**Distribution Agreement**



Signature:

____________________________ ____________________

Eunhee (Emily) Ko Date

Two Essays on Content Engineering with Unstructured Data:
Business Insights from User-Generated Content

By

Eunhee (Emily) Ko
Doctor of Philosophy
Business

Douglas Bowman, Ph.D.
Committee Chair

Diego Klabjan, Ph.D.
Committee Member

Zhongjian Lin, Ph.D.
Committee Member

Daniel McCarthy, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Two Essays on Content Engineering with Unstructured Data:
Business Insights from User-Generated Content

By

Eunhee (Emily) Ko
B.A. in Psychology, Yonsei University
M.S. in Marketing, University of Cincinnati
M.S. in Analytics, Northwestern University

Advisor: Douglas Bowman, Ph.D.

An abstract of
A Dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Marketing
2019

Abstract

Two Essays on Content Engineering with Unstructured Data:
Business Insights from User-Generated Content
By Eunhee (Emily) Ko

A primary driver behind the topics in my dissertation essays is the desire to address the challenges that marketing practitioners front into the market environment, where consumer behaviors are changing quickly with the expansion of platforms into new media that are native to computers or mobile devices, which have prompted continuous growth in marketing expenditures. While there is a wide range of research that studies user-generated content (UGC) and its impact on marketing or consumer purchasing behavior, few studies highlight the content characteristics with large-scale data from the field. Moreover, most of the existing empirical research that studies the semanticity of UGC pays limited attention to content beyond the text. To fill this gap, I have initiated and advanced several projects to investigate the content features not only from *texts* but *images* in my Ph.D. program. In doing so, I bring a variety of methodological approaches to my research (natural language processing, machine learning, and image processing techniques), having merged public and proprietary datasets – both longitudinal and cross-sectional.

The first essay of my dissertation examines consumer engagement, measured as the number of likes and comments tied to a brand-themed social media post on Instagram. I study consumer engagement with brand-themed user-generated content – imaged-based social media posts tagged with #*brandname* – an increasingly common way that consumers engage with brands. I describe consumer engagement using characteristics of the image and the text of a post – visual sentiment, visual complexity, text sentiment, and text complexity – which I craft using techniques that include deep convolutional neural networks (Deep CNNs), and both a computer vision application programming interface (API) and natural language processing (NLP). Using data from over 86,000 Instagram posts collectively hashtagged with 86 product brand names, I find that visual sentiment and text sentiment are positively associated with higher levels of consumer engagement. Visual complexity and text complexity both positively affect consumer engagement at low and moderate levels, and become negative at high levels. Too much information either from images or from texts attenuates consumer engagement. Around the middle of the range of visual complexity there is an optimal level that makes a post rich and engaging.

The second essay of my dissertation investigates factors that characterize manipulated reviews by concentrating on unstructured text data and brand strength as a factor associated with suspicious online review incidences. Studying over 270,000 Amazon.com reviews from 16 product categories, I find that approximately 3% of reviews are ones consumers would be suspicious about. Extreme emotions (e.g., fear, joy) account for a review being viewed as suspicious better than mixed emotions (e.g., anticipation, surprise) or low-arousal emotions (e.g., sadness). I argue that weaker brands have an incentive for review manipulation. I find that a weak brand status, described by lower advertising effort, is associated with suspicious reviews that are promotional (positive) in nature. Though, the effect fades away for suspicious reviews that are denigrating (negative).

Two Essays on Content Engineering with Unstructured Data:
Business Insights from User-Generated Content

By

Eunhee (Emily) Ko
B.A. in Psychology, Yonsei University
M.S. in Marketing, University of Cincinnati
M.S. in Analytics, Northwestern University

Advisor: Douglas Bowman, Ph.D.

A Dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Marketing
2019

Acknowledgement

This dissertation is the beginning of my journey in an academic career, not the end of the process obtaining my Ph.D. degree. I have not been going through this journey in a vacuum. In this accomplishment, I am forever indebted to many people, including my well-wishers, my family, my friends, and colleagues.

