks all the time (Kahneman, 1973). Two paradigms are proposed: prospective paradigm and retrospective paradigm (Block & Zakay, 1997). Users being notified to estimate time passage before experiencing the wait is known as the prospective paradigm. In contrast, users being requested to approximate the time duration after the wait is considered as the retrospective paradigm (Block & Zakay, 1997).

During the prospective paradigm, when the majority of users' attention capacity is used for estimating the time, nontemporal information such as music is able to distract users' attention from the cognitive timer, therefore reducing the estimation of time passage. During the prospective paradigm, attention-based models dominate the waiting process. In contrast, when a user is asked to evaluate the duration after the occurrence of the wait, any information or memory cues can lead to the increase of waiting duration estimation. Memory-based models dominate the waiting procedure during the retrospective paradigm. More time reminders during the wait increase the memory cues. As a result, longer perceived waiting time is assessed.

Because users are not requested to evaluate the wait before the actual waiting in most of the cases, the retrospective paradigm takes place in most of the waiting environments. Hong et al. (2013) propose a paradigm shift within a longer wait. During the shorter wait, as mentioned in RAM (Block & Zakay, 1997), the retrospective paradigm and memory-based model dominate the waiting evaluation. Throughout the longer wait, people start to wonder about the amount of time they have already waited, question the additional time they still need to wait, and determine if they should continue to wait. At this point, the retrospective paradigm shifts to the prospective paradigm with the attention-based model determining the waiting perception. Therefore, any nontemporal information that distracts people's attention away from the wait reduces the duration perception more than any temporal information that directs people's attention toward time (Hong et al., 2013).

3.3.2 Reactance Theory

Reactance theory (Brehm, 1966) proposes that a perceived threat or restriction to expected freedom can lead to a motivational state directed at engaging in the threatened free behavior. This theory indicates the fact that users have certain freedoms to their behavior. If these behavioral freedoms are reduced or threatened, the individual wants to regain them. Because online video advertisements are a goal impediment for users, users can perceive the loss of freedom and control. Therefore, online users increasingly start to avoid advertisements and generate an overall negative attitude toward the ads.

3.3.3 Selective Attention Theories

The Broadbent's Filter Model proposes the scenario (Broadben, 1958) that users have a certain capacity of attention to process information. In order to select information to process, users utilize a filter to determine which information to pay attention to. Unlike Broadbent's Filter Model (Broadbent, 1958), memory selection models pass both attended and unattended message through the initial filter. Both messages are then sorted at a second-level filter based on the actual meaning of the message's contents. Attended messages can then further be stored in users' short-term memory.

3.4 Research Model

3.4.1 Advertisement Duration Perception

When comparing the duration perception of advertisement videos with different lengths, corresponding to memory-based time perception models (Ornstein, 1969), longer advertisements can provide additionally more memory cues than shorter advertisements. In line with attention-based time perception models (McClain, 1983; Zakay, 1993), there is a cognitive timer in their mind that people use to track the passage of time. Therefore, more memory cues can result in additional estimation of time duration. People track the longer passage of time during the longer advertisement than the shorter advertisement. In addition, there is a general consensus that the actual waiting time significantly influences the perceived waiting duration (Dabholkar & Sheng, 2008). We first hypothesize that people perceive a longer waiting period during the longer advertisement video than during the shorter advertisement.

H1: People perceive a longer video duration during the longer advertisement video than during the shorter advertisement video.

Under ordinary online circumstances, users' estimation of duration falls into the retrospective paradigm. They are not forced to estimate the duration before watching the video. Based on the resource allocation model (Block & Zakay, 1997), for the retrospective paradigm, the memory-based time perception model (Ornstein, 1969) is dominating the estimation process, and additional nontemporal information can increase the estimate of time. This is consistent with the prior studies that nontemporal information such as music (Chebat et al., 2010; Hui et al., 1997) negatively affects waiting.

Corresponding to the attention-based model (McClain, 1983; Zakay, 1993), the resource allocation model (Block & Zakay, 1997), and the paradigm shift proposition (Hong et al., 2013), a paradigm shift occurs from the retrospective paradigm to the prospective paradigm during the longer waiting environment. Users start to wonder about the passage of waiting time and question the waiting environment (Hong et al., 2013). For the prospective paradigm, the attention-based model and temporal information are leading the waiting evaluation so that nontemporal information that distracts attention from time passage estimation can reduce the approximation of the waiting length. This is in line with another collection of studies that nontemporal information, such as music (Cameron et al., 2003), TV (Pruyn & Smidts, 1998) and an electronic news board (Katz et al., 1991), negatively affects waiting and managing consumer wait perceptions.

There are three different temporal designs we discuss in this study. The first temporal design is the counting down timer. A clock embedded in the online video advertisements can count down as the advertisement video proceeds. The remaining time presented to viewers serves as a temporal reminder prompting the remaining video duration as displayed in Figure 3.1. The second reminder design is the progressing bar indicator. The progressing bar indicator visually indicates the proportion of video completed as demonstrated in Figure 3.2. The last design is the remaining number of advertisements reminder. Sponsors and video website operators select a banner presenting the remaining number of sponsored ads as the reminder shown in Figure 3.3. Remaining advertisement number information allows people to estimate the remaining duration of the advertisement video roughly. However, without providing the time for each remaining advertisement, it is hard to calculate the remaining advertisement period accurately. On the other hand, remaining ad number information can serve as a distractor converting users' attention to the content of the advertisement.

A shorter video ad can be considered as a waiting period with shorter duration. According to RAM (Block & Zakay, 1997), for certain online circumstances, users can focus on their own tasks instead of estimating the duration of the video before watching. This indicates the retrospective paradigm during the video. In the retrospective paradigm, the memory-based model (Ornstein, 1969) dominates the waiting perception. Based on the model (Ornstein, 1969), by giving an additional temporal reminder to users, such as counting down clock, progressing bar indicator, and remaining number of ad reminder, users can receive more temporal memories than with no temporal reminder. More memory cues can result in a longer perception of video duration. We propose that during the shorter online video advertisement, providing temporal reminder designs such as counting down clock, progressing bar indicator, and remaining number of ad reminder can increase the perceived video duration.

H2a: For a shorter video advertisement, temporal reminder information such as counting down clock, progressing bar indicator, and remaining number of advertisements reminder can result in a longer perceived advertisement duration than without any information.

During the longer video advertisement, giving temporal information such as counting down clock, progressing bar indicator, and remaining number of ads can arouse a temporal stimulus that can motivate paradigm shift (Hong et al., 2013). After a paradigm shift, the retrospective paradigm can shift to the prospective paradigm. The evaluation of the time passage of the video becomes the primary task of users (Block & Zakay, 1997). In the prospective paradigm, the attention-based model (McClain, 1983, Zakay, 1993) and temporal information control the overall duration estimation. If nontemporal information such as music, pictures, and TV is given to users and distracts users' attention from the estimate of time, the perceived duration can be reduced (Cameron et al., 2003; Katz et al., 1991; Pruyn & Smidts, 1998). Different from online waiting and the traditional waiting environment (Cameron et al., 2003; Hui et al., 1997; Katz et al., 1991; Pruyn & Smidts, 1998), the content of an online advertisement video is combined with music, graphics, and animation (Ducoffe, 1995; Hoffman & Novak, 1996). Any cues that direct people's attention back to the video can also be considered as nontemporal information that reduces the duration perception. However, giving temporal information such as counting down clock and progressing bar indicator, in comparison to no temporal information, these temporal reminder designs can lead users' attention toward time, which result in overestimating duration time. Therefore, we hypothesize that temporal reminder designs such as counting down clock and progressing bar indicator during the longer advertisement duration can positively affect the perceived video duration.

H2b: For a longer video advertisement, temporal reminder information such as counting down clock and progressing bar indicator can result in a longer perceived advertisement duration than without any information.

In terms of the remaining number of advertisements reminder, it provides an estimation of remaining advertisement duration. More importantly, it directs users' attention back to the ad video by displaying the number of advertisements "advertisement n out of m." This reminder, therefore, can also serve as a nontemporal cue and lead more users' attention toward nontemporal information than without any reminder designs. Therefore, we hypothesize that temporal reminder designs such as the remaining number of advertisement reminder, during the longer ads duration, can negatively affect the perceived video duration more than no temporal design.

H2c: For a longer video advertisement, temporal reminder information such as the remaining number of advertisements reminder can result in a shorter perceived advertisement duration than without any information.

Furthermore, we compare the video duration assessment from different temporal reminder designs. For shorter advertisement videos, based on the memory-based model (Ornstein, 1969), all temporal reminders can increase additional temporal memory cues, which lead to longer video duration perception than no reminder. During the longer advertisement video, when there is a paradigm shift (Hong et al., 2013), the attention-based model (McClain, 1983, Zakay, 1993) dominates the video estimation. In line with the attention-based model (McClain, 1983, Zakay, 1993), additional nontemporal information such as pictures and music can reduce the duration perception. As we discussed above, because an advertisement video consists of animations, graphics, and music (Ducoffe, 1995), the video can also be considered as nontemporal information.

Any indicators that direct users' attention to advertisement videos can distract users' attention from passage of time, and therefore can reduce the overall perception.

In comparison to counting down clock and progressing bar indicator that direct users' attention toward time, remaining number of advertisements is a mixed reminder. On the one hand, it provides estimation of remaining advertisement duration. On the other hand, it leads users' attention back to the advertisement video. Therefore, we hypothesize that temporal reminder designs such as remaining number of advertisement reminder, during the longer ad duration, can negatively affect the perceived video duration more than reminders such as counting down clock and progressing bar indicator.

H3: For longer video advertisements, the remaining number of advertisements reminder can result in shorter perceived advertisement durations than temporal information such as counting down clock and progressing bar indicator.

3.4.2 Negative Attitudes Toward Advertisements

According to Osuna (1985), waiting can be detrimental to users, not only from the time loss perspective but also because of psychological costs due to stress and frustration. Besides, online and traditional waiting is described as frustrating, annoying, and aggravating (Dube, Schmitt, & Leclerc, 1991; Gardner, 1985; Katz et al., 1991) for user experience. A longer advertisement video can be considered as a longer waiting duration, which can increase the frustrating, annoying, and aggravating level perceived by users. In line with psychological reactance theory (Brehm, 1996), because inserted advertisements interrupt users' behavior freedom, general attitudes toward advertisements are negative (Krugman, 1983). A longer advertisement video can plunder users' behavior freedom longer than a shorter video, which can result in additional negative and aggravating attitudes. There are compelling reasons to investigate the potential of short-form advertisements to lessen the consumer annoyance factor. Therefore, we hypothesize people generate more negative attitudes toward the longer advertisement video than toward the shorter ad.

H4: Longer online video advertisements can result in more negative attitudes toward the online video advertisements than shorter online video advertisements.

On the one hand, waiting duration information is considered to reduce the uncertainty and frustration of the waiting, therefore enhance the evaluation of the wait (Dellaert & Kahn, 1999; Hui & Tse, 1996). On the other hand, waiting duration information reminds people of the time and converts attention back to the length of the wait. A longer period of the wait can arouse more negative attitude toward the waiting process (Hong et al., 2013; Houston, Bettencourt, & Wenger, 1998; Katz et al., 1991). Therefore, some of the waiting duration information does not affect the wait evaluation (Antonides et al., 2002; Chebat et al., 2010; Katz et al., 1991).

These two potential effects are in opposite directions and can counteract one another (e.g., a temporal reminder can make the wait feel longer, but also comfort people by reducing the uncertainty of the waiting).

For shorter advertisements, the negative attitudes toward the advertisements are hypothesized to be less severe than the longer ads (Dube et al., 1991; Gardner, 1985; Katz et al., 1991). During the shorter advertisement, when the memory-based model (Ornstein, 1969) dominates the duration estimation perception, temporal reminder designs can increase the temporal perception, which lead people's attention to time and increase additional temporal memory. Further attention to time and temporal memories can increase the perception of waiting time duration (Ornstein, 1969). Increasing perception of waiting time can generate more negative attitudes toward the advertisement and the waiting duration, which is consistent with psychological reactance theory (Brehm, 1996) and related studies that longer queue waits can lead to negative affective responses (Houston et al., 1998; Hui & Tse, 1996). Therefore, we hypothesize people generate more negative attitudes toward the shorter advertisement videos with temporal reminder designs than toward shorter ads with no temporal reminders.

H5a: For a shorter advertisement, the temporal reminder designs such as counting down clock, progressing bar indicator, and remaining number of advertisement reminder can result in more negative attitudes toward the advertisement than an advertisement without any reminder.

For longer advertisements, the negative attitudes toward the advertisements are hypothesized to be more severe than the shorter advertisements (Dube et al., 1991; Gardner, 1985; Katz et al., 1991). During the longer advertisement, there is a paradigm shift where the attention-based model directs the waiting assessment (Hong et al., 2013). Temporal reminder designs can increase the temporal perception, which leads people's attention toward time duration and increases additional temporal memory (Block & Zakay, 1997; McClain, 1983; Zakay, 1993). Additional attention to time can increase the perception of waiting time duration. Increasing perception of waiting time can generate more negative attitudes toward the advertisement and the waiting duration (Brehm, 1996; Krugman, 1983; Osuna, 1985).

However, during the longer advertisement, people perceive incremented unsureness, uncertainty, and frustration (Dellaert & Kahn, 1999; Hui & Tse, 1996) when they do not receive the information of how much longer they have to stay with the advertisement. Temporal reminder designs such as counting down clock and progressing bar indicator can serve as an important temporal reminder that reduces the uncertainty and unsureness of the waiting for advertisement. As for the remaining number of advertisements reminder, we analyze and assume it can serve as both temporal information and nontemporal information. Even though it does not provide accurate remaining time estimation, it still reduces the uncertainty and frustration by providing the approximated remaining time duration. Furthermore, it can additionally distract people's attention from time passage to advertisements, and therefore reduces the overall video duration perception. In reviewing the waiting filters-related studies, the majority findings verify that by giving a filler (i.e., temporal information and nontemporal information), regardless of whether the filler makes the wait even longer (Hui et al., 1997), shorter (Cameron et al., 2003), or no effect (Hui & Tse, 1996). It improves people's attitude toward the wait. Therefore, temporal reminder designs reduce the uncertainty and unsureness of the waiting for an advertisement. They further reduce the negative attitudes toward the advertisement.

H5b: For a longer advertisement, temporal reminder designs such as counting down clock, progressing bar indicator, and remaining number of advertisements reminder can result in less negative attitudes toward an advertisement than an advertisement without any reminder.

3.4.3 Contents Recall of Online Video Advertisements

Based on the attention theories (Eriksen & Hoffman, 1972), the temporal time reminder control can be recognized as distractors that distract the attention from the advertisement video.

We hypothesize the temporal time reminders can influence the attitudes toward ads in the prior section. Based on Mehta (2000), the attitudes toward ads can negatively affect the advertisement content recall and effectiveness. We also hypothesize the temporal time reminders can influence the perceived length of the ads. Based on Stanton and Burke (1998), people spontaneously recall more content with longer advertisements than shorter.

Based on the attention theories (Eriksen & Hoffman, 1972), the temporal time reminder control can be recognized as distractors that interfere with attention. From an advertiser's perspective, the contents that the advertiser wish customers to attend to, process, and memorize are the visual and audio. According to RAM (Kahneman, 1973), people's central attention has limited capacity. The attention can allocate to different tasks all the time. However, in line with the selective attention theories, the design components such as counting down clock, progress bar indicator, and the number of advertisements remaining are stimuli that can take up people's attention capacity (Broadbent, 1958). In addition, counting down clock and progress bar indicator can be sorted by the memory selection (Broadben, 1958) and direct the attention to time. The number of remaining advertisement can be arranged by the memory selection and directs the attention to both time and ad.

From customers' perspective, as the advertisement gets longer, in line with the attention-based model (Block & Zakay, 1997; McClain, 1983; Zakay, 1993) and paradigm shift (Hong et al., 2013), when the attention-based model and temporal information dominate the process, users pay more attention to the temporal reminder with shorter advertisements. Therefore, we hypothesize that the temporal reminder designs can distract users' attention from ad videos and therefore reduce the contents recall.

H6: For both longer and shorter advertisements, customers perceive and recall fewer advertisement contents with temporal information designs including counting down clock, progressing bar, and remaining advertisement reminder than without any reminder.