At this moment of accomplishment, first and foremost I would like to thank my academic advisor, Douglas Bowman, for his guidance and encouragement, not only during the course of my Ph.D. study but throughout my graduate experience here at Emory University. I still recall the first day when I visited his office as a research assistant. He welcomed me with a smile and asked me in which areas I would be interested in doing research, instead of forcing me to do the research that he wanted me to do. Throughout this process he has put my research interests first, giving me the freedom to pursue various projects without objection. He has also engaged with me in insightful discussions about the research. This work would not have been possible without his critical insight and expertise in marketing and marketing models, which always helps to strengthen and refine my work. His unflinching courage and conviction are not limited to academic work. He understands my challenging situation as a doctoral student and as a mom and knows what my needs will be in advance. I look forward to continuing this academic journey with him and his noble thoughts.

I will forever be thankful to my dissertation committee, Professors Diego Klabjan, Daniel McCarthy, and Zhongjian Lin. Professor Klabjan was my former advisor at Northwestern University and has been guiding me ever since. He has been helpful in providing me with advice many times during my graduate school career. His ambidextrous expertise in machine learning and big data always greatly inspires me and advances my research. I would also like to thank Professor Daniel McCarthy for his scientific advice and knowledge, and for many insightful discussions and suggestions. He also provided endless help and support during my graduate career and job search process. My gratitude also goes to Professor Zhongjian Lin for a great econometrics course, valuable advice, and constructive criticism of my work.

Dozens of people in the Marketing Department, as well as at Goizueta Business School, have helped me and supported me. Professor Sandy Jap offered a wonderful seminar for Marketing Ph.D. students in my first semester, starting me off in my Ph.D. study with widened eyes. Her seminar incorporating a variety of important subjects both broadened and deepened my knowledge of marketing. She gave valuable feedback and suggestions on my research through one-on-one discussions and shared her lifetime experience as a mom and a professor. Professor Ryan Hamilton not only provided me with a rigorous foundation in behavior theories through his Ph.D. seminar but also gave me a nice and breezy smile whenever I met him in the fifth floor aisle. I am also grateful to Professors Chae and Andrews in the Marketing Department, in particular for their comments and encouragement surrounding my preparation for the job market. I also would like to thank my many student colleagues for providing a stimulating and fun-filled environment. Suh-Yeon, who joined the Ph.D. program the same year that I did, has been going through the challenging process with me since bootcamp. Her beautiful eyes and witty jokes are always refreshing and cheerful. Bright Hong, the accounting Ph.D. student who started this program with me, is like a sister to me. We have had honest conversations not only about Ph.D. life but about personal life as well. I believe that she will remain a lifelong friend, as we have been going through both hardship and delight together. I will miss having lunch and coffee with Da Young and Woo Young, Ph.D. students in Marketing. I also owe much gratitude to Mrs. Sharon Bowman, the wife of my academic advisor. She has been so nice and warm-hearted to me whenever I met her. I was very pleasantly surprised by her presents to my son, Joseph.

I also must thank Allison Gilmore. A good support system is critical in surviving and staying sane during graduate school. Her support of Ph.D. students is unparalleled. I will miss not only her spectacular administrative help and guidance but also her amusing jokes and beautiful smiles.

Finally, and above all, I cannot begin to express my bottomless gratitude and love for my husband, Dr. Daewook Kim, and my beautiful son, Joseph. My beloved husband has given me endless support and constant encouragement throughout the Ph.D. process—and even before this process, during my master's work at Northwestern University. We were in graduate school at the same time but in different places across the US throughout our master and Ph.D. programs. Although the days were hard, and there were lots of tears from me, my husband, and especially from my son, Joseph, we trusted each other and committed to each other in the hope of better days ahead. I know that my son, Joseph, who is the best thing to happen to me, is definitely the one who shed the most tears whenever his dad left when he lived with me and when I left when he lived with his dad. I am very proud of his mature soul and endless curiosity. I would not have completed this Ph.D. journey without my sweet cookie's loving words, "I love you, mom. You are the best." I would like to recognize my parents and in-laws for their unwavering support and prayers throughout my graduate study. Combining my husband's Ph.D. with my master's and Ph.D., we have been in graduate school for 10 years. Throughout the long journey, my parents and in-laws stood by us and trusted us. I am sending my love to them all.

Eunhee (Emily) Ko
*Emory University*
April 2019

Table of Contents

Table of Tables

Table of Figures

# Chapter 1

## Overview

### 1.1. Introduction

As consumers have a completely different purchasing journey in today's digital era than they used to, the function of marketing has evolved rapidly. The greatest challenges come from drastic changes in consumer behaviors and proper usage of new techniques to better understand the new consumer purchasing journeys and improve management practices. Consumers' voices have great power nowadays; consumer-generated stories and views are everywhere, such as in e-commerce or on social media platforms. Consumers—especially millennials, who have the most purchasing power—are greatly affected by peers' opinions and make purchase decisions based on user-generated content (UGC) (Bazaar Voice 2012). In recent years, consumers have generated not only critical ratings of products but also various types of content (e.g., images, videos) to express their opinions. In many cases, this emerging content is large-scale and multi-dimensional in nature. Firms competitively adopt advanced technologies to incorporate diverse forms of UGC into their marketing strategies. Artificial intelligence (AI), as opposed to the natural intelligence displayed by humans or animals, and machine learning (ML), a subset of AI, are essential techniques that firms are leveraging as machines become increasingly capable.

Marketing researchers have been also mining the UGC for more than a decade and adopting AI technology more recently to make the UGC more scalable. While a wide range of papers (Goh et al. 2013, Nielsen 2015) reveal that consumer-generated content is more engaging and trustworthy than brand-generated content, we do not know much about content characteristics of UGC and how the content features are linked to market outcomes or consumer behaviors.

Under this context, where there is no clear consensus regarding what types of content drive more effective results or what is more lucrative digital properties, we probably should delay such marketing actions as how to determine user-generated content mix or what is a successful UGC activation plan. In these two essays, I showcase the content features of UGC and link it to consumer reaction or market outcome. In doing so, I employ traditional machine learning techniques and recent progress of AI that can bring a level of efficiency to my work where multidimensional big data is involved. The next three subsections of Chapter 1 are an overview user-generated content, AI in marketing and agenda of this dissertation.

## 1.2. User-Generated Content

Although there is no consensus on the definition of UGC, most people agree that it is any form of content (e.g., text, audio, video, images) that is created by general users on diverse online platforms or websites, including e-commerce websites, social media, and online communities. UGC has become an increasingly important source of data for marketers due to its significant impact on consumers' decisions. Thus, both the marketing industry and academia have paid significant attention to UGC, concentrating on its impact on consumer behaviors and how to embrace it in marketing strategies. Accordingly, UGC is rapidly becoming an important cornerstone of integrated marketing strategies. Due to the rising importance of UGC, firms and marketers have been focusing on new customer metrics, such as likes, engagement, comments, or impressions with which consumers express their opinions about content on online platforms.

Although marketing researchers have extensively examined this new type of input from the consumer side (e.g., customer reviews, social media posts) and how it impacts consumers or market outcomes, there is still limited research on intermediary outcome variables, such as consumer engagement. In addition, although numerous papers have investigated UGC (Agnieszka 2018, Goh et al. 2013, John et al. 2017, Lee et al. 2018, Malthouse et al. 2013) and found interesting associations between content features and consumer metrics, the results regarding whether and how other types of content (e.g., visual content) affect consumer behaviors have been largely inconclusive. Furthermore, existing marketing research on UGC commonly focuses on outcomes instead of the motivations for consumer activities (Toubia and Stephen 2013). Also, in spite of the growing amount of literature on the features of UGC, so far researchers in marketing have only focused on numeric or textual forms of UGC. Throughout this dissertation, the focus is on why consumers pay more attention to certain types of UGC, including not only numeric- and text-based UGC but also image-based UGC.