The research model is demonstrated in Figure 3.4. The hypotheses proposed in the research model are concluded in Table 3.1.

3.5 Experimental Study

One experimental study is proposed to examine the hypotheses of the impact of temporal designs on the perception of advertisement duration, negative attitude toward the advertisement, and advertisement content recall.

We conducted a controlled experiment with a 2 (actual length of advertisements: short/long) by 4 (temporal reminder design: no temporal design, counting down clock, progressing bar indicator, the remaining number of advertisements) full-factorial between-subject design. Subjects were Amazon Turk workers who have received at least 95% project approval rate and finished at least 50 projects before. We consider the use of Amazon Turk workers as subjects to be appropriate because these workers use the Internet for a long time each day. Because we are examining a usual psychological phenomenon (i.e., watching online video advertisements), people can give a similar reaction. Workers received $0.50 for participating.

3.5.1 Desired Video, Advertisement Video, and Task

An experimental video was designed for the study. We chose an NBA game collection video gathering best NBA shots as our desired video. The reason to choose a NBA game collection video as our designed video was that it was popular, had a significant amount of views, and online users could quickly get involved in the video. The total length of the NBA video was around 3 minutes and 30 seconds.

For the advertisement video, in order to evaluate the effects of the temporal reminder design for the remaining number of advertisements, we distinguished the short advertisement video and the long advertisement video by the total number of ads. In this case, the short ad video was composed of one advertisement video with a single brand and product, while the long advertisement video was a combination of several videos with different brands and products.

We provided workers with a cover story that they were watching an NBA game collection video. After watching the video without skipping and fast forwarding, several questions regarding the games would be asked. We also told them there was a manipulation check of whether they have finished the video. Submissions would be rejected if they failed the verification test.

3.5.2 Pilot Study

We conducted a pilot study with 72 workers to identify appropriate advertisement video duration embedded in the desired NBA collection video. We preselected four duration intervals from 15 seconds, 1 minute, 2 minutes, and 3 minutes based on the review of the literature. We used the similar cover story and asked the worker to watch the video with different length of advertisement videos. After they finished watching the video, we asked them to evaluate the perception of the advertisement duration. They believed they had watched a short, moderate, or long advertisement video. Based on the responses, we selected two intervals in the experiment, which were 15 seconds for the shorter advertisement, and 3 minutes for the longer advertisement.

3.5.3 Variable Manipulation

Actual advertisement duration was manipulated as short and long (15 seconds and 3 minutes). The short ad video was composed by a video promoting glass jars manufactured by the brand ball, while the long advertisement video was constituted by eight different advertisement videos with the first four of them lasting about 15 seconds each and rest of them lasting about 30 seconds. The brands and products promoted by the advertisements were randomly selected but were irrelevant to the NBA game video.

Different placements of online advertisement generate different completion rate (Krishnan & Sitaraman, 2013). Completion rate is highest when the advertisement video is mid-roll. A mid-roll ad is inserted between users' desired video contents. Completion of mid-roll ads is 18.1% higher than pre-roll ads, while completion of pre-roll ads is 14.3% higher than post-roll ads (Krishnan & Sitaraman, 2013). We, therefore, embedded the advertisement videos in the middle of the desired video to ensure a better advertisement completion rate. Since the total length of the desired video was around 3 ½ minutes, we inserted the advertisement video at around the 2 minutes time point.

For the control group that does not incorporate any temporal reminder design during the ad video, a simple line of text (i.e., "Please watch the following advertisement videos before continuing your video."), as shown in Figure 3.5, was displayed on the bottom of the video before proceeding to the desired video. For the experimental group with counting down clock as temporal reminder design, besides the line of text shown on the bottom of the video, a counting down clock with a blue background, as shown in Figure 3.6, was presented at the upper left corner of the advertisement video. If the video lasted 2 minutes and 58 seconds, it was counting down from "2:58" to "0:00" while the video was playing. For the temporal reminder design progressing bar indicator, besides the line of text, as demonstrated in Figure 3.7, a progressing bar indicator highlighting a horizontal bar from left to right was displayed on the bottom of the video. As the video proceeded, the progressing bar changed from completely dim to fully highlighted in a uniform velocity. The last temporal reminder design was the remaining number of advertisements. We indicated the number of ad videos in the format as shown in Figure 3.8. If in total we had eight advertisement videos, and we were currently playing the second video, there was an indicator presenting "2 of 8 advertisements" on the top of the video, which was consistent with many video websites such as ABC.com. As displayed in Figure 3.9, the number of indicators changed when a new advertisement video started. All desired video and video advertisements were derived from the original commercials and tailored to fit the experiment by editing the length and look of the video.

3.5.4 Variable Measured

All dependent variables were measured using existing scales from published studies. Specifically, the perceived advertisement duration was measured similar to Gorn, Chattopadhyay, Sengupta, and Tripathi (2004) with a one-item semantic differential scale. Negative attitudes toward the advertisement were measured by a four-item, seven-point Likert-type scale based on the existing instruments (Hong et al., 2013; Hui et al., 1997; Hui & Tse, 1996). Instead of using open-ended and closed-ended questions for querying user's advertisement product and brand recall (Jeong & King, 2010), we provided multiple choice questions asking workers to select the correct watched products and brands out of 15 different selections.

Additionally, two control variables, impatience and attribution, were also evaluated (Hong et al., 2013). Impatience was one of the most significant time-related individual difference factors that impacted the perception of time duration estimation (Francis-Smythe & Robertson,1999). Attribution of the advertisements was studied affecting online user's evaluation toward the performance of websites (Houston et al., 1998; Rose et al., 2005). Impatience was measured with three items and attribution was measured with two items using existing instruments based on the literature (Houston et al., 1998; Rose et al., 2005; Spence, Helmreich, & Pred, 1987).

Furthermore, to ensure workers finished the entire video without skipping, we inserted a screenshot with a number displaying for five seconds asking users to remember it as a manipulation check as shown in Figure 3.10. The duration of the desired video embedded with the long advertisement video was around 5 ½ minutes, while the one with the short advertisement video was around 3 minutes. The time duration workers used for finishing the task was another manipulation check.

3.5.5 Subjects and Procedure

A total of 1,368 subjects completed the experiment, with 77.57% being male, 22.03% being female, and 0.39% choosing rather not to tell. Amongst all workers, 51.51%, 39.87%, and 8.61% are between age 18 to 29, age 30 to 45, and over 45, respectively. A series of MANOVA tests showed that age and gender did not affect the dependent variables.

Workers were shown an instruction that contained a cover story. They were then directed to the NBA game video. After they clicked the start button, the NBA game collection video played. At around the 2 minutes time point, the desired video stopped. The advertisement video started with various temporal reminder designs or no design. After the advertisement video, the NBA game collection continued to play. Upon finishing up the desired video, the subjects were then redirected to an online survey measuring the perceived duration, negative attitudes toward advertisements, as well as the brand and product recall from the video. In addition, workers were asked to respond to additional questions regarding their general impatience, attribution, and demographic questions such as gender and age. The empirical study is summarized in Table 3.2.

3.6 Data Analysis Results

3.6.1 Manipulation and Control Checks

We tested the manipulation and control check and hypotheses testing. The data analysis results of study 1 are presented in the following.

We incorporated in total 1,368 Amazon Turk workers in our study 1. Based on our manipulation check, there were two criteria we used to discard the results by the subject. The first criterion was when workers failed to see the number embedded near the end of the advertisement video. The second condition was if the total time workers spent finishing the task was less than the video duration. After rejecting the results from subjects who failed the manipulation checks, 1,266 workers were kept in study 1. Amongst 1,266 subjects, 639 were assigned for finishing the video inserted with the longer advertisement, while 627 completed the video with the shorter advertisement. The statistics of subjects completing the different temporal reminder designs are displayed in Table 3.3.

Impatience and Attribution significantly related to perceived advertisement duration (F value = 3.785, p value = 0.01017; F value = 422.9, p value = 0.0000), negative attitudes toward advertisement (F value = 109.82, p value = 0.0000; F value = 108.52, p value = 0.0000), and advertisement content recall (F value = 34.429, p value = 0.0000; F value = 15.085, p value = 0.0000). This indicated subjects attribute the advertisement duration to the advertisement itself during the video.

3.6.2 Hypotheses Testing

Following Brambor, Clark, and Golder (2006), because we use one item to measure perceived duration of advertisement videos, four items to evaluate negative attitudes toward the advertisement, and two items to assess the content recall from the ad, we, therefore, adopt analysis of variance (ANOVA) for testing perceived duration hypotheses testing and multivariate analysis of variances (MANOVAs) for testing negative attitudes and content recall of advertisements.

In Table 3.4, we present the mean and standard deviation of perceived advertisement duration and negative attitudes toward the ad. The results are consistent with our hypothesis 1 and hypothesis 4. The average of perceived advertisement duration of the longer advertisement video is generally longer than for the shorter video. The average negative attitudes toward the shorter ad video is less than toward the longer video. In Table 3.5, based on ANOVA and MANOVAs testing results, a longer advertisement length significantly leads to more negative attitude from users and a longer perceived ad duration.

According to the values of mean and standard deviation from Table 3.6, during the longer advertisement video, subjects perceive the duration of the advertisement video to be longer if the video is designated with temporal reminder designs such as counting down clock and progressing bar indicator rather than the video with no temporal reminder design, which is consistent with our hypothesis 2b. Subjects perceive the duration of the advertisement video to be shorter if the video is embedded with temporal reminder designs such as the remaining number of advertisements rather than the video with no temporal reminder design, which is consistent with our hypothesis 2c. The duration is perceived longer for temporal reminder designs counting down clock and progressing bar indicator than the video with the remaining number of advertisements, which is in line with hypothesis 3. In addition, we assume that the different temporal reminder designs can serve as distractions of the attention when users are watching the video and therefore reduce the number of video contents remembered. Based on the results, the average of recalls with temporal reminder designs such as counting down clock and progressing bar indicator are fewer than without any temporal reminder. However, if the video is inserted with the number of remaining advertisements design, subjects can recall more contents than without any temporal reminder. Furthermore, the results displayed in Table 3.6 are opposite from hypothesis 5b. We assume time reminder designs can reduce the negative attitudes toward the advertisement video because they can relieve people's uncertainty of the time passage. The results indicate that the different temporal reminder designs can generate more negative attitudes than without any designs.

The ANOVA and MANOVAs testing results from Table 3.7 support hypothesis H2b, indicating the perception duration differences between the counting down clock and no temporal design (F =8.5162, p-value= 0.0037), and progressing bar indicator and no temporal design (F = 4.8743, p-value = 0.028) are significant. Even though subjects generate a shorter perception of duration with the remaining number of advertisements reminder, in comparison to no design, the difference is not significant. The testing results support H3, demonstrating the perception duration differences between the counting down clock and remaining number of advertisements reminder (F =10.787, p-value= 0.0011), and progressing bar indicator and remaining number of advertisements reminder (F = 6.7976, p-value = 0.0096), are significant. The hypotheses regarding the negative attitude aroused by different temporal reminder designs counting down clock (F = 0.8466, p-value = 0.4966), progressing bar indicator (F = 1.017, p-value =0.3987), and remaining number of advertisements (F =1.2204, p-value = 0.3021) are not significantly different from no temporal design and are not supported by the results either. In terms of content recall from advertisement videos, subjects recall significantly fewer brands and items (F =2.357, p-value = 0.0964) with a counting down clock in the video. The differences of content recall between designs such as progressing bar (F = 0.423, p-value = 0.6555) and no design, number of remaining advertisements (F = 0.5732, p-value =0.5643) and no design are not significant. Therefore, the results do not support hypothesis H6.

Similarly, as indicated in Table 3.8, user's perceived video duration can increase when they watch the desired video with the three different temporal reminder designs, in comparison to no reminder, for the shorter advertisement video as well. The differences are statistically significant based on the ANOVA results (F =7.631, p-value = 0.0061; F= 5.4676, p-value =0.02; F= 4.7772, p-value = 0.03), which support the hypothesis H2a. Even though the negative attitudes generated after watching the advertisement video with counting down clock, progressing bar indicator, and number of remaining advertisement are greater than with no design, the MANOVA results demonstrate the differences are not statistically significant, which does not support the hypothesis H5a. Unlike the longer advertisement video, subjects recall more items and brands when the video is designed with progressing bar indicator, which is not in agreement with our hypothesis. The differences are not significant either (F = 0.423, p-value = 0.6555). The summary of hypotheses supported and rejected by the study results is presented in Table 3.10.

3.6.3 Study 1 Discussion

The results of study 1 confirm the unfavorable impact of the longer advertisement video. Importantly, the study 1 findings (H2a, H2b, and H3) provide support for the perception theories memory-based model (Ornstein, 1969), attention-based model (McClain, 1983; Zakay, 1993), and paradigm shift during the longer waits (Hong et al., 2012). Providing additional temporal content (e.g., the counting down clock) in the shorter advertisement makes subjects feel the duration is longer, which is consistent with the memory-based models (Ornstein, 1969). The additional temporal content can serve as the memory cue that increases the perception of the duration. The results of H2a support the temporal designs including counting down clock, progressing bar, and remaining number of advertisements as additional memories can increase the duration perception. On the other hand, providing the counting down clock and progressing bar indicator designs in the longer ad video can lead user's attention to time passage itself, which can make users feel the video is longer. Giving the remaining number of advertisements design in the longer advertisement video can lead more user's attention to the nontemporal information (i.e., content of advertisements) than the purely temporal designs (i.e., counting down clock and progressing bar indicator), which can make users feel the video is shorter than the other two temporal designs. This result is consistent with the attention-based models (McClain, 1983; Zakay, 1993). The domination of the attention-based model during the longer advertisement video and memory-based model during the shorter ad video, respectively, also verifies the paradigm shift from the retrospective paradigm to the prospective paradigm (Block & Zakay, 1997) during the longer video duration evaluation (Hong et al., 2013).

The expected beneficial effects of temporal reminder design reducing negative affect in longer advertisement videos, and the expected detrimental effects of temporal reminder design increasing negative influence in shorter ad videos, are not confirmed. It is most likely because time reminders can lead the attention direction toward time, making the duration even longer, but also assures users with the remaining time. During the longer waits, when the attention-based model (McClain, 1983; Zakay, 1993) dominates, temporal reminder designs can lead users' attention to time, which increases the duration perception. Similarly, during the shorter advertisement video and when the memory-based model (Ornstein, 1969) dominates the duration perception, additional memory cues such as temporal reminders can increase the duration perception. On the other hand, when users are watching the advertisement video, as time goes by, they can be anxious, wondering about the remaining time of the video. Temporal reminders provide the specific or estimation of time information, which can relieve users' uncertainty and mitigate the negative attitudes toward the video. The two effects interact with each other and one cannot overweigh the other. The interaction of temporal designs that increase and reduce negative effects during both longer and shorter advertisement videos leads to the insignificant results of our hypotheses.

A surprising finding is that there are no significant differences in advertisement content recall between different temporal reminder designs and no temporal reminder. Based on the selective attention theory (Broadben, 1958), people have limited process capacity and they filter the information they can pay attention to when a large amount of information is being processed. Corresponding to the memory selection models (Broadben, 1958), the temporal reminders are not the relevant information that people pay attention to and they do not significantly distract people's attentions from the advertisement contents.

3.7 Conclusion and Discussion

This study examines various forms of temporal reminder designs that are unique to the online video advertisements. Since online video advertising is still a new advertisement type compared to the traditional paper advertisement, TV commercials, and online pop-in and banners, only a few studies have focused on the content level of advertisement videos and analyzed particular features such as skippable advertisements and ad placement. In our study, we employ time perception theories, reactance theories, and selective attention theories to examine how online users' perception of advertisement duration, negative attitudes, and contents recall from shorter and longer online video advertisements are impacted by providing different time reminder designs including counting down clock, progressing bar, and remaining number of ads. One empirical study was conducted on Amazon Turk for evaluating different impacts from various temporal reminder designs.