### 1.3. Artificial Intelligence in Marketing

AI and ML are powerful techniques that have been increasingly employed by the marketing industry in recent years due to their ability to improve forecasting models and assist in management decision-making. With the advent of the Internet and diversity of inputs from various sources, marketers have adopted methods such as AI and ML, a subset of AI, to more accurately predict consumer behaviors and build a successful branding strategy, create engaging content that amuses their customers, or optimize operations and the supply chain.

AI has been widely applied in many areas of marketing, but there are several specific areas in which AI and ML can be leveraged to improve marketing or sales performance. First, AI can increase forecasting or predictive maintenance power. Marketing or sales managers face the challenges of accurately predicting their teams' revenues or profits in each quarter as well as very large amounts of multi-dimensional data from various sources. Deep learning can be used to handle a large amount of multi-dimensional data (e.g., audio, images) from numerous areas, including those that are rather new in the marketing or sales fields, it can enhance the power of predicting revenues or reducing operating costs, and it can help firms better manage their inventory and resources. Second, AI can be employed for customer relationship management or personalized selling. These days, consumers are exposed to and purchase products through various devices and platforms. AI's capability to combine various types of data, from demographics to past transactions with social media, helps firms to build individualized product recommendation systems (e.g., Netflix's recommendation system) or develop better sales strategies, such as up-selling or cross-selling. Finally, AI-driven optimization can solve diverse marketing or logistic problems. For example, AI can produce optimized prices or discount rates to ensure that customers are mostly likely to make a purchase decision or find ideal logistical solutions by leveraging various features, such as delivery traffic, drivers' behaviors, fuel consumption, or maintenance costs.

Throughout the dissertation, both AI and ML techniques are employed to solve various marketing problems. In the first essay, a neural network is applied to determine whether the content features of customer opinions improve the predictive power to distinguish between suspicious reviews and genuine reviews. In the second essay, various deep learning techniques, such as a computer vision application programming interface (API) based on an AI algorithm, a deep learning tool that extracts sentiment features from images, and a neural network algorithm named Word2vec, are used to process text-based data and identify interesting and relevant features in the various types of data (e.g., images, texts).

## 1.4. Agenda of the Dissertation

This dissertation investigates why certain consumer activities occur and the motivations behind the behaviors with which they are associated, employing various types of UGC, including numeric, textual, and visual data. To properly process the unstructured data that are incorporated into the models as focal features, various AI and ML techniques are used, including a computer vision API, deep learning, and machine learning algorithms. Specifically, the first essay examines consumer engagement with brand-themed UGC—image-based social media posts tagged #*brandname*—and asks which content characteristics drive more consumer engagement. To answer this question, the study identifies four focal variables (i.e., visual sentiment, visual complexity, text sentiment, and text complexity) in social media posts including images and text and empirically examines how these four visual and text features drive more consumer engagement, employing various AI and ML techniques to process and manipulate the unstructured data. The second essay delineates the underlying mechanisms in manipulative reviews on an e-commerce website as well as the motivations of such opportunistic behaviors. Specifically, this study considers the relationship between a brand's advertising efforts and reviews that a consumer suspects are fake. Although numerous papers focus on the impact of UGC on market outcomes or consumer spending decisions, as consumer-generated online reviews are a major source of information that can be used in the decision-making process, limited attention has been paid to manipulative reviews and their link to economic motivations. After studying over 270,000 customer reviews on Amazon, the paper finds an interesting relationship between a brand's advertising expenditure, the incidence of manipulative reviews, and the semantic features that characterize these manipulative reviews.