Our study effectively applies time perception theories to the online video advertisements' context and extends the analysis to temporal reminder designs during advertisement videos that have not been studied before. In general, our results show support for RAM and the distinctive time perception models during the retrospective paradigm and the prospective paradigm. Our study also verifies the paradigm shift during the longer online advertisement video. During the shorter advertisement video, in the retrospective paradigm, the memory-based time perception model dominates the duration perception, and our temporal designs are considered as additional memory cues that increase the video duration estimation. During the longer advertisement video, responding to the paradigm shift, the retrospective paradigm can change to the prospective paradigm. In the prospective paradigm, the attention-based time perception model directs that the duration estimation and temporal designs such as counting down clock and progressing bar indicator lead attention toward time passage and increase users' perception of video duration. The remaining number of ads reminder distract users' attention away from time and to the advertisement video during the longer video, and therefore results in shorter video duration perceptions than the other two temporal reminder designs. On the basis of reactance theory, we assume temporal reminder designs can increase the negative attitudes during the shorter advertisement since they are the factors that enlarge the perceived duration and reduce the negative attitudes during the longer advertisement video due to time uncertainty assurance. The empirical study does not support our hypotheses, possibly due to the two unmeasured conflicting attitude effects of the temporal designs. In line with selective attention theories, we presume the temporal reminder designs to be users' attention distractions from advertisement videos and therefore reduce the advertisement content recalls. However, the experimental results do not support our hypothesis. This is probably because the temporal reminders are not the important information that people filtered to pay attention to and they do not significantly distract people's attentions from the advertisement contents.

There are a few limitations of this study. First, we use a simulated desired video and a simulated advertisement insertion scenario. Users may not fully enjoy and get involved in the NBA video as their desired video. Therefore, the negative attitudes toward the impedimental advertisement video are not as negative as in their real life. Second, during this experimental study, we require participants to watch the video carefully without skipping and forwarding. They understand this is an experiment and there are related questions afterward. Therefore, they may pay much more attention to watching the advertisements, which leads to indistinguishable good ads recall results under different circumstances.

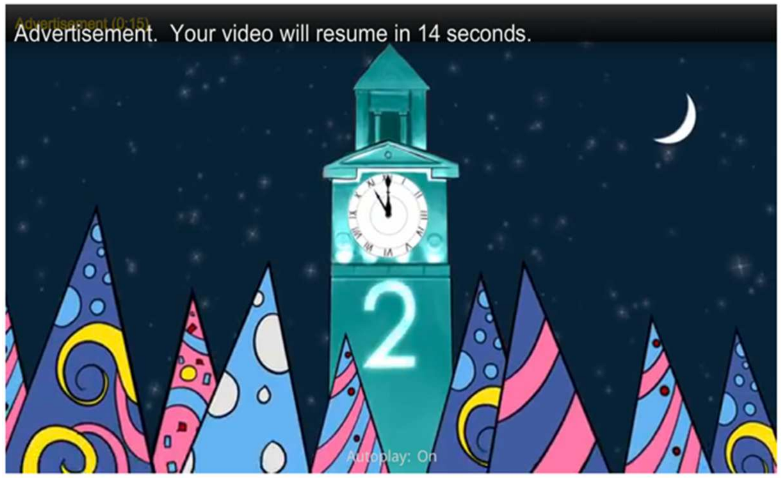

Figure 3.1: Counting Down Clock During the Advertisement Video

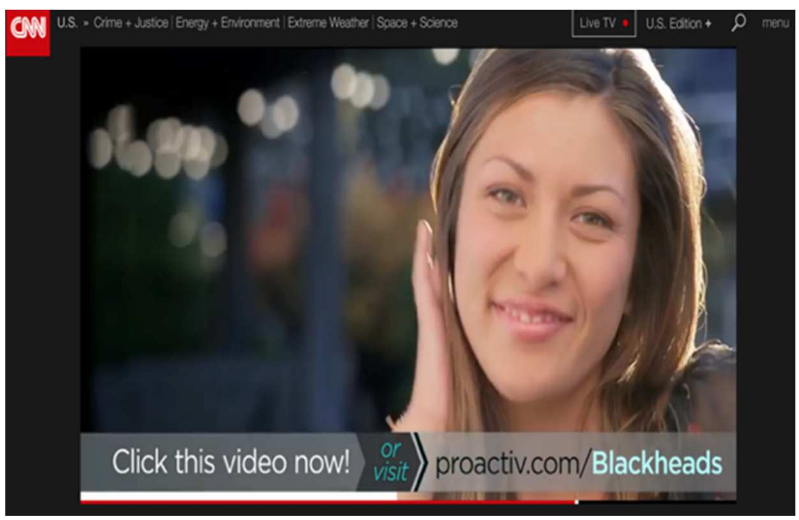

Figure 3.2: Progress Bar Indicator During the Advertisement Video

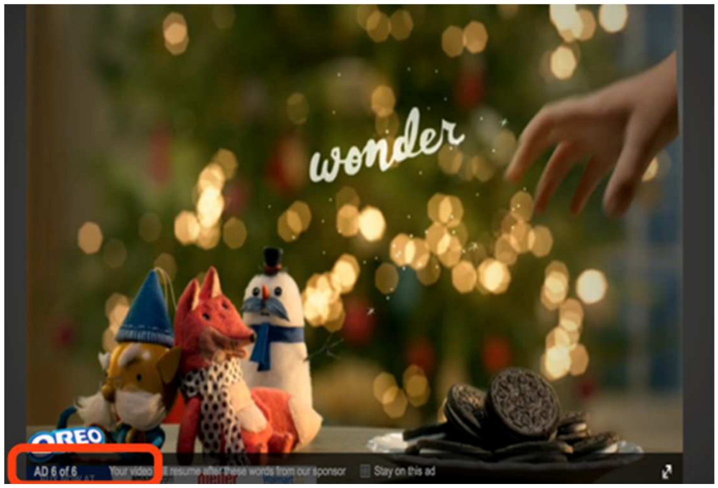

Figure 3.3: Remaining Number of Ads During the Advertisement Video

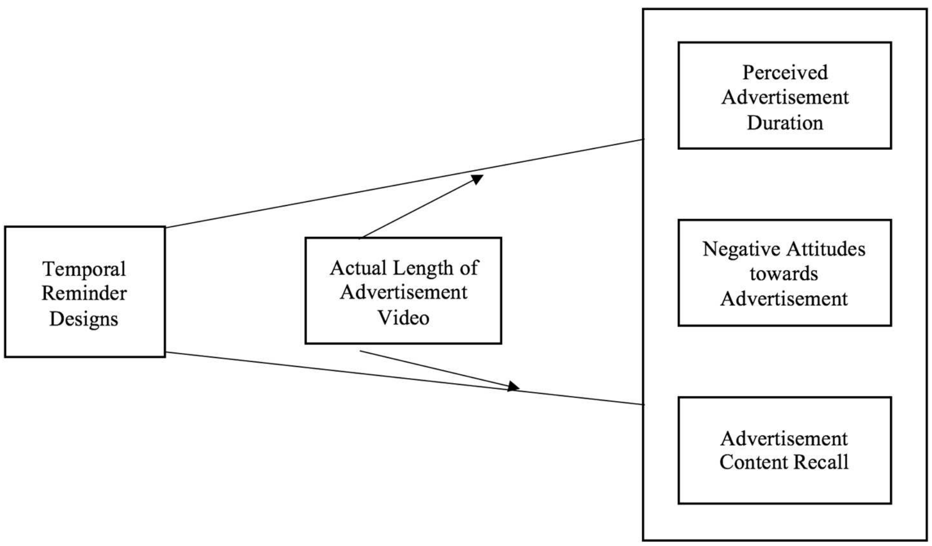

Figure 3.4: Research Model Design

Figure 3.5: Short and Long Online Video Ads With No Temporal Designs

Figure 3.6: Short and Long Online Video Ads With Counting Down Clock

Figure 3.7: Short and Long Online Video Ads With Progressing Bar Indicator

Figure 3.8: Short and Long Online Video Ads With Remaining Number of Ads

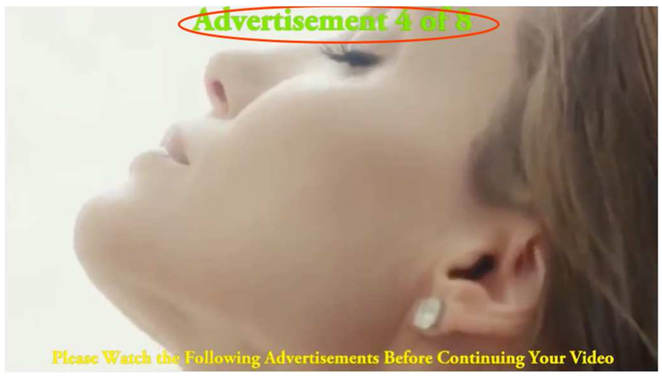

Figure 3.9: Change of Remaining Number of Advertisements

Figure 3.10: Manipulation Check

Table 3.1: Summary of Hypotheses

Hypothesis	Independent Variables	Dependent Variables	Hypotheses
H1	Length of the advertisement video (short vs. long)	Perceived duration of the advertisement video	H1: People perceive longer video duration during the longer advertisement video than during the shorter advertisement video.
H2a	Temporal Reminder Designs; Without any reminder; (short ad duration)	Perceived duration of the advertisement video	H2a: For a shorter video advertisement, temporal reminder information such as counting down clock, progressing bar indicator and remaining number of advertisements reminder can result in a longer perceived advertisement duration than without any information.
H2b	Temporal Reminder Designs: Counting down clock and progressing bar; Without any reminder; (long ad duration)	Perceived duration of the advertisement video	H2b: For a longer video advertisement, temporal reminder information such as counting down clock and progressing bar indicator can result in a longer perceived advertisement duration than without any information.
H2c	Temporal Reminder Designs: Remaining number of ads; Without any reminder; (long ad duration)	Perceived duration of the advertisement video	H2c: For a longer video advertisement, temporal reminder information such as the remaining number of advertisements reminder can result in a shorter perceived advertisement duration than without any information.

Table 3.1: Continued

Hypothesis	Independent Variables	Dependent Variables	Hypotheses
H3	Temporal Reminder Designs: Remaining number of advertisements reminder; Temporal Reminder Designs: Counting down clock, progressing bar; (long ad duration)	Perceived duration of the advertisement video	H3: For longer video advertisements, the remaining number of advertisements reminder can result in a shorter perceived advertisement duration than temporal information such as counting down clock and progressing bar indicator.
H4	Length of the advertisement video (short vs. long)	Negative attitude toward the advertisement video	H4: Longer online video advertisements can result in more negative attitude toward the online video advertisements than shorter online video advertisements.
H5a	Temporal Reminder Designs; Without any reminder; (short ad duration)	Negative attitudes toward the advertisement video	H5a: For a shorter advertisement, the temporal reminder designs such as counting down clock, progressing bar indicator, and remaining number of advertisement reminder can result in more negative attitudes toward the advertisement than an advertisement without any reminder.
H5b	Temporal Reminder Designs; Without any reminder; (long ad duration)	Negative attitudes toward the advertisement video	H5b: For a longer advertisement, temporal reminder designs such as counting down clock, progressing bar indicator, and remaining number of advertisements reminder can result in less negative attitudes toward the advertisement than an advertisement without any reminder.

Table 3.1: Continued

Hypothesis	Independent Variables	Dependent Variables	Hypotheses
H6	Counting down clock; Progress bar indicator; Number of remaining ad; Without any reminder; (long and short ad)	Advertisement Contents Recall	H6: For both longer and shorter advertisements, customers perceive and recall fewer advertisement contents with temporal information designs including counting down clock, progressing bar, and remaining advertisement reminder than without any reminder.

Table 3.2: The Summary of Study 1

	Study 1
Experiment Design	2 × 4 between subject
Focus	Examine the impacts of temporal reminder designs on the advertisement duration perception, negative attitudes toward advertisements and advertisement contents recall
Independent Variables	Actual Waiting Length (Shorter/Longer) Designs (Counting Down Clock, Progress Bar Indicator, Number of remaining advertisements)
Dependent Variables	The Perceived Duration of the Advertisement Video; The Negative Attitude Toward the Advertisement; The Content Recall of the Advertisement Video
Control Variables	Impatience Attribution
Analysis Method	ANOVA, MANOVA
Number of Subjects	1,266

Table 3.3: Subjects Belong to Each Design Group

Length/Temporal Design	No Design	Counting Down Clock	Progressing Bar Indicator	Number of Remaining Ads
Long	158	159	163	159
Short	157	157	157	156

Table 3.4: Descriptive Statistics of Perceived Duration and Negative Attitudes

Factors (Actual Advertisement Duration)	Perceived Duration	Negative Attitudes
	Mean (Standard deviation)	Mean (Standard deviation)
Longer Advertisement Video	4.8277 (1.4846)	3.8068 (1.8329)
Shorter Advertisement Video	3.1303 (1.4429)	2.785 (1.6529)

Table 3.5. MANOVA and ANOVA Results of Actual Advertisement Duration

Factors	ANOVA		MANOVA	
	Perceived Duration		Negative Attitudes	
	F	Sig.	*F*	Sig.
Actual Advertisement Duration	422.9	< 2.2e-16 ***	30.973	< 2.2e-16 ***

*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: Descriptive Statistics for Longer Ad Video

Factors (Longer Ad Duration)	Perceived Duration Mean (Standard Deviation)	Negative Attitudes Mean (Standard Deviation)	Content Recall Mean (Standard Deviation)
No Temporal Design	4.6346 (1.4372)	3.6687 (1.7845)	5.1329 (2.4078)
Counting Down Clock	5.1075 (1.4347)	3.9309 (1.9097)	4.9623 (2.4074)
Progressing Bar Indicator	4.9876 (1.4142)	3.8758 (1.8226)	5.0307 (2.4252)
Number of Remaining Ads	4.5541 (1.5541)	3.7885 (1.7671)	5.3899 (2.3731)

Table 3.7: MANOVA and ANOVA Results for Long Ad Video

Factors (Longer Advertisement Duration)	ANOVA		MANOVA		MANOVA	
	Perceived Duration		Negative Attitudes		Content Recall	
	F	Sig.	*F*	Sig.	*F*	Sig.
Counting Down Clock and No Temporal Design	8.5162	0.0037**	0.8466	0.4966	2.357	0.0964*
Progressing Bar Indicator and No Temporal Design	4.8743	0.028*	1.017	0.3987	0.423	0.6555
Remaining Number of Ads and No Temporal Design	0.2261	0.6347	1.2204	0.3021	0.5732	0.5643
Counting Down Clock and Remaining Number of Ads	10.787	0.0011**				
Progressing Bar Indicator and Remaining Number of Ads	6.7976	0.0096**				

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.8: Descriptive Statistics for Shorter Ad Video

Factors (Shorter Ad Duration)	Perceived Duration Mean (Standard Deviation)	Negative Attitudes Mean (Standard Deviation)	Content Recall Mean (Standard Deviation)
No Temporal Design	3.1178 (1.095)	2.6026 (1.5562)	0.7643 (0.4258)
Counting Down Clock	3.4679 (1.1468)	2.7516 (1.5049)	0.7325 (0.444)
Progressing Bar Indicator	3.4268 (1.2414)	2.7325 (1.7408)	0.7898 (0.4087)
Number of Remaining Ads	3.4123 (1.2759)	2.9355 (1.7826)	0.7051 (0.4574)

Table 3.9: MANOVA and ANOVA Results for Shorter Ad Video

Factors (Shorter Advertisement Duration)	ANOVA		MANOVA		MANOVA	
	Perceived Duration		Negative Attitudes		Content Recall	
	F	Sig.	*F*	Sig.	*F*	Sig.
Counting Down Clock and No Temporal Design	7.631	0.0061**	0.821	0.5125	0.7182	0.4884
Progressing Bar Indicator and No Temporal Design	5.4676	0.02*	0.5512	0.6983	0.423	0.6555
Remaining Number of Ads and No Temporal Design	4.7772	0.03*	1.132	0.3414	1.2965	0.275

Table 3.10: Summary of Hypotheses Results

Hypothesis	Hypotheses	Results
H1	H1: People perceive longer video duration during the longer advertisement video than during the shorter advertisement video.	Supported
H2a	H2a: For a shorter video advertisement, temporal reminder information such as counting down clock, progressing bar indicator and remaining number of advertisements reminder can result in a longer perceived advertisement duration than without any information.	Supported
H2b	H2b: For a longer video advertisement, temporal reminder information such as counting down clock and progressing bar indicator can result in a longer perceived advertisement duration than without any information.	Supported
H2c	H2c: For a longer video advertisement, temporal reminder information such as the remaining number of advertisements reminder can result in a shorter perceived advertisement duration than without any information.	Rejected
H3	H3: For a longer video advertisements, the remaining number of advertisements reminder can result in a shorter perceived advertisement duration than temporal information such as counting down clock and progressing bar indicator.	Supported
H4	H4: Longer online video advertisements can result in more negative attitude toward the online video advertisements than shorter online video advertisements.	Supported
H5a	H5a: For a shorter advertisement, the temporal reminder designs such as counting down clock, progressing bar indicator and remaining number of advertisement reminder can result in more negative attitudes toward the advertisement than an advertisement without any reminder.	Rejected
H5b	H5b: For a longer advertisement, temporal reminder designs such as counting down clock, progressing bar indicator and remaining number of advertisements reminder can result in less negative attitudes toward the advertisement than an advertisement without any reminder.	Rejected
H6	H6: For both longer and shorter advertisements, customers perceive and recall fewer advertisement contents with temporal information designs including counting down clock, progressing bar and remaining advertisement reminder than without any reminder.	Rejected

CHAPTER 4

ONLINE VIDEO ADVERTISEMENTS: HOW DOES CONTENT RELEVANCE AFFECT USER'S OPINIONS?

4.1 Introduction

Online video advertising is the fastest growing advertising format (eMarketer, 2012). It is estimated that 74% of all Internet traffic is video-based by 2017 and video ads can account for 35.6 % of total videos viewed. According to the latest infographic released by invespcro.com, estimated digital video advertisement spending in the United States can reach $9.59 billion dollars in 2016 from $5.96 billion dollars in 2014, while the worldwide revenue from online video advertising is estimated to achieve $11.4 billion by 2016 from $8.3 billion in 2014.

Online video advertisements look very similar to traditional TV commercials but are also different in many ways. Firstly, online users tend to be more goal-oriented, when TV as a passive media pushes the advertisement to the mass audience. Secondly, as TV viewers are used to the TV commercial breaks as a standardized and inherent process of TV watching, online users are not aware of when and where an online video advertisement can interrupt the video that they are currently watching. As a result, online video advertisements receive more negative evaluations than TV commercials and other types of traditional ads (Logan, 2013). In general, since viewers are goal- and search-oriented, and advertisements can interrupt their goal, the attitudes toward the online advertisements are more negative. Social effects of advertising are generally negative (Krugman, 1983; Previte & Forrest, 1998).

The online video advertisement is still a relatively new form of advertising, so limited studies have focused on customers' responses and how to reduce the negative attitudes toward the video. The majority of existing studies emphasize the impact of several unique features of online video advertising such as the skippable control option (Hegner et al., 2016; Kusse, 2013) and the advertisement video placement (Krishnan & Sitaraman, 2013). Fewer studies incorporate other factors of online video ads, such as length (Kusse, 2013), emotional appeal (Hegner et al., 2016; Kim, 2015), content relevance (Hegner et al., 2016; Kim, 2015; Kusse, 2013), and their effects on customers' attitudes and perceived intrusiveness.

Online video advertisement content relevance refers to the extent to which an online video ad is similar to the video content that users are presently watching, in terms of contexts, topics, or execution styles (Li & Lo, 2014; Mei, Hua, & Li, 2009; Moorman et al., 2002; Tutaj & van Reijmersdal, 2012). The online video advertisement is normally embedded or inserted in the video that users are currently watching. Since users select to watch the video based on their preferences, we refer to the video that they choose as the desired video. There are two different types of content relevance. One is referred to as thematically or contextually similar (Li & Lo, 2014; Mei et al., 2007; Moorman et al., 2002). The other is considered as execution similar based on the aspect of native advertising (Van Reijmersdal et al., 2005). Similarity in terms of context and topic can be defined as the extent to which an advertisement and its surrounding media content are similar regarding their topic, content, context, or theme (Jeong & King, 2010; Kim, 2005; Kim & Sundar, 2012; Li & Lo, 2014; Moore, Stammerjohan, & Coulter, 2005). For example, there can be a similarity level (high or low) between the contextual content of a banner ad and its surrounding media. A website (Jeong & King, 2010) or a print ad match the topic with magazines, where the ads are placed in based on the themes and contents of the magazines (Moorman et al., 2002; Simola et al., 2013; Van Reijmersdal et al., 2005). On the other hand, execution similar refers to native advertising. Based on IAB (2013), native advertising is "paid ads that are so cohesive with the page content, assimilated into the design, and consistent with the platform behavior that the viewer simply feels that they belong" (p. 3). For example, a suggested post or people on Facebook and promoted tweet on Twitter are considered as native advertising, which highlights the key features of native advertising: sponsored and similar to the page content.

Apart from the online video advertising format, for traditional advertisements, such as banner advertisements and newspaper ads, prior studies have evaluated various impacts of the content relevance on purchasing intentions (Goldfarb & Tucker, 2011), return intentions (McCoy et al., 2004), customer attitudes (Jeong & King, 2010; Kim & Sundar, 2012) and effectiveness (Jeong & King, 2010; Moorman et al., 2002; Porta et al., 2013; Rodgers, 2003; Simola et al., 2013; Yaveroglu & Donthu, 2008; Zanjani et al., 2011). As for the online video advertising type, existing studies limit primary focuses of video content relevance to evaluation of customer attitudes and intrusiveness of the video (Hegner et al., 2016; Kim, 2015; Kusse, 2013).

In this study, we apply time perception theories, contextual priming effect, and schema incongruent effect to evaluate the influence of different content relevance levels of online video advertisements with the desired video on ad duration perception, perceived negative users' attitudes, and recall of advertisement contents. Based on the theories, we address the following research questions in this study: How does (1) content relevance affect online users' perception of advertisement duration, the attitude, and message recall? We further analyze: (2) how does the length of the online video advertisement change the effect that content relevance has on the perceptions of ad duration time, the attitudes, and content recall? One experimental study is conducted to answer the above questions.

The remainder of the study is organized as follows. We first review related studies and highlight key differences between our study and previous representative research. We then formally propose the theoretical foundations including time perception theories, contextual priming effect, and schema incongruent effect of this study. Furthermore, we hypothesize a few hypotheses based on the existing literature and theories. We describe one experimental study with research design, data collection and analysis, and a discussion of the results. Finally, we conclude the chapter with contributions, implications, and suggestions for future research.

4.2 Literature Review

Two streams of research are relevant to our study, including the effect of advertisement content congruent on perceived customer attitudes and advertisement intrusiveness, and the impact of advertisement content congruent on advertisement effectiveness. In this section, we review representative studies in each stream and highlight the gaps that motivate our research.

4.2.1 Perceived Attitudes and Advertisement Intrusiveness

Prior studies suggest that matching the content of advertisements to the media context tend to generate more positive attitudinal feedback (Braun-LaTour, Puccinelli, & Mast, 2007; Coulter, 1998; Edwards, Li, & Lee 2002; Hegner et al., 2016; Kamins, Marks, & Skinner, 1991; Kim, 2015; Kim & Sundar, 2012; Kusse, 2013; Mandler, 1975; Moore et al., 2005; Tutaj & van Reijmersdal, 2012; Van Reijmersdal et al., 2005; Ying, Korneliussen, & Grønhaug, 2009) in terms of traditional mailing advertisements, TV commercials, and online advertisements such as banners and pop-ups, as well as online video ads. For example, Van Reijmersdal et al. (2005) examine the perceived informativeness and amusement from magazine subscribers if they received advertisements with contextual contents well blended with magazine content. As for TV commercials, studies discover if there is contextual overlap between TV commercials and TV programs, and if both commercials and programs share the same emotional level, viewers elicit more favorable attitudes (Coulter, 1998; Kamins et al., 1991). In the context of online advertising, website congruent ads are perceived as less intrusive, create less psychological reactance, and receive more positive attitudes than incongruent ads (Edwards et al., 2002; Moore et al., 2005). Specifically, by inserting a computer store banner ad (relevant ad) and a student loan banner ad (irrelevant ad) to computer websites, Jeong and King (2010) discover that the users evaluate the relevant ad more favorably than the irrelevant ad. When displaying the ads with the contextual relevance and irrelevance, Kim and Sundar (2012) observe that the relevant ads are evaluated more positively than irrelevant ads. Regarding online video advertisements, Kusse (2013) generate four video conditions, matching an Adidas advertisement video with a soccer game as the relevant online ad and matching an Adidas advertisement video with the African savanna as the irrelevant online ad. The results indicate that the relevant online ad led to a more positive attitude toward brands than the irrelevant video. Two other studies (Hegner et al., 2016; Kim, 2015) construct high and low content similarities between the advertisement video and the desired video environment and obtain a less negative feedback with the ad video with higher content similarity.

Some studies, on the other hand, receive opposite results (Campbell, 1995; Lunardo & Mbengue, 2013). Even though greater contextual similarity of advertisements attracts additional attention (Van Reijmersdal et al., 2005), some studies reveal that customers find these ads misleading and evaluate them more negatively than irrelevant advertisements (Campbell, 1995; Lunardo & Mbengue, 2013).

4.2.2 Advertisement Effectiveness

A large number of prior studies show support for better memory outcomes of customers when the content of advertisements is similar to the media context thematically, contextually, and executionally (Moorman et al., 2002; Porta et al., 2013; Rodgers, 2003; Simola et al., 2013; Yaveroglu & Donthu, 2008; Zanjani et al., 2011). For example, concerning traditional mailing advertisement, Simola et al. (2013) find better advertisement recognition with a high contextually similarity between advertisement and article when they place a beer ad in an article about beer in a magazine than the irrelevant ad. Similarly, placing an apparel advertisement in a lifestyle magazine can generate a higher recognition score than inserting it in a healthcare magazine (Moorman et al., 2002).

In terms of online advertising environment, Yaveroglu and Donthu (2008) place ads promoting technical products in a Personal Computer World website as a thematically relevant ad-media relationship, in comparison with placing the same ad in CNN website as the irrelevant relationship. The results indicate a better brand name recall with the relevant relationship than the irrelevant relationship. When comparing promoting technical products in an e-magazine regarding technologies (thematically congruent ads) and promoting the same set of products in an e-magazine about travel (thematically incongruent ads), Zanjani et al. (2011) also demonstrate that the information seekers generate better advertisement recall and recognition with thematically congruent ads.

Besides item and brand recall and recognition, in this literature review, we also consider purchase intentions and return plans as advertisement effectiveness. By matching the advertisement to website content, Goldfarb and Tucker (2011) discover greater purchase intentions among users. However, when combining the contextual relevance with advertisement obtrusiveness, the intention was not significantly higher than without the relevance. Unlike the prior mentioned studies, when comparing the return intention of users after they watch the online advertisements with inline versus pop-up banner design and congruent content versus incongruent content, McCoy et al. (2004) discover a higher return intention level of users after they watch advertisements with incongruent content. Also, opposite from the studies previously discussed, Jeong and King (2010) assume irrelevant online banners can motivate better users' brand recall and recognition since they require a longer time to process in users' memory. The results, however, indicate the differences are not statistically significant.

As for online video advertising format, only Kim (2015) tried to evaluate the impact of content relevance between the advertisement video and the desired video on the ad outcome (brand recall) and hypothesize relevant ad videos can lead to higher brand memory. However, the results demonstrate no significant differences between brand recall with relevant and irrelevant advertisement videos.

4.2.3 Research Gap Summary

A large number of studies emphasize evaluating the impact of content relevance of advertisements on negative customer attitudes and perceived ad intrusiveness (Braun-Latour et al., 2007; Coulter, 1998; Edwards et al., 2002; Hegner et al., 2016; Kamins et al., 1991; Kim & Sundar, 2012; Kusse, 2013; Mandler, 1982; Moore et al., 2005; Tutaj & van Reijmersdal, 2012; Van Reijmersdal et al., 2005; Ying et al., 2009), and advertisement effectiveness (Moorman et al., 2002; Porta et al., 2013; Rodgers, 2003; Simola et al., 2013; Yaveroglu & Donthu, 2008; Zanjani et al., 2011) in traditional formats such as magazine, TV commercials, and online advertisements such as pop-in and banners.

The studies related to online video advertising, however, are sparse. The studies mainly focus on the negative customer attitude effects and perceived advertisement intrusiveness (Hegner et al., 2016; Kim, 2015; Kusse, 2013). Only a few studies have tried evaluating the effectiveness of ads with different levels of content relevance (Kim, 2015), but no significant results have been discovered. Furthermore, different from other advertising formats, the duration of the advertisement video can be considered as a waiting process. Our study, therefore, aims at investigating the influence of ad content relevance on users' perception of advertisement video duration, negative attitudes toward both the ads and the brands, and the effectiveness of the advertisements (brand and product recall), that complements the existing literature. Additionally, because the length of the advertisement video is a significant factor impacting all three dependent variables, our study introduces the actual length of the ad video as the mediated factor for the measurement.

4.3 Theory Foundation

4.3.1 Time Perception Theories

The two conflict models regarding time perception are memory-based models and attention-based models. Memory-based models propose that during a time interval, the more information cues that users remember, the longer they estimate time duration to be (Ornstein, 1969). On the opposite side, attention-based models suggest there is a cognitive timer in mind that people use to track the passage of time. When additional information that is not related to a temporal reminder requires people's attention, people spend less attention on the cognitive timer and passage of time. Therefore, they estimate shorter time duration (McClain, 1983; Zakay, 1993). Time is linearly but negatively associated with the experimental complexity with which given intervals are filled (Priestly, 1968).

One reconciliation between the two models is the resource-allocation model

(RAM) (Kahneman, 1973; Zakay, 1989). The RAM assumes people's central attention has limited capacity and the attention allocates to different tasks all the time (Kahneman, 1973). Two paradigms are proposed: prospective paradigm and retrospective paradigm (Block & Zakay, 1997). Users being notified to estimate time passage before experiencing the wait is known as the prospective paradigm. In contrast, users being requested to approximate the time duration after the wait is considered as the retrospective paradigm (Block & Zakay, 1997).

During the prospective paradigm, when the majority of user's attention capacity is used for estimating the time, nontemporal information such as music is able to distract users' attention from the cognitive timer, therefore reducing the estimation of time passage. During the prospective paradigm, attention-based models dominate the waiting process. In contrast, when a user is asked to evaluate the duration after the occurrence of the wait, any information or memory cues can lead to the increasing of waiting duration estimation. Memory-based models dominate the waiting procedure during the retrospective paradigm. More time reminders during the wait increase the memory cues. As a result, longer perceived waiting time is assessed.

Because users are not requested to evaluate the wait before the actual waiting in most of the cases, the retrospective paradigm takes place in most of the waiting environments. Hong et al. (2013) propose a paradigm shift within the longer wait. During the shorter wait, as mentioned in RAM (Block & Zakay, 1997), the retrospective paradigm and memory-based model dominate the waiting evaluation. Throughout the longer wait, people start to wonder about the amount of time they already waited, question the additional time they still need to wait, and determine if they should continue to wait. At this point, the retrospective paradigm shifts to the prospective paradigm with the attention-based model determining the waiting perception. Therefore, any nontemporal information that distracts people's attention away from the wait reduces the duration perception more than any temporal information that directs people's attention toward time (Hong et al. 2013).

4.3.2 Contextual Priming Effect

When contexts and the advertisement contents share a high contextual similarity, the contextual priming effect can appear. According to Yi (1990a, 1990b), when the website context provides information of a particular feature, while an advertisement emphasizes the similar feature at the same time, the feature can be enhanced, and this feature can be used to interpret advertisement message and evaluate the ad product.

Furthermore, since people understand information mostly depending on current stimuli, the immediately available contexts are critical when they interpret new information such as the advertisement video (Erdley & D'Agostino, 1988; Higgins, Bargh, & Lombardi, 1985; Wyer & Srull, 1981). Because of the high contextual similarity, the website context can increase the information accessibility to people when they look at the advertisement, which can increase people's excitation level (Higgins et al., 1985; Wyer & Srull, 1981).

4.3.3 Schema Incongruent Effect

The schema incongruent effect occurs when a new stimulus appears, and it can be used to understand the contextual effect when the content of an advertisement is unexpected and incongruent to the website context (Berlyne, 1960; Houston, Childers, & Heckler, 1987; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989)

This effect can be explained in two different ways. Firstly, there is a novelty effect of an unexpected stimulation (Berlyne, 1960). According to Berlyne (1960), novel information can be processed more elaborately. Secondly, incongruent information is harder to comprehend and is held in working memory longer than congruent stimulations (Houston et al., 1987; Pezdek et al., 1989).