# Chapter 2

## Content Engineering of Images: The Effect of Sentiment and Complexity on Consumer Engagement with Brand-Themed User-Generated Content

### 2.1. Introduction

Image-based posts have become dominant content on social media, and many users (people who post content) generate considerable "buzz" around them. Increasingly, such posts are brand-themed being tagged with #*brandname*. Brand-themed posts on social media are an emerging way that consumers express their relationships with brands.

The importance of image-based posts in social media can be seen through the rise of Instagram, the dominant image-centric social media platform. eMarketer.com (2017) reports that global market penetration of Instagram jumped from 18.8% in 2016 to 24.0% as of December 2017, and is expected to increase to 30.1% by 2021. A survey by Bloglovin (2016) found that approximately 60% of micro-influencers (defined as mid-size social media users with a large following of daily engaged users) consider Instagram to be the most effective platform for generating user engagement. Marketers citing the disruptive power of advertising's digital revolution often point to Instagram. Jeffery Dachis, a Razorfish co-founder, asks, "*What am I doing filming cars driving through the desert when brands are being built on Instagram*?" (Wall Street Journal, 2018). In spite of the growing dominance of visual content in social media and firms' increasing interest in its power to engage consumers, visual content remains thus far poorly understood in marketing.

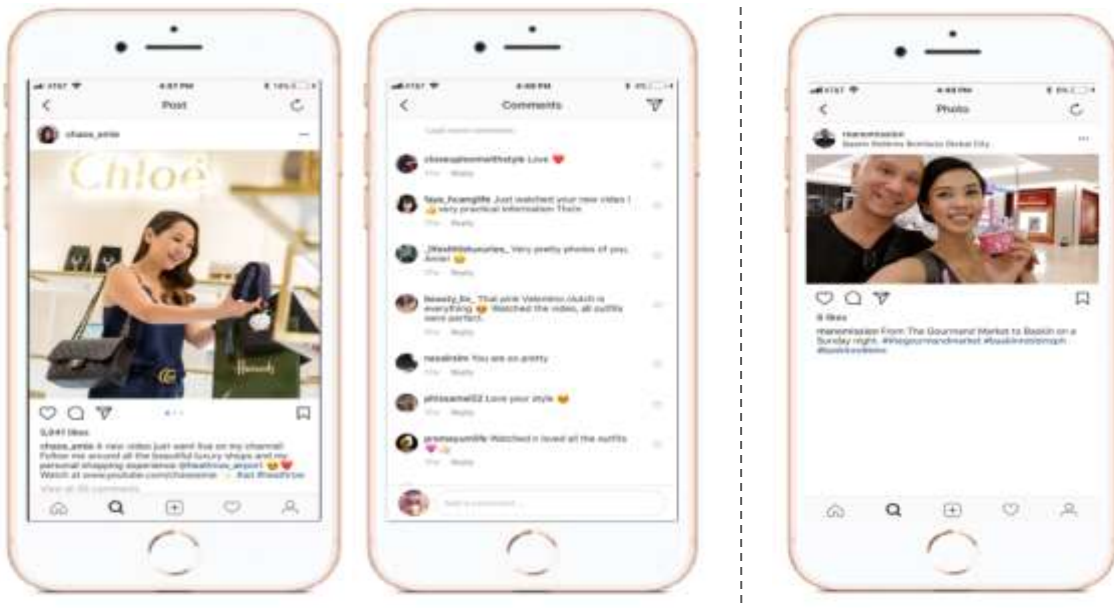
In this paper, we study consumer engagement, measured as the number of likes and comments tied to a brand-themed social media post on Instagram. Higher consumer engagement with a brand-related social media post is associated with more visits to the brand's website (Socialbakers, 2014), which can help drive sales. Earned social media engagement volume affects brand awareness and purchase intent (Colicev et al. 2018), and it also elicits a positive causal effect on offline customer behavior (Mochon et al. 2017).