4.4 Research Model

4.4.1 Advertisement Duration Perception

Under ordinary online circumstances, users' estimation of duration is in the retrospective paradigm, in which case, they are not forced to estimate the length before watching the video. Based on the resource allocation model (Block & Zakay, 1997), for the retrospective paradigm, the memory-based time perception model (Ornstein, 1969) is dominating the estimation process. Additional nontemporal information can increase the estimate of time. The nontemporal information can be considered as extra memory cues. When users are asked to evaluate the waiting length afterward, more memory cues remembered can lead to longer estimation of duration time. This paradigm is consistent with prior studies, in which nontemporal information such as music (Chebat et al., 2010; Hui et al., 1997) negatively affects waiting perception.

Corresponding to the attention-based model (McClain, 1983; Zakay, 1993), the resource allocation model (Block & Zakay, 1997), and the paradigm shift proposition (Hong et al., 2013), a paradigm shift occurs from the retrospective paradigm to the prospective paradigm during the longer waiting environment. Users may start to wonder about the passage of waiting time, if the waiting environment is functioning well, and whether they should continue to wait (Hong et al., 2013). For the prospective paradigm, users start to track time duration using their cognitive timer in mind. The attention-based model and temporal information are dominating the waiting evaluation so that nontemporal information can distract attention away from time passage estimation, which is in line with the past studies that nontemporal information, such as music (Cameron et al., 2003), TV (Pruyn & Smidts, 1998) and electronic news board (Katz et al., 1991), negatively affects waiting perceptions.

A shorter video advertisement can be considered as a waiting period with shorter duration. According to RAM (Block & Zakay, 1997), for normal online activities, users can focus on their tasks instead of estimating the duration of a video before watching it, which indicates the retrospective paradigm is the main paradigm for most online video watching cases. In the retrospective paradigm, the memory-based model (Ornstein, 1969) directs the waiting perception. Based on the memory-based model (Ornstein, 1969), according to the schema incongruent effect (Berlyne, 1960; Houston et al., 1987; Pezdek et al., 1989), compared to advertisements with relevant content, irrelevant information with a prior expectation is more complicated to process and interpret, and therefore held in memory longer than advertisements with relevant context. Longer memory process indicates more memory cues (Ornstein, 1969). Therefore, we hypothesize that the irrelevant advertisement video content can generate a longer duration perception of the ad video than the relevant advertisement video.

H1a: For shorter video advertisements, an advertisement video with irrelevant content can lead to a longer perception of advertisement duration than an advertisement video with relevant content.

During the longer video advertisement, there can be a paradigm shift in that the attention-based model can direct the duration perception (Hong et al., 2013). After paradigm shift, the retrospective paradigm can shift to the prospective paradigm. The evaluation of the time passage of the video becomes the primary task of users (Block & Zakay, 1997). In the prospective paradigm, the attention-based model (McClain, 1983; Zakay, 1993) and temporal information control the overall duration estimation. If nontemporal information such as music, pictures, and TV is given to users and distracts users' attention from the estimation of time, the perceived duration can be reduced (Cameron et al., 2003; Katz et al., 1991; Pruyn & Smidts, 1998). Because relevant advertisement content shares high contextual similarity with the desired video, according to contextual priming effect (Erdley & D'Agostino, 1988; Higgins et al., 1985; Wyer & Srull, 1981), the desired video context can increase the information accessibility to people when they look at the advertisement video, which can increase people's excitation level (Higgins et al., 1985; Wyer & Srull, 1981). The excitation level of the relevant advertisement content can direct more people's attention away from time and toward the ad video than irrelevant advertisement context. Therefore, we hypothesize that ad videos with irrelevant content during the longer ads duration can positively affect the perceived video duration.

H1b: For longer video advertisements, an advertisement video with relevant content can lead to a shorter perception of advertisement duration than an advertisement video with irrelevant content.

4.4.2 Negative Attitudes Toward Advertisements

When advertisement videos are embedded in a topically relevant desired video, the desired video and ads share a higher level of contextual similarity. Under this circumstance, the contextual priming effect is likely to occur. According to the contextual priming effect (Erdley & D'Agostino, 1988; Higgins et al., 1985; Wyer & Srull, 1981), the pre-existing information of the desired video can be the prior exposure about information shared with an advertisement video, which can motivate easier interpretation and better evaluation of the ad contents. Therefore, when the contextual priming effect occurs, the temporal availability of information from recent memory can increase people's excitation level and impact people's attitude toward the content of an advertisement video (Higgins et al., 1985; Wyer & Srull, 1981).

Furthermore, based on a study Wyer and Srull (1981) proposed, when embedding a banner advertisement on a website, the website contextual information can be saved in people's top layer storage bin and can be immediately retrieved for interpreting relevant banner information. In this case, the immediate information to the content of the advertisement video can be useful for evaluating the advertisement video. According to Sallam and Algammash (2016) and Gardner (1985), the customers' attitude toward ads has a significant and positive effect on their attitude toward the brand. The better attitude toward the advertisement video can result in a better attitude toward brands promoted. Therefore, we hypothesize that the advertisement video with relevant content can result in better evaluation of the ad than an advertisement video with irrelevant content. The advertisement video with relevant content can improve the assessment of the brand promoted more than the advertisement video with irrelevant content.

H2a: For the long advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the advertisement than the advertisement with irrelevant content.

H2b: For the short advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the advertisement than the advertisement with irrelevant content.

H3a: For the long advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the brand promoted than the advertisement with irrelevant content.

H3b: For the short advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the brand promoted than the advertisement with irrelevant content.

4.4.3 Recall of Online Video Advertisements

Based on the study proposed by Calvo, Castilo, and Estevez (1999), context-induced expectations are generated by websites. These expectations can assist people in comprehending new information. Thus, when people watch the desired video for a while, they may induce a schema and be able to process new information based on the video-related schema. Then if the following advertisement video is different from the desired video-caused schema, the schema incongruent effect appears.

In line with the schema incongruent effect (Berlyne, 1960; Houston et al., 1987; Pezdek et al., 1989), because the content of the advertisement video is incongruent with prior context-induced expectation, information from the ad video can be more complicated to comprehend and interpret. People require longer processing time and larger memory space (Houston et al., 1987; Pezdek et al., 1989). Since the information conveyed in the advertisement video can be considered as a new subject, people process it more elaborately to reduce emotional arousal (Jeong & King, 2010). Therefore, we hypothesize that people can recall more contents from advertisement videos with irrelevant content than ad videos with relevant content.

H4a: For the long advertisement video, customers perceive and recall fewer advertisement contents from the relevant advertisement video than the irrelevant advertisement video.

H4b: For the short advertisement video, customers perceive and recall fewer advertisement contents from the relevant advertisement video than the irrelevant advertisement video.

The research model is demonstrated in Figure 4.1. The hypotheses in the research model are concluded in Table 4.1.

4.5 Experimental Study

One experimental study is proposed to examine the hypotheses of the impact of the content relevance of the advertisements on the perception of advertisement duration, negative attitudes toward the advertisement(s)/brand(s), and advertisement content recall.

4.5.1 Desired Video, Advertisement Video, and Task

We conducted a controlled experiment with a 2 (actual length of ads: short/long) by 2 (advertisement video with relevant contextual content, advertisement video with irrelevant contextual content) full-factorial between-subject design. Subjects were Amazon Turk workers who have received at least 95% project approval rate and finished at least 50 projects before. We consider the use of Amazon Turk workers as subjects to be appropriate because these workers use the Internet for a long time each day. Because we are examining a usual psychological phenomenon (i.e., watching online video advertisements), people can give a similar reaction. Workers received $0.50 for participating.

We used the same NBA game collection video gathering best NBA shots as our desired video. The reason to choose the NBA game collection video as our designed video was that it was popular, had a significant amount of views, and online users could quickly get involved in the video. The total length of the NBA video was around 3 minutes and 30 seconds.

For the advertisement video, in order to evaluate the effects of the temporal reminder design with the remaining number of advertisements, we distinguished the short advertisement video and the long advertisement video by the total number of ads. In this case, the short ad video was composed of one advertisement video with a single brand and product, while the long advertisement video was a combination of several videos with different brands and products. To measure the effects of advertisements with relevant content and irrelevant content in this study, we selected eight advertisement videos with irrelevant content and another eight ad videos with relevant content (e.g., each of the first four videos was 15 seconds long and each of the other four videos was 30 seconds).

We provided workers with a cover story that they were watching an NBA game collection video. After watching the video without skipping and fast forwarding, several questions regarding the games would be asked. We also told them there was a manipulation check of whether they have finished the video. Submissions would be rejected if they failed the verification test.

4.5.2 Pilot Study

We conducted a pilot study with 72 workers to identify appropriate advertisement video duration embedded in the desired NBA collection video. We preselected four duration intervals from 15 seconds, 1 minute, 2 minutes, and 3 minutes based on the review of the literature. We used the similar cover story and asked the workers to watch the videos with different lengths of advertisement videos. After they finished watching the video, we asked them to evaluate the perception of the advertisement duration. They believed they had watched a short, moderate, or long advertisement video. Based on the responses, we selected two intervals in the experiment, which were 15 seconds for the shorter advertisement, and 3 minutes for the longer advertisement.

4.5.3 Variable Manipulation

Actual advertisement duration was manipulated as short and long (15 seconds and 3 minutes). The irrelevant short advertisement video was composed of a video promoting glass jars manufactured by the brand ball as shown in Figure 4.2, while the irrelevant long advertisement video was constituted of eight different advertisement videos with the first four of them lasting about 15 seconds each and the other four of them lasting about 30 seconds as displayed in Figure 4.3. The related short advertisement video was composed of a video promoting sneakers manufactured by Peak featuring NBA player Tony Parker as demonstrated in Figure 4.4. Similarly, eight additional relevant advertisement videos were used for constituting the long video. To ensure the contextual relevance, we determined to use advertisements that are all featuring NBA players and promoting products such as sneakers, sports drinks, automobiles, and Footlocker stores as presented in Figure 4.5. All desired videos and video advertisements were derived from the original commercials and tailored to fit the experiment by editing the length and look of the video. We embedded the advertisement video in the middle of the desired video to ensure a better advertisement completion rate. Since the total length of the desired video was around 3 ½ minutes, we inserted the advertisement video at the around the 2 minutes time point.

4.5.4 Variable Measured

All dependent variables were measured using existing scales from published studies. Specifically, the perceived ad duration was measured similar to Gorn et al. (2004) with a one-item semantic differential scale. Negative attitudes toward the advertisement and promoted brands were measured by a four-item, seven-point Likert-type scale based on the existing instruments respectively (Hong et al., 2013; Hui et al., 1997; Hui & Tse, 1996). Instead of using open-ended and closed- ended questions for querying user's advertisement product and brand recall (Jeong & King, 2010), we provided multiple choice questions asking workers to pick the correct watched goods and brands out of 15 different selections.

Additionally, two control variables, impatience and attribution, were also evaluated (Hong et al., 2013). Impatience was one of the most significant time-related individual difference factors that impacted the perception of time duration estimation (Francis-Smythe & Robertson, 1999). Attribution of the advertisements was studied affecting online users' evaluation toward the performance of websites (Houston et al.,1998; Rose et al., 2005). Impatience was measured with three items and attribution was measured with two items using existing instruments based on the literature (Houston et al., 1998; Rose et al., 2005; Spence et al., 1987).

Furthermore, to ensure workers finished the entire video without skipping, we inserted a screenshot with a number displaying for 5 seconds asking users to remember it as a manipulation check as shown in Figure 4.6. The duration of the desired video with the long advertisement video was around 5 ½ minutes, while the one with the short advertisement video was around 3 minutes. The time duration workers used for finishing the task was another manipulation check.

4.5.5 Subjects and Procedure

A total of 2,736 subjects completed this study, with 76.89% being male, 22.79% being female, and 0.32% choosing rather not to tell. Amongst all workers, 47.83%, 43.80%, and 8.29% were between age 18 to 29, age 30 to 45, and over 45, respectively.

Workers were shown an instruction that contained a cover story. They were then directed to the NBA game video. After they clicked the start button, the NBA game collection video played. At around the 2 minutes time point, the desired video stopped. The advertisement video started with relevant and irrelevant contents. After the advertisement video, the NBA game collection continued to play. Upon finishing up the desired video, the subjects were then redirected to an online survey measuring the perceived duration, negative attitudes toward advertisements, as well as the brand and product recall from the video. In addition, workers were asked to respond to additional questions regarding their general impatience, attribution, and demographic questions such as gender and age. The empirical study is summarized in Table 4.2.

4.6 Data Analysis Results

We tested the manipulation and control check and hypotheses testing. The data analysis results of the study are presented in the following sections.

4.6.1 Manipulation and Control Checks

We incorporated in total 1,368 Amazon Turk workers to watch the irrelevant advertisement video and another 1,268 workers to watch the relevant advertisement video. Based on our manipulation check, there are two criteria we used to discard the results by the subject. The first criterion is when workers failed to see the number embedded near the end of the advertisement video. The second criterion is when the total time workers spent finishing the task is less than the video duration. After rejecting the results from subjects who failed the manipulation check, 1,266 workers who watched the irrelevant advertisement video were kept in this study. Amongst the 1,266 subjects, 639 were assigned for finishing the video inserted with the irrelevant longer advertisement, while 627 completed the video with the irrelevant shorter advertisement. After rejecting the results from subjects who failed the manipulation check, 1,242 workers who watched the relevant advertisement video were kept in the study. Amongst 1,242 subjects, 611 were assigned for finishing the video inserted with the longer advertisement, while 631 completed the video with the shorter advertisement. The statistics of subjects are displayed in Table 4.3.

Impatience and Attribution significantly related to perceived advertisement duration (*F* value = 2.3328, *p*-value = 0.07249; *F* value = 62.338, *p*-value = 0.0000) and negative attitudes toward advertisement (*F* value = 84.645, *p* value = 0.0000; *F* value = 50.096, *p*-value = 0.0000), but were not significantly impacting advertisement content recall (*F* value = 1.1543, *p*-value = 0.3156; *F* value = 0.2666, *p*-value = 0.766).

4.6.2 Hypotheses Testing

Following Brambor et al. (2006), we use one item to measure the perceived duration of the advertisement video, four items to evaluate negative attitudes toward the advertisement, and two items to assess the content recall from the advertisement. We adopt ANOVA for testing perceived duration hypotheses and MANOVAs for testing negative attitude effects and content recall of advertisements.

In Table 4.4, we present the mean and standard deviation of perceived advertisement duration and negative attitudes toward the advertisement with both relevant and irrelevant video content. The average of perceived advertisement duration of the longer advertisement video is longer than the shorter video. The average negative attitudes toward the shorter advertisement video are smaller than toward the longer video.

Based on the results from Table 4.5, with both relevant and irrelevant video contents, the differences in perceived duration (*F* = 62.338, *p*-value = 0.0000) and negative attitudes toward advertisements (*F* = 23.386, *p*-value = 0.0000) between longer and shorter videos are significant.

As demonstrated in Table 4.6, during both longer and shorter advertisement videos, subjects perceive longer duration, generate more negative attitudes toward both advertisement videos and brands, and recall more contents from irrelevant advertisement videos than from relevant advertisement videos.

The differences of our measured dependent variables perceived duration (F = 3.452, p-value =0.06354; F = 70.775, p-value =0.0000;), negative effects for advertisement videos (F =4.1766, p-value = 0.0413; F = 3.4713, p-value = 0.0079), and item and brand recall (F = 230.8, p-value = 0.0000; F = 333.87, p-value = 0.0000) were statistically significant based on our ANOVA and MANOVA results in Table 4.7. The proposed hypotheses H1a, H1b, H2a, H2b, H4a, and H4b are all supported by the results. The difference of negative attitudes generated toward promoted brands during the long advertisements between relevant and irrelevant ad contents is significant. The difference for the short advertisement video is not statistically significant (F =5.1499, p-value = 0.0234; F = 0.747, p-value = 0.3876). Therefore, the only hypothesis that is not supported is H3b. The summary of hypotheses testing results is demonstrated in Table 4.8.

4.6.3 Study Results Discussion

According to time perception theories, memory-based model (Ornstein, 1969) can dominate people's perception during the shorter wait. Based on the schema incongruent effect (Berlyne, 1960; Houston et al., 1987; Pezdek et al., 1989), irrelevant and inconsistent content of advertisement videos can occupy more memory space of people than relevant content. Therefore, they create more memory cues for increasing people's duration and wait perception. Furthermore, when the attention-based model (McClain, 1983; Zakay, 1993) led people's wait perception (Block & Zakay, 1997; Hong et al., 2013), any distractions from the time can reduce the perception. Based on the contextual priming effect (Erdley & D'Agostino, 1988; Higgins et al., 1985; Wyer & Srull, 1981), the relevant content of advertisement videos can increase people's excitation level and therefore distract people's attentions away from time passage more than the irrelevant video content. The findings of this empirical study show that for both longer and shorter advertisement videos, if the contents of the video are highly related to the original desired video content, users can perceive less duration than irrelevant ad videos, which confirmed the referred time perception theories (Block & Zakay, 1997; Hong et al., 2013; McClain, 1983; Ornstein, 1969; Zakay, 1993), contextual priming effect (Erdley & D'Agostino, 1988; Higgins et al., 1985; Wyer & Srull, 1981), and schema incongruent effect (Berlyne, 1960; Houston et al., 1987; Pezdek et al., 1989).

Furthermore, based on the contextual priming effect (Erdley & D'Agostino, 1988; Higgins et al., 1985; Wyer & Srull, 1981), because the relevant content of advertisements can promote the excitation level of people, the negative effects of advertisements can be mitigated to some extent. The results from this study indicate that for both longer and shorter advertisement videos, if the contents of the video are related to the desired video content, users can generate less negative attitudes toward the advertisement than irrelevant advertisement videos, which supports the referred contextual priming effect (Higgins et al., 1985; Wyer & Srull, 1981).

Regarding the attitudes toward promoted brands, we only discover significant differences of negative attitudes between relevant and irrelevant advertisement videos during the long advertisement video. Even though users still generate more negative attitudes toward the brands promoted in short irrelevant than short relevant advertisement videos, the difference is not significant. This result does not verify the conclusion from prior studies (Gardner, 1985; Sallam & Algammash, 2016). This demonstrates that the negative attitudes toward advertisements do not necessarily impact the attitudes toward the brand, especially for the short advertisement video.

Even though the higher relevance level of advertisement video contents can result in shorter perceived waiting duration and less negative effects, the detrimental effect of the higher relevance level has also been strengthened. The results of this study confirm that users who watch irrelevant advertisement contents can recall more items and brands than users who watch relevant ad videos. This result further verifies the impact of the schema incongruent effect (Berlyne, 1960; Houston et al., 1987; Pezdek et al., 1989), which claims irrelevant information from the advertisement video can be harder to comprehend and interpret. Therefore, it requires longer processing time and larger memory space, and should be more elaborately processed (Houston et al., 1987; Pezdek et al., 1989).

4.7 Conclusion and Discussion

This study examines the impacts of different relevance levels of online advertisement video contents with the desired videos. The contextual relevance of traditional advertisements and online advertisements such as paper advertisements, TV commercials, and online pop-ins and banners have been widely discussed in prior studies. Since online video advertising is still a new advertising form, limited studies have examined the effects of the content relevance of online video advertisement on perceived video duration, negative attitude effects, and content recall. In our study, we employ time perception theories, contextual priming effect, and schema incongruent effect to evaluate the influence of content relevance levels of online video advertisements on duration perception, perceived negative users' attitude toward both advertisement and brands promoted, and recall of advertisement contents. One empirical study is conducted on Amazon Turk for evaluating different impacts.

In general, our results show support for RAM and the different time perception models during the retrospective paradigm and the prospective paradigm. Our study also verifies the paradigm shift during the longer online advertisement video. During the shorter advertisement video, in the retrospective paradigm, the memory-based time perception model dominates the duration perception. Our manipulated irrelevant advertisement video content requires larger memory space and longer processing time, and therefore is considered as additional memory cues that increase the video duration estimation. During the longer advertisement video, responding to the paradigm shift, the retrospective paradigm can change to the prospective paradigm. In the prospective paradigm, the attention-based time perception model directs the duration estimation. Relevant advertisement content can arouse a higher excitation level for users and direct attention away from time passage and reduce users' perception of video duration. On the basis of contextual priming effect and schema incongruent effect, because relevant advertisement content can arouse a higher excitation level for users, and the irrelevant advertisement video content requires larger memory space and longer processing time, we hypothesize the irrelevant advertisement video can result in more negative attitudes toward the video. The irrelevant advertisement video can also lead to more brand/product recalls due to longer elaborative processing time. Our experimental results support the higher level of negative effects toward the irrelevant advertisement video and confirm the beneficial impact of the irrelevant advertisement video on advertisement effectiveness.

In terms of practical implications, from advertisers' perspective, the purpose of creating and purchasing an advertisement video and placing it between users' desired video is to promote the product and brand to the viewers. Higher product and brand recall and recognition rate can be optimal and is the most beneficial result for advertisers. From users' perspective, when they are interrupted by the advertisement video during their desired video, since users are goal-oriented, they can generate generally negative and aggravated attitudes toward the advertisement video. The users can perceive longer advertisement duration and higher level of negative attitudes toward the video when the content of the advertisement video they watched does not match their desired video. To decrease the perceived duration, negative attitudes toward the advertisement, and even the negative impacts for promoted brands during the longer advertisement video, the advertiser might consider sacrificing the beneficial effects of the irrelevant advertisement video and selecting a contextual or thematically similar desired video in which to place their advertisement video.

There are a few limitations of this study. First, users may not fully enjoy and get involved in the NBA video as their desired video. Therefore, the negative attitudes toward the impedimental advertisement video are not as negative as in their real life. Second, we consider the NBA game video as our desired video and a few advertisement videos featuring NBA players promoting products such as sports drink, sneakers, and snacks as relevant advertisements. We, therefore, only match the similarity between the desired video and the advertisement video at contextual and thematically levels. The second type of relevance, which is the execution similarity between the videos, can be investigated in future studies.

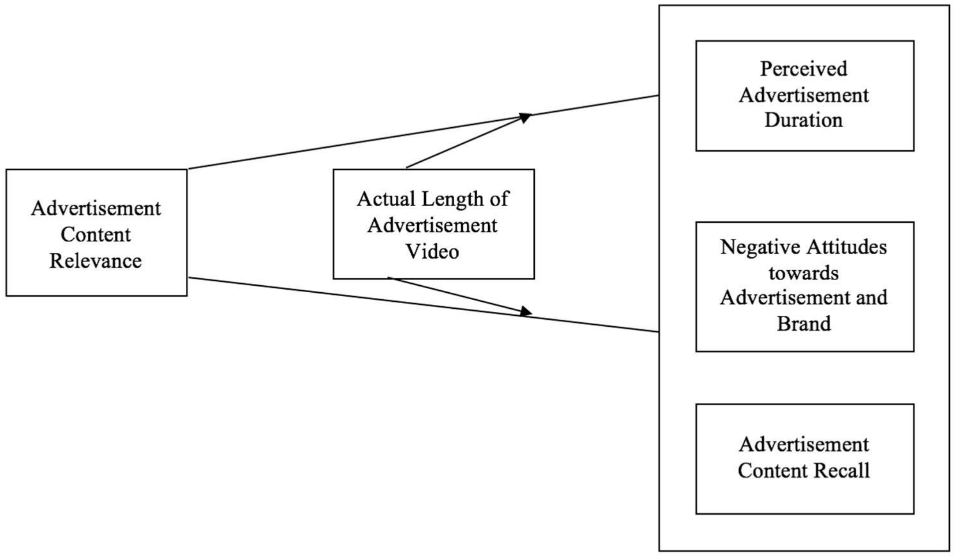

Figure 4.1: Research Model

Figure 4.2: Short Irrelevant Ad Video

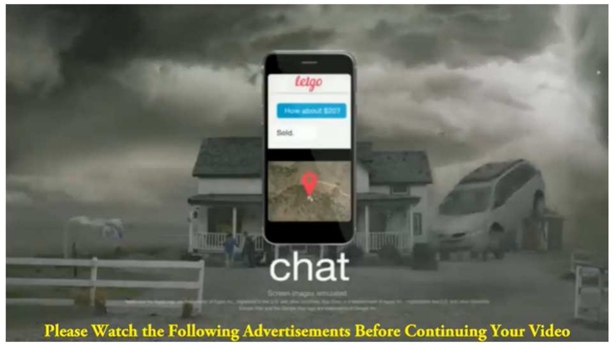

Figure 4.3: Long Irrelevant Ad Video

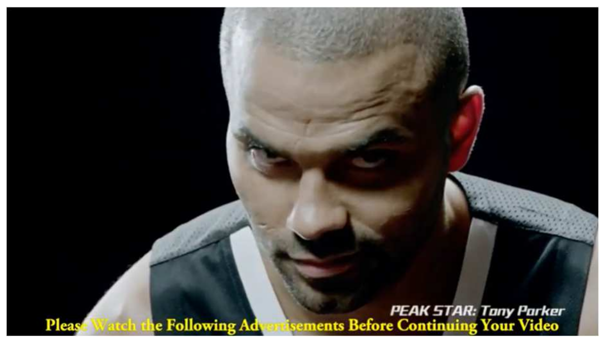

Figure 4.4: Short Relevant Ad Video

Figure 4.5: Long Relevant Ad Video

Figure 4.6: Manipulation Check of the Study

Table 4.1: Summary of Hypotheses

Hypothesis	Independent Variables	Dependent Variables	Hypotheses
H1a	Irrelevant ad content; Relevant ad content; (short ad)	Perceived duration of the ad video	H1a: For shorter video advertisements, an advertisement video with irrelevant content can lead to a longer perception of advertisement duration than an advertisement video with relevant content.
H1b	Irrelevant ad content; Relevant ad content; (long ad)	Perceived duration of the ad video	H1b: For longer video advertisements, an advertisement video with relevant content can lead to a shorter perception of advertisement duration than an advertisement video with irrelevant content.
H2a	Irrelevant ad content; Relevant ad content; (long ad)	Negative attitudes toward the ad video	H2a: For the long advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the advertisement than the advertisement with irrelevant content.
H2b	Irrelevant ad content; Relevant ad content; (short ad)	Negative attitudes toward the ad video	H2b: For the short advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the advertisement than the advertisement with irrelevant content.
H3a	Irrelevant ad content; Relevant ad content; (long ad)	Negative attitudes toward the brand promoted	H3a: For the long advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the brand promoted than the advertisement with irrelevant content.

Table 4.1: Continued

Hypothesis	Independent Variables	Dependent Variables	Hypotheses
H3b	Irrelevant ad content; Relevant ad content; (short ad)	Negative attitudes toward the brand promoted	H3b: For the short advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the brand promoted than the advertisement with irrelevant content.
H4a	Irrelevant ad content; Relevant ad content; (long ad)	Advertisement Contents Recall	H4a: For long advertisement video, customer perceive and recall less advertisement contents from the relevant advertisement video than the irrelevant advertisement video.
H4b	Irrelevant ad content; Relevant ad content; (short ad)	Advertisement Contents Recall	H4b: For the short advertisement video, customer perceive and recall less advertisement contents from the relevant advertisement video than the irrelevant advertisement video.

Table 4.2: Summary of the Empirical Study

	Study
Experiment Design	2 × 2 between subject
Focus	Examine the impacts of content relevance of the advertisement video on advertisement duration perception, negative attitudes toward advertisements/brands and advertisement contents recall
Independent Variables	Actual Advertisement Video Length (Shorter/Longer) Advertisement Video Content (Irrelevant, relevant)
Dependent Variables	The Perceived Duration of the Advertisement Video; The Negative Attitude Toward the Advertisement; The Negative Attitude Toward the Promoted Brand; The Content Recall of the Advertisement Video
Control Variables	Impatience Attribution
Analysis Method	ANOVA, MANOVA
Number of Subjects	2,508

Table 4.3: The Statistics of Subjects

Length	Number of Subjects
Long (relevant)	611
Long (irrelevant)	639
Short (relevant)	631
Short (irrelevant)	627

Table 4.4: Descriptive Statistics of Perceived Duration and Negative Attitudes

Factors (Actual Advertisement Duration)	Perceived Duration	Negative Attitudes
	Mean (Standard deviation)	Mean (Standard deviation)
Longer Advertisement Video	4.1021 (1.407)	3.8068 (1.8329)
Shorter Advertisement Video	3.4936 (1.3)	2.785 (1.6529)

Table 4.5: MANOVA and ANOVA Results of Actual Advertisement Duration

Factors (Advertisements with relevant content)	ANOVA		MANOVA	
	Perceived Duration		Negative Attitudes	
	F	Sig.	*F*	Sig.
Actual Advertisement Duration	62.338	6.355e-15***	23.386	< 2.2e-16 ***

Note. *** p<0.01, ** p<0.05, * p<0.1

Table 4.6: Descriptive Statistics of Irrelevant and Relevant Video Contents

Factors	Perceived Duration	Negative Attitudes Toward Ad Video	Negative Attitudes Toward Brand	Content Recall
	Mean (Standard Deviation)	Mean (Standard Deviation)	Mean (Standard Deviation)	Mean (Standard Deviation)
Irrelevant (long)	4.8407 (1.5154)	3.8915 (1.744)	3.8318 (1.8042)	5.2175 (2.088)
Relevant (long)	3.9668 (1.3744)	3.6382 (1.74)	3.6013 (1.766)	4.887 (2.058)
Irrelevant (short)	3.6561 (1.3761)	3.0049 (1.8503)	2.9757 (1.7784)	0.8501 (0.3573)
Relevant (short)	3.4789 (1.3425)	2.7593 (1.6659)	2.89 (1.7262)	0.6862 (0.4641)

Table 4.7: ANOVA and MANOVA Results of Irrelevant and Relevant Video Contents

Factors	ANOVA		MANOVA		MANOVA		MANOVA	
	Perceived Duration		Negative attitudes toward Ad		Negative attitudes toward Brand		Content Recall	
	F	Sig.	*F*	Sig.	*F*	Sig.	*F*	Sig.
Shorter Ad	3.452	0.0635*	4.1766	0.0413*	0.747	0.3876	230.8	<2.2e-16 ***
Longer Ad	70.775	< 2.2e-16 ***	3.4713	0.0079**	5.1499	0.0234*	333.87	<2.2e-16 ***

Note. *** p<0.01, ** p<0.05, * p<0.1

Table 4.8: Summary of Hypotheses Results

Hypothesis	Hypotheses	Results
H1a	H1a: For shorter video advertisements, an advertisement video with irrelevant content can lead to a longer perception of advertisement duration than an advertisement video with relevant content.	Supported
H1b	H1b: For longer video advertisements, an advertisement video with relevant content can lead to a shorter perception of advertisement duration than an advertisement video with irrelevant content.	Supported
H2a	H2a: For the long advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the advertisement than the advertisement with irrelevant content.	Supported
H2b	H2b: For the short advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the advertisement than the advertisement with irrelevant content.	Supported
H3a	H3a: For the long advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the brand promoted than the advertisement with irrelevant content.	Supported
H3b	H3b: For the short advertisement video, the relevant advertisement video content can motivate less negative attitudes toward the brand promoted than the advertisement with irrelevant content.	Rejected
H4a	H4a: For long advertisement video, customer perceive and recall less advertisement contents from the relevant advertisement video than the irrelevant advertisement video.	Supported
H4b	H4b: For the short advertisement video, customer perceive and recall less advertisement contents from the relevant advertisement video than the irrelevant advertisement video.	Supported

CHAPTER 5

CONCLUSION

In this dissertation, we propose three studies to learn users' opinions from different sources and to address specific research questions. We are firstly interested in discovering the significant factors that influence chain restaurants' rating differences from users' opinion learned from their generated reviews. To assist companies and their marketing departments to better design online video advertisements and receive optimal feedbacks from users, the two other studies focus on the impact of different temporal reminder designs and content relevance designs of advertisement videos on users' video duration perception, negative attitude effects, and ad content recalls. Two empirical, experimental studies are proposed to collect users' opinion and feedback after they watched advertisement videos with different temporal reminder designs and content relevance levels.