Users post image content as well as text content which can include hashtags. Other Instagram users who view the post may respond by clicking "like" (a heart shape under the post) or leaving comments. Figure 1 presents two typical brand-themed posts (viewed on a mobile interface). The post on the left by user (A) includes the hashtag *#chanel* (for *Chanel*) and the post on the right by user (B) includes *#baskinrobbins* (for *Baskin Robbins*) in its text content. We ask what visual content (e.g., post (A): *a smiling girl holding a bag; positive mood in the image*) affects consumer engagement. We investigate whether higher consumer engagement with the post by user (A) compared with the post by user (B) is associated systematically with differences in their visual and text characteristics, and the characteristics of the hashtagged brands, while accounting for the user's social media network size and activity.

We posit that perceptual characteristics of images and text affect consumer engagement. More specifically, we propose that semantic aspects of an image-based post in social media – emotional, figural, and contextual – are associated with consumer engagement. We focus on the content of a post and its characteristics rather than on the structure of the networks through which the information is moving.

Our unit of analysis is a brand-themed post on Instagram. The raw data include images that Instagram users upload, text that is associated with the images, and user-related variables such as the number of followers. To ensure that the posts are brand-related, we use brand names (e.g., *#bmw, #pepsi, #nike*) as source tags when extracting the posts. For each brand we append several brand characteristics (e.g., brand knowledge, brand involvement) taken from a dataset posted by Lovett, Peres, and Shachar (2014). We study over 72,000 Instagram posts that are collectively associated with 86 different product brands.

Figure 1. Example Posts from the Instagram Data



(a) A post by a user (A) (b) A post by a user (B)

We describe a post by four focal aspects: visual sentiment, visual complexity, text sentiment and text complexity. Since we analyze large-scale data, we use automated methods that employ various techniques including Deep Convolutional Neural Networks (Deep CNNs), Natural Language Processing (NLP), and hierarchical clustering. After crafting the variables, we model consumer engagement, using characteristics of a post's image and text, as well as characteristics of the focal brand, while accounting for the extent of the user's social media network and activity.

The analyses reveal several interesting findings. First, we find that consumer engagement is higher with a brand-related post when the post contains visually positive images and when its text content is emotionally divergent. Second, we find a *S*-shaped pattern between consumer engagement and both visual complexity and text complexity. Visual complexity affects consumer engagement positively at low and moderate levels and negatively at very high levels. There is an optimal point that drives the most consumer engagement, while images that are too visually cluttered attenuate consumer engagement. For text complexity, we find a similar effect: excessive text information attenuates consumer engagement. Finally, we find that several brand characteristics affect consumer engagement. Brand visibility and brand involvement are positively associated with higher levels of consumer engagement.

The article is organized as follows: We first review prior studies in the areas of content marketing, consumer engagement, and machine learning applications in marketing that are relevant to our research. Next, we present our theoretical model that describes how characteristics of the post, the focal brand, and the user together affect consumer engagement with a brand-themed post. We then describe how we extracted the raw data from Instagram and created the variables that are used in our models. We then test our theory using a negative binomial model and zero-inflated negative binomial model for the number of likes and the number of comments, respectively, our dependent variables. We conclude with a discussion of the empirical results, managerial insights gained, and suggestions for future research.

## 2.2. Background

A user-generated brand-themed post can result from a user's intrinsic desire to express her or his relationship with the brand, possibly motivated by self-presentation purpose often underlying social media posting behavior (Jensen Schau and Gilly 2003). A brand-themed user-generated content can also be motivated by actions taken by marketers. Marketers encourage brand-themed user-generated content (UGC) in a number of ways including experiential marketing campaigns, by collaborating with general users in social media, and by conducting influencer campaigns, realizing various benefits such as content authenticity and cost-effective efficiency. A successful example of the latter is a Hewlett Packard Australia product launch campaign that leveraged Australian fashion influencers (Scrunch 2016). The influencers were given an opportunity to use an HP Spectre for a week, then share their trial experiences with the new laptop on Instagram. With only 20 collaborating influencers, the brand recorded 62,943 direct engagements with the campaign's content and reached 941,300 consumers. A well-known example of collaboration with users is the ALS Association's Ice Bucket Challenge, which promoted awareness of the disease amyotrophic lateral sclerosis (ALS) and generated over $100 million in exposure in 2014 through the viral nature of social media.

As UGC including online reviews and social media posts has emerged as an important source of interactions between consumers, a large body of literature has investigated the relationship between the UGC components (e.g., content, volume) and relevant outcomes (e.g., product sales, virality, engagement) (Godes and Mayzlin 2004; Chevalier and Mayzlin 2006; Liu 2006; Archak et al. 2011; Sonnier et al. 2011; Lee et al. 2018). While numeric (e.g., number of hashtags, review ratings) and textual information has been studied, visual characteristics have not been considered.

Images are an integral component of social media posts. The power of visual depiction is well-established by the social psychology and consumer behavior literatures (Goldberg 1999; Pieters and Warloop 1999; Pieters and Wedel 2007; Rayner et al. 2008; Townsend and Kahn 2014). Much of this work was eye movement (lab) studies in (firm-generated) advertising contexts. Advertising features (e.g., brand names, portion of text or picture) or contextual cues (e.g., purchasing goal) were found to affect viewers' fixation durations (Rayner et al. 2001 and 2008; Wedel and Pieters 2000; Radach et al. 2003; Li et al. 2016). Different from this literature, we examine visual content from large-scale field data.

Our study is relevant to three streams of research in the marketing and data analytics fields: consumer engagement in social media, content marketing with visual and text information, and machine learning applications in content-based image retrieval.

### 2.2.1. Consumer Engagement in Social Media

Consumer engagement in social media is a core metric for firms monitoring the breadth and engagement of their customers. Colicev et al. (2018) provide a recent and comprehensive summary of studies of brand social media engagement, which we won't repeat here.

Prior research has investigated both the antecedents and outcomes of engagement on social media (Goh et al. 2013; Malthouse et al. 2013; John et al. 2017; Lee et al. 2018). For example, Goh et al. (2013) connect user-marketer interactions with content data in a brand community to transaction data, uncovering a stronger effect of UGC than marketer-generated content (MGC) on consumer purchase behavior. Lee et al. (2018) focus on the textual part of advertising content in social media to explore what text features affect customer engagement using large-scale field data-set on Facebook. They find that content related to brand personality (e.g., emotion, humor) increases users' engagement whereas informative content (e.g., prices) is related to lower-levels of users' engagement. Agnieszka (2018) reveals that certain brand-generated message strategies in social media (e.g., emotional appeal, informative content) are positively associated with engagement or viral behaviors. A common aspect of all these studies is that, although they find interesting relationships between content factors and consumer engagement, their focus is limited to text. Different from these studies, we study consumer engagement more comprehensively, accounting for both visual and text content.

The extant literature has associated a consumer engagement with various consumer behavioral outcomes or market results. Brodie et al. (2013) empirically reveal that engaged consumers demonstrate enhanced consumer loyalty, satisfaction, empowerment, connection, emotional bonding, trust and commitment in the context of brand communities. Cheung et al. (2011) address customer relationship management through consumer engagement in social media platform has become an increasingly important component of marketing or brand strategies. Several papers directly link consumer engagement to economic performance. Oh et al. (2015) examines how consumer engagement around social media platforms such as Facebook and YouTube is associated with box-office gross revenue and find empirical support for the positive relationship between consumer engagement and the economic performance. Coursaris et al. (2016) show the positive effect of engaging brand content on purchase intention and consumer brand equity. Cheung et al. (2015) reveal the empirical association of consumer engagement with their spending on online games in online game industry.