In the first study, to recognize the significant features affecting Yelp ratings, our study first aims at discovering the representative topics/aspects from unstructured user-generated reviews. We adopt a local LDA model that treats individual sentences as documents in order to avoid extracting general topics such as brand names and product names. To determine the optimal aspect number, we employ the cluster validation scheme to select the number of aspects with consistent performance. Our study then tries to discover the adjectives and verbs that can contain sentiment polarities. After constructing a word graph network connecting words with similar word embedding, and by proposing an objective function on the basis of label propagation method, we assign each extracted adjective and verb with a sentiment polarity probability distribution. The sentiment scores for individual aspects are further integrated from representative nouns modified by adjectives and verbs with sentiment probability distributions. Food quality, service, location, value, drink and dessert, and general opinion are identified as the significant impactors of the rating difference by using three different linear regression models. Furthermore, we employ predictive regressions for future Yelp rating prediction. The predictive models incorporating the sentiment scores from significant aspects as predictors achieve comparable lower error measures than including all possible aspects' sentiment score. By further extracting restaurant-related attributes from Yelp such as price range, credit cards, kids, and group as additional features for constructing predicative models, we can obtain the best performing models for rating prediction.

Besides the academic value shown here, the sentiment score of each aspect for restaurants also provides important practical implications for the chain restaurants. For example, knowing the sentiment score of drink and dessert aspect from specific Applebee's restaurants is lower than the average sentiment score of this certain aspect from all Applebee's restaurants, restaurant managers can hire new bakers with better performance so as to improve the performance of this particular aspect. Training programs can also be designated for employees if the restaurant knows it is the service that pulls the ratings back.

The limitations of the first study are discussed in three aspects. First of all, we only evaluate the top 81 chain restaurants' reviews and ratings from Yelp.com. Limited reviews are extracted from certain restaurants. The data sparsity problem might affect our model performance and further empirical analysis. In addition, Yelp.com is the only review website from which we extract ratings and reviews. The rating differences of chain restaurants might result from other factors if we use alternative business review websites such as TripAdvisor, Yahoo Local, and Angie's List. It is therefore important to extend the variety of data source. Furthermore, even though the aspect sentiment analysis is an unsupervised learning approach, our study does not require any supervised data at the model building phase. Labeled data of aspects and sentiment scores for model evaluation are still necessary. There are possibilities that the performance of our proposed aspect sentiment model is not good. Therefore, the model performance can be improved if the method can be evaluated by labeled data.

In the second study, our study effectively applies time perception theories to the online video advertisements context and extends the analysis to temporal reminder designs during advertisement videos that have not been studied before. In general, our results show support for RAM and the distinctive time perception models during the retrospective paradigm and the prospective paradigm. Our study also verifies the paradigm shift during the longer online advertisement video. During the shorter advertisement video, in the retrospective paradigm, the memory-based time perception model dominates the duration perception, and our temporal designs are considered as additional memory cues that increase the video duration estimation. During the longer advertisement video, responding to the paradigm shift, retrospective paradigm can change to the prospective paradigm. In the prospective paradigm, the attention-based time perception model directs that the duration estimation and temporal designs such as counting down clock and progressing bar indicator lead attention toward time passage and increase users' perception of video duration. The remaining number of ads reminder distracts users' attention away from time and to an advertisement video during the longer video, and therefore result in shorter video duration perceptions than the other two temporal reminder designs. On the basis of reactance theory, we assume temporal reminder designs can increase the negative attitudes during the shorter advertisement since they are the factors that enlarge the perceived duration and reduce the negative attitudes during the longer advertisement video due to time uncertainty assurance. The empirical study does not support our hypotheses, possibly due to the two unmeasured conflicting attitude effects of the temporal designs. In line with selective attention theories, we presume the temporal reminder designs as users' attention distractions from advertisement videos and therefore reduce the advertisement content recalls. However, the experimental results do not support our hypothesis. This is probably because the temporal reminders are not the important information that people filtered to pay attention to and they do not significantly distract people's attentions from the advertisement contents.

There are a few limitations of the second study. First, we use a simulated desired video and a simulated advertisement insertion scenario. Users may not fully enjoy and get involved in the NBA video as their desired video. Therefore, the negative attitudes toward the impedimental advertisement video are not as negative as in their real life. Second, during this experimental study, we require participants to watch the video carefully without skipping and forwarding. They understand this is an experiment and there are related questions afterward. Therefore, they may pay much more attention to watching the advertisements, which leads to indistinguishable good ads recall results under different circumstances.

In the last study, we employ time perception theories, contextual priming effect, and schema incongruent effect to evaluate the influence of content relevance levels of online video advertisements on duration perception, perceived negative users' attitude toward both advertisements and brands promoted, and recall of advertisement contents. One empirical study is conducted on Amazon Turk for evaluating different impacts.

In general, the results from the last study show support for RAM and the different time perception models during the retrospective paradigm and the prospective paradigm. Our study also verifies the paradigm shift during the longer online advertisement video. During the shorter advertisement video, in the retrospective paradigm, the memory-based time perception model dominates the duration perception. Our manipulated irrelevant advertisement video content requires larger memory space and longer processing time, and is therefore considered as additional memory cues that increase the video duration estimation. During the longer advertisement video, responding to the paradigm shift, the retrospective paradigm can change to the prospective paradigm. In the prospective paradigm, the attention-based time perception model directs the duration estimation. Relevant advertisement content can arouse a higher excitation level for users and direct attention away from time passage and reduce users' perception of video duration. On the basis of the contextual priming effect and the schema incongruent effect, because relevant advertisement content can arouse a higher excitation level for users, and the irrelevant advertisement video content require larger memory space and longer processing time, we hypothesize the irrelevant advertisement video can result in more negative attitudes toward the video. The irrelevant advertisement video can also lead to more brand/product recalls due to longer elaborative processing time. Our experimental results support the higher level of negative effects toward the irrelevant advertisement video and confirm the beneficial impact of the irrelevant advertisement video on advertisement effectiveness.

In terms of practical implications, from advertisers' perspective, the purpose of creating and purchasing an advertisement video and placing it between users' desired video is to promote the product and brand to the viewers. Higher product and brand recall and recognition rate can be optimal and is the most beneficial result for advertisers. From users' perspective, when they are interrupted by an advertisement video during their desired video, since users are goal-oriented, they can generate generally negative and aggravated attitudes toward the advertisement video. The users can perceive longer advertisement duration and higher level of negative attitudes toward the video when the content of the advertisement video they watched does not match their desired video. To decrease the perceived duration, negative attitudes toward the advertisement, and even the negative impacts for promoted brands during the longer advertisement video, the advertiser might consider sacrificing the beneficial effects of the irrelevant advertisement video and selecting a contextual or thematically similar desired video in which to place their advertisement video.

Regarding practical implications, from advertisers' perspective, the purpose of creating and purchasing an advertisement video and placing it between users' desired video is to promote the product and brand to the viewers. Higher product and brand recall and recognition rate can be optimal and is the most beneficial result for advertisers. On the other hand, from users' perspective, when they are interrupted by an advertisement video during their desired video, since users are goal-oriented, they can generate generally negative and aggravated attitudes toward the advertisement video. The users can perceive longer advertisement duration and the higher level of negative attitudes toward the video when the content of the advertisement video they watched does not match their desired video. To decrease the perceived duration, negative attitudes toward the advertisement, and even the negative impacts for promoted brands during the longer advertisement video, the advertiser might consider sacrificing the beneficial effects of the irrelevant advertisement video (better brand and product recall) and selecting a contextual or thematically similar desired video in which to place their advertisement video.

The last study also suffers from a few limitations. First, users may not fully enjoy and get involved in the NBA video as their desired video. Therefore, the negative attitude toward the impedimental advertisement video are not as negative as in their real life. Second, we consider the NBA game video as our desired video and a few advertisement videos featuring NBA players promoting products such as sports drink, sneakers, and snacks as relevant advertisements. We, therefore, only match the similarity between the desired video and the advertisement video at the contextual and thematically level. The second type of relevance, which is the execution similarity between the videos, can be investigated in the future studies.

REFERENCES

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems (TOIS)*, *26*(3), 12.

Andaleeb, S. S., & Caskey, A. (2007). Satisfaction with food services: Insights from a college cafeteria. *Journal of Foodservice Business Research*, *10*(2), 51-65.

Antonides, G., Verhoef, P. C., & Van Aalst, M. (2002). Consumer perception and evaluation of waiting time: A field experiment. *Journal of Consumer Psychology*, *12*(3), 193-202.

Azeem, A., & ul Haq, Z. (2012). Perception towards internet advertising: A study with reference to three different demographic groups. *Global Business and Management Research*, *4*(1), 28.

Barsky, J. D. (1992). Customer satisfaction in the hotel industry: Meaning and measurement. *Hospitality Research Journal*, *16*(1), 51-73.

Berlyne, E. D. (1960). *Conflict, arousal, and curiosity*. New York: McGraw-Hill.

Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G. A., & Reynar, J. (2008). Building a sentiment summarizer for local service reviews. *Proceedings of Workshop on NLP in the Information Explosion Era*, *14,* 339-348. doi: 10.1.1.182.4520

Blair, M. H., & Kuse, A. R. (2004). Better practices in advertising can change a cost of doing business to wise investments in the business. *Journal of Advertising Research*, *44*(1), 71-89.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993-1022.

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, *4*(2), 184-197.

Brackett, L. K., & Carr, B. N. (2001). Cyberspace advertising vs. other media: Consumer vs. mature student attitudes. *Journal of Advertising Research*, *41*(5), 23-32.

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, *14*(1), 63-82.

Braun-LaTour, K. A., Puccinelli, N. M., & Mast, F. W. (2007). Mood, information congruency, and overload. *Journal of Business Research*, *60*(11), 1109-1116.

Brehm, J. W. (1966). *A theory of psychological reactance.* New York: Academic Press.

Broadbent, D (1958). *Perception and communication*. London: Pergamon Press.

Brody, S., & Elhadad, N. (2010, June). An unsupervised aspect-sentiment model for online reviews. *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 804-812. doi: 10.1007/978-3-642-38824-8_12

Calvo, M. G., Castillo, M. D., & Estevez, A. (1999). On-line predictive inferences in reading: Processing time during versus after the priming context. *Memory & Cognition*, *27*(5), 834-843.

Cameron, M. A., Baker, J., Peterson, M., & Braunsberger, K. (2003). The effects of music, wait-length evaluation, and mood on a low-cost wait experience. *Journal of Business Research*, *56*(6), 421-430.

Campbell, M. C. (1995). When attention-getting advertising tactics elicit consumer inferences of manipulative intent: The importance of balancing benefits and investments. *Journal of Consumer Psychology*, *4*(3), 225-254.

Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., & Shneiderman, B. (2004). Determining causes and severity of end-user frustration. *International Journal of Human-Computer Interaction*, *17*(3), 333-356.

Chebat, J. C., Salem, N. H., Poirier, J. F., & Gélinas-Chebat, C. (2010). Reactions to waiting online by men and women. *Psychological Reports*, *106*(3), 851-869.

Coulter, K. S. (1998). The effects of affective responses to media context on advertising evaluations. *Journal of Advertising*, *27*(4), 41-51.

Dabholkar, P. A., & Sheng, X. (2008). Perceived download waiting in using web sites: A conceptual framework with mediating and moderating effects. *Journal of Marketing Theory and Practice*, *16*(3), 259-270.

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the 12th International Conference on World Wide Web,* 519-528. doi: 10.1145/775152.775226

Dellaert, B. G., & Kahn, B. E. (1999). How tolerable is delay: Consumers' evaluations of internet web sites after waiting. *Journal of Interactive Marketing*, *13*(1), 41-54.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629.

Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 231-240. doi: 10.1145/1341531.1341561

Dube, L., Schmitt, B. H., & Leclerc, F. (1991). Consumers' affective response to delays at different phases of a service delivery. *Journal of Applied Social Psychology*, *21*(10), 810-820.

Ducoffe, R. H. (1995). How consumers assess the value of advertising. *Journal of Current Issues & Research in Advertising*, *17*(1), 1-18.

Edwards, S. M., Li, H., & Lee, J. H. (2002). Forced exposure and psychological reactance: Antecedents and consequences of the perceived intrusiveness of pop-up ads. *Journal of Advertising*, *31*(3), 83-95.

eMarketer. Data and Research on Digital for Business Professionals. (2012). *Ads disguised as content mislead, annoy.* Retrieved from http://www.mediabrix.com/wp-content/uploads/2014/02/eMarketer_Ads_Disguised_as_Content_Mislead_Annoy_1009478.pdf

Erdley, C. A., & D'Agostino, P. R. (1988). Cognitive and affective components of automatic priming effects. *Journal of Personality and Social Psychology*, *54*(5), 741–747.

Eriksen, C. W., & Hoffman, J. E. (1972). Some characteristics of selective attention in visual perception determined by vocal reaction time. *Attention, Perception, & Psychophysics*, *11*(2), 169-171.

Francis-Smythe, J., & Robertson, I. (1999). Time-related individual differences. *Time & Society*, *8*(2-3), 273-292.

Gardner, M. P. (1985). Does attitude toward the ad affect brand attitude under a brand evaluation set? *Journal of Marketing Research*, *22*(2), 192-198.

Goldberg, A. B., & Zhu, X. (2006, June). Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, 45-52. doi: 10.1145/585265.585268

Goldfarb, A., & Tucker, C. (2011). Online display advertising: Targeting and obtrusiveness. *Marketing Science*, *30*(3), 389-404.

Goodrich, K., Schiller, S. Z., & Galletta, D. (2015). Consumer reactions to intrusiveness of online-video advertisements. *Journal of Advertising Research*, *55*(1), 37-50.

Gorn, G. J., Chattopadhyay, A., Sengupta, J., & Tripathi, S. (2004). Waiting for the web: How screen color affects time perception. *Journal of Marketing Research*, *41*(2), 215-225.

Griffiths, T. L., & Steyvers, M. (2003). Prediction and semantic association. *Advances in Neural Information Processing Systems*, *15*, 11-18.

Groth, M., & Gilliland, S. W. (2006). Having to wait for service: Customer reactions to delays in service delivery. *Applied Psychology*, *55*(1), 107-129.

Ha, J., & Jang, S. S. (2010). Effects of service quality and food quality: The moderating role of atmospherics in an ethnic restaurant segment. *International Journal of Hospitality Management*, *29*(3), 520-529.

Haghighi, M., Dorosti, A., Rahnama, A., & Hoseinpour, A. (2012). Evaluation of factors affecting customer loyalty in the restaurant industry. *African Journal of Business Management*, *6*(14), 5039.

Han, H., & Ryu, K. (2009). The roles of the physical environment, price perception, and customer satisfaction in determining customer loyalty in the restaurant industry. *Journal of Hospitality & Tourism Research*, *33*(4), 487-510.

Hegner, S. M., Kusse, M. D. C., & Pruyn, A. T. (2016). Watch it! The influence of forced pre-roll video ads on consumer perceptions. *Proceedings of Advances in Advertising Research, VI,* 63-73. doi: 10.1007/978-3-658-10558-7_6

Higgins, E. T., Bargh, J. A., & Lombardi, W. (1985). Nature of priming effects on categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 59–69.

Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *The Journal of Marketing*, *60*, 50-68.

Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 50-57. doi: 10.1145/312624.312649

Hong, W., Hess, T. J., & Hardin, A. (2013). When filling the wait makes it feel longer: A paradigm shift perspective for managing online delay. *Management Information Systems Quarterly*, *37*(2), 383-406.

Houston, M. B., Bettencourt, L. A., & Wenger, S. (1998). The relationship between waiting in a service queue and evaluations of service quality: A field theory perspective. *Psychology & Marketing*, *15*(8), 735-753

Houston, M. J., Childers, T. L., & Heckler, S. E. (1987). Picture-word consistency and the elaborative processing of advertisements. *Journal of Marketing Research*, *24*(4), 359–369.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177. doi: 10.3115/1118693.1118704

Hui, M. K., Dube, L., & Chebat, J. C. (1997). The impact of music on consumers' reactions to waiting for services. *Journal of Retailing*, *73*(1), 87-104.

Hui, M. K., & Tse, D. K. (1996). What to tell consumers in waits of different lengths: An integrative model of service evaluation. *The Journal of Marketing*, *60*(April), 81-90.

IAB, Internet Advertising Revenue Report. (2013). *The native advertising playbook*. Retrieved from http://www.iab.net/media/file/IAB- Native-Advertising-Playbook2.pdf

Ikeda, D., Takamura, H., Ratinov, L. A., & Okumura, M. (2008). Learning to shift the polarity of words for sentiment classification. *Proceedings of the International Joint Conference on Natural Language Processing*, 296-303. doi: 10.1527/tjsai.25.50

Jakob, N., & Gurevych, I. (2010). Extracting opinion targets in a single-and cross-domain setting with conditional random fields. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1035-1045. doi: 10.1016/j.ipm.2005.09.002

Jang, S. S., & Namkung, Y. (2009). Perceived quality, emotions, and behavioral intentions: Application of an extended Mehrabian–Russell model to restaurants. *Journal of Business Research*, *62*(4), 451-460.

Jeong, Y., & King, C. M. (2010). Impacts of website context relevance on banner advertisement effectiveness. *Journal of Promotion Management*, *16*(3), 247-264.

Jia, L., Yu, C., & Meng, W. (2009). The effect of negation on sentiment analysis and retrieval effectiveness. *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 1827-1830. doi: 10.1145/1645953.1646241

Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). Target-dependent twitter sentiment classification. *Proceedings of the 49th Annual Meeting of the*

Association for Computational Linguistics: Human Language Technologies-Volume 1, 151-160. doi: 10.1023/A:1009715923555

Jin, W., Ho, H. H., & Srihari, R. K. (2009). A novel lexicalized HMM-based learning framework for web opinion mining. *Proceedings of the 26th Annual International Conference on Machine Learning,* 465-472. doi: 10.1145/1553374.1553435

Jo, Y., & Oh, A. H. (2011). Aspect and sentiment unification model for online review analysis. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 815-824. doi: 10.1145/1935826.1935932

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Kamins, M. A., Marks, L. J., & Skinner, D. (1991). Television commercial evaluation in the context of program induced mood: Congruency versus consistency effects. *Journal of Advertising*, *20*(2), 1-14.

Katz, K. L., Larson, B. M., & Larson, R. C. (1991). Prescription for the waiting-in-line blues: Entertain, enlighten, and engage. *MIT Sloan Management Review*, *32*(2), 44.

Kelman, H. C. (1958). Compliance, identification, and internalization three processes of attitude change. *Journal of Conflict Resolution*, *2*(1), 51-60.

Kim, N. Y., & Sundar, S. S. (2012). Personal relevance versus contextual relevance. *Journal of Media Psychology*, *24*(3), 89-101.

Kim, S. (2015). *Effects of ad-video similarity, ad location, and user control option on ad avoidance and advertiser-intended outcomes of online video ads.* (Unpublished doctoral dissertation) University of Minnesota, Minneapolis, USA.

Kim, S. M., & Hovy, E. (2004). Determining the sentiment of opinions. *Proceedings of the 20th International Conference on Computational Linguistics*, 1367. doi: 10.3115/1220355.1220555

Kim, W., & Han, H. (2008). Determinants of restaurant customers' loyalty intentions: A mediating effect of relationship quality. *Journal of Quality Assurance in Hospitality & Tourism*, *9*(3), 219-239.

Kobayashi, N., Iida, R., Inui, K., & Matsumoto, Y. (2006). Opinion mining on the web by extracting subject-attribute-value relations. *Proceedings of AAAI 2006 Spring Sympoia on Computational Approaches to Analyzing Weblogs, 2006.* doi: 10.1007/11780496_49

Kobayashi, N., Inui, K., & Matsumoto, Y. (2007). Extracting aspect-evaluation and aspect-of relations in opinion mining. *Proceedings of the Conference on Empirical Methods on Natural Language Processing and Computational Natural Language Learning*, *7*, 1065-1074. doi: 10.1.1.76.1394

Krishnan, S. S., & Sitaraman, R. K. (2013). Understanding the effectiveness of video ads: A measurement study. *Proceedings of the 2013 Conference on Internet Measurement Conference*, 149-162. doi: 10.13140/2.1.3719.4562

Krugman, H. E. (1983). Television program interest and commercial interruption. *Journal of Advertising Research*, *23*(11), 21-23.

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

Kusse, D. C. (2013). *Forced pre-roll video-advertising-Length, control, emotional appeal and contextual overlap of pre-roll video-ads on perceived intrusiveness, attitudes and skipping.* (Unpublished master's thesis) University of Twente, Enschede, Netherlands.

Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the Eighteenth International Conference on Machine Learning, 1*, 282-289. doi: 10.1.1.26.803

Lange, T., Roth, V., Braun, M. L., & Buhmann, J. M. (2004). Stability-based validation of clustering solutions. *Neural Computation*, *16*(6), 1299-1323.

Levine, E., & Domany, E. (2001). Resampling method for unsupervised estimation of cluster validity. *Neural Computation*, *13*(11), 2573-2593.

Lewis, R. C. (1981). Restaurant advertising-appeals and consumers intentions. *Journal of Advertising Research*, *21*(5), 69-74.

Li, H., & Lo, H. Y. (2015). Do you recognize its brand? The effectiveness of online in-stream video advertisements. *Journal of Advertising*, *44*(3), 208-218.

Li, S., Wang, Z., Zhou, G., & Lee, S. Y. M. (2011). Semi-supervised learning for imbalanced sentiment classification. *Proceedings of International Joint Conference on Artificial Intelligence, 22*(3), 1826. doi: 10.5591/978-1-57735-516-8/IJCAI11-306

Lin, C., & He, Y. (2009). Joint sentiment/topic model for sentiment analysis. *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 375-384. doi: 10.1145/1361684.1361685

Logan, K. (2013). And now a word from our sponsor: Do consumers perceive advertising on traditional television and online streaming video differently? *Journal of*

Marketing Communications, *19*(4), 258-276.

Luca, M. (2016). *Reviews, reputation, and revenue: The case of Yelp. com*. Working Paper (Harvard Business School).

Lunardo, R., & Mbengue, A. (2013). When atmospherics lead to inferences of manipulative intent: Its effects on trust and attitude. *Journal of Business Research*, *66*(7), 823-830.

Maliov, L. (2012). *The effects of musical tempo on time perception and memory*. (Unpublished doctoral dissertation) Manchester Metropolitan University, Manchester, UK.

Mandler, G. (1975). *Mind and emotion*. Malabar, India: Krieger Publishing Company.

Manevitz, L. M., & Yousef, M. (2001). One-class SVMs for document classification. *Journal of Machine Learning Research*, *2*(Dec), 139-154.

Mcclain, L. (1983). Interval estimation: Effect of processing demands on prospective and retrospective reports. *Perception & Psychophysics*, *34*(2), 185-189.

McCoy, S., Everard, A., Galletta, D., & Polak, P. (2004). A study of the effects of online advertising: A focus on pop-up and in-line ads. *Proceedings of Special Interest Group on Human-Computer Interaction 2004*, 11. doi: 10.1.1.93.5614

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. Cambridge: The MIT Press.

Mehta, A. (2000). Advertising attitudes and advertising effectiveness. *Journal of Advertising Research*, *40*(3), 67-72.

Mei, T., Hua, X. S., & Li, S. (2009). VideoSense: A contextual in-video advertising system. *IEEE Transactions on Circuits and Systems for Video Technology*, *19*(12), 1866-1879.

Mei, T., Hua, X. S., Yang, L., & Li, S. (2007). VideoSense: towards effective online video advertising. *Proceedings of the 15th ACM International Conference on Multimedia*, 1075-1084. doi: 10.1145/1291233.1291467

Meng, J. G., & Elliott, K. M. (2008). Predictors of relationship quality for luxury restaurants. *Journal of Retailing and Consumer Services*, *15*(6), 509-515.

Moore, R. S., Stammerjohan, C. A., & Coulter, R. A. (2005). Banner advertiser-web site context congruity and color effects on attention and attitudes. *Journal of Advertising*, *34*(2), 71-84.

Moorman, M., Neijens, P. C., & Smit, E. G. (2002). The effects of magazine-induced psychological responses and thematic congruence on memory and attitude toward the ad in a real-life setting. *Journal of Advertising*, *31*(4), 27-40.

Morante, R., Schrauwen, S., & Daelemans, W. (2011). Corpus-based approaches to processing the scope of negation cues: An evaluation of the state of the art. *Proceedings of the Ninth International Conference on Computational Semantics*, 350-354. doi: 10.1016/j.jbi.2010.08.003

Namkung, Y., & Jang, S. (2008). Are highly satisfied restaurant customers really different? A quality perception perspective. *International Journal of Contemporary Hospitality Management*, *20*(2), 142-155.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2010). Recognition of affect, judgment, and appreciation in text. *Proceedings of the 23rd International Conference on Computational Linguistics*, 806-814. doi: 10.1.1.175.2795

Ng, A. Y., Jordan, M. I., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Proceedings of Neural Information Processing Systems, 14*(2), 849-856. doi: 10.1.1.19.8100

Niu, Z. Y., Ji, D. H., & Tan, C. L. (2007). I2r: Three systems for word sense discrimination, Chinese word sense disambiguation, and English word sense disambiguation. *Proceedings of the 4th International Workshop on Semantic Evaluations*, 177-182. doi: 10.3115/1118693.1118699

Ornstein, R. E. (1969). *On the experience of time*. London: Penguin.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, 271. doi: 10.3115/1218955.1218990

Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 115-124. doi: 10.3115/1219840.1219855

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up: Sentiment classification using machine learning techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, 10,* 79-86. doi: 10.3115/1118693.1118704

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

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing*, *14,* 1532-1543. Retrieved from https://www.aclweb.org/anthology/D14-1162

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 587– 595.

Popescu, A. M., & Etzioni, O. (2005). Extracting product features and opinions from reviews. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing,* 339-346. doi: 10.3115/1220575.1220618

Porta, M., Ravarelli, A., & Spaghi, F. (2013). Online newspapers and ad banners: An eye tracking study on the effects of congruity. *Online Information Review*, *37*(3), 405-423.

Potamias, M. (2012). The warm-start bias of Yelp ratings. *arXiv preprint arXiv:1202.5713*. Retrieved from https://arxiv.org/abs/1202.5713

Previte, J., & Forrest, E. (1998). *Internet advertising: An assessment of consumer attitudes to advertising on the Internet.* Paper presented at the 1998 Communications Research Forum, Canberra, Australia.

Priestly, J. B. (1968). *Man and time*. New York: Dell.

Pruyn, A., & Smidts, A. (1998). Effects of waiting on the satisfaction with the service: Beyond objective time measures. *International Journal of Research in Marketing*, *15*(4), 321-334.

Qu, L., Ifrim, G., & Weikum, G. (2010). The bag-of-opinions method for review rating prediction from sparse text patterns. *Proceedings of the 23rd International Conference on Computational Linguistics*, 913-921. doi: 10.1.1.229.1053

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the Institute of Electrical and Electronics Engineers*, *77*(2), 257-286. doi: 10.1109/5.18626

Ribeiro Soriano, D. (2002). Customers' expectations factors in restaurants: The situation in Spain. *International Journal of Quality & Reliability Management*, *19*(8/9), 1055-1067.

Rodgers, S. (2003). The effects of sponsor relevance on consumer reactions to internet sponsorships. *Journal of Advertising*, *32*(4), 67-76.

Rose, G. M., Meuter, M. L., & Curran, J. M. (2005). On-line waiting: The role of download time and other important predictors on attitude toward e-retailers. *Psychology & Marketing*, *22*(2), 127-151.

Russell, J. A., & Pratt, G. (1980). A description of the affective quality attributed to environments. *Journal of Personality and Social Psychology*, *38*(2), 311.

Sallam, M. A., & Algammash, F. A. (2016). The effect of attitude toward advertisement on attitude toward brand and purchase intention. *International Journal of Economics, Commerce and Management*, *4*(2), 509-520.

Simola, J., Kivikangas, M., Kuisma, J., & Krause, C. M. (2013). Attention and memory for newspaper advertisements: Effects of ad–editorial congruency and location. *Applied Cognitive Psychology*, *27*(4), 429-442.

Snyder, B., & Barzilay, R. (2007). Multiple aspect ranking using the good grief algorithm. *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 300-307. doi: 10.1.1.129.4132

Somasundaran, S., & Wiebe, J. (2009). Recognizing stances in online debates. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 1,* 226-234. doi: 10.3115/1225733.1225750

Spence, J. T., Helmreich, R. L., & Pred, R. S. (1987). Impatience versus achievement strivings in the type a pattern: Differential effects on students' health and academic achievement. *Journal of Applied Psychology*, *72*(4), 522.

Stanton, J. L., & Burke, J. (1998). Comparative effectiveness of executional elements in TV advertising: 15- versus 30-second commercials. *Journal of Advertising Research*, *38*(6), 7-8.

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of Latent Semantic Analysis*, *427*(7), 424-440.

Teixeira, T., Wedel, M., & Pieters, R. (2012). Emotion-induced engagement in internet video advertisements. *Journal of Marketing Research*, *49*(2), 144-159.

Titov, I., & McDonald, R. (2008). Modeling online reviews with multi-grain topic models. *Proceedings of the 17th International Conference on World Wide Web*, 111-120. doi: 10.1145/1367497.1367513

Tutaj, K., & van Reijmersdal, E. A. (2012). Effects of online advertising format and persuasion knowledge on audience reactions. *Journal of Marketing Communications*, *18*(1), 5-18.

Van Reijmersdal, E., Neijens, P., & Smit, E. (2005). Readers' reactions to mixtures of advertising and editorial content in magazines. *Journal of Current Issues & Research in Advertising*, *27*(2), 39-53.

Whiting, A., & Donthu, N. (2006). Managing voice-to-voice encounters: Reducing the agony of being put on hold. *Journal of Service Research*, *8*(3), 234-244.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 347-354. doi: 10.3115/1220575.1220619

Wu, Y., Zhang, Q., Huang, X., & Wu, L. (2009). Phrase dependency parsing for opinion mining. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 3,* 1533-1541. Retrieved from https://dl.acm.org/citation.cfm?id=1699700

Wyer, Jr., R. S., & Srull, T. K. (1981). Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In E. T. Higgins, C. P. Herman, & M. P. Zanna, (Eds.), *Social cognition: The Ontario symposium* (pp. 161–197). Hillsdale, NJ: Lawrence Erlbaum.

Yaveroglu, I., & Donthu, N. (2008). Advertising repetition and placement issues in on-line environments. *Journal of Advertising*, *37*(2), 31-44.

Yi, Y. (1990a). Cognitive and affective priming effects of the context for print advertisements. *Journal of Advertising*, *19*(2), 40–48.

Yi, Y. (1990b). The effects of contextual priming in print advertisements. *Journal of Consumer Research*, *17*(2), 215–222.

Ying, L., Korneliussen, T., & Grønhaug, K. (2009). The effect of ad value, ad placement and ad execution on the perceived intrusiveness of web advertisements. *International Journal of Advertising*, *28*(4), 623-638.

Yüksel, A., & Yüksel, F. (2003). Measurement of tourist satisfaction with restaurant services: A segment-based approach. *Journal of Vacation Marketing*, *9*(1), 52-68.

Zakay, D. (1989). Subjective time and attentional resource allocation: An integrated model of time estimation. *Advances in Psychology*, *59*, 365-397.

Zakay, D. (1993). Time estimation methods - do they influence prospective duration estimates? *Perception*, *22*(1), 91-101.

Zanjani, S. H., Diamond, W. D., & Chan, K. (2011). Does ad-context congruity help surfers and information seekers remember ads in cluttered e-magazines? *Journal* *of Advertising*, *40*(4), 67-84.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). *Delivering quality service: Balancing customer perceptions and expectations*. New York: Simon and Schuster.

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

Zhu, X., & Ghahramani, Z. (2002). *Learning from labeled and unlabeled data with label propagation* (Technical Report CMU-CALD-02-107). Retrieved from Carnegie Mellon University: http://mlg.eng.cam.ac.uk/zoubin/papers/CMU-CALD-02-107.pdf

Zhuang, L., Jing, F., & Zhu, X. Y. (2006). Movie review mining and summarization. *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, 43-50. doi: 10.1145/944012.944013

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.