### 2.2.2. Content Marketing

We use the term content marketing to describe the strategies behind MGC on the Internet. While many MGC studies are laboratory settings using small samples (Percy and Rossitier 1983; Pieters et al. 2002; Peracchio and Meyers-Levy 2005; Pieters et al. 2007; Rayner et al. 2008; Deng and Poole 2010; Xiao and Ding 2014; Kumar et al. 2016), more recent studies have employed large-scale field data to investigate the effects of advertising content (Albuquerque et al. 2012; Sun and Zhu 2013; Lee et al. 2018). Broadly, these works focus on two types of content in advertisements: visual or text information. For example, Xiao and Ding (2014) focus exclusively on the effect of faces (visual stimuli) on a viewer's reaction to an advertisement, while Pieters et al. (2007) investigate five key elements of advertisements (brand, text, pictorial, price and promotion), suggesting alternative design approaches to increase consumers' attention paid to the ad. Goldberg et al. (1999) focus on the design aspect of visual stimuli studying the ability of nutrition labels on food to support fast and accurate visual searches for nutrition information. Certain types of label manipulations (e.g., center vs top or bottom of the label, thinner alignment lines vs. thicker anchoring lines) strengthen a consumer's ability to search for targeted information.

More recently, researchers have investigated the effect of content components in advertisements with large-scale datasets. Teixeira et al. (2014) examine the downside of too much positive entertainment (e.g., visual imagery, upbeat music) in TV advertisements through a large-scale study, revealing that the level of entertainment elicits a non-monotone (U-shaped) effect on purchase intent. Ryan et al. (2017) connect visual tokens with descriptions of firms in textual form to study how much a logo can explain a brand's personality, as well as key visual components from logos that elicit brand and firm relevant associations. Lee et al. (2018) study textual forms of advertising content in social media finding that brand-personality related components are associated with higher consumer engagement levels, whereas informative content is related to lower levels of engagement.

Although several papers contribute to growing extant literature that investigates content features with large-scale observational data sets, the focus so far has been limited to textual forms of content. Further, most of the work is in advertisement settings, dealing with MGC rather than UGC although consumers increasingly play a pivotal role in creating and sharing brand stories using such emerging dynamic networks as social media (Gensler et al. 2013). Liu et al. (2017) study content beyond text (images) using large-scale field data, but the study focuses on measurement development, which does not connect visual features with outcome variables. We attempt to fill this gap in extant literature by examining visual features, as well as text, using large-scale observational data, linking them to consumer engagement.

### 2.2.3. Machine Learning and Social Media

Machine learning has been popularly used to retrieve information from images as huge quantities of images have become available in various online media platforms, accompanied by explosive growth in research involving large-scale social multimedia analysis. Numerous studies, mostly in the computer science field, have used machine learning to understand, index, and annotate images to represent a wide range of concepts (Wan et al. 2014; Cappallo et al. 2015; Kalayeh et al. 2015; Hu et al. 2018). Deep convolutional neural networks (CNNs) have been adopted increasingly by scholars for their improved performance in classifying large-scale web datasets (Chen et al. 2014; Chen et al. 2015; Xu et al. 2014; Gelli et al. 2015). We draw on a deep learning method developed in the computer vision field to process the images in our study. Frequently, classifying visual images using a deep learning approach requires a very large number of images for training data. To overcome this challenge, many researchers use models that are pre-trained on a large dataset in a similar domain. We follow this approach, using a visual sentiment classifier based on Deep CNNs, *DeepSentiBank* (Chen et al. 2014).

## 2.3. Conceptual Framework

We conceptualize consumer engagement with a brand-themed user-generated post on social media as a function of three broad drivers: characteristics of the post (*content*), characteristics of the brand (*motives*), and characteristics of the user (*network structure*). Our framework is consistent with Peters et al. (2013), who conceptualize social media content as having three distinct aspects: (1) content quality; (2) content valence; and, (3) content volume. Our conceptual framework is presented in Figure 2.

### 2.3.1. Characteristics of the Post

Posts are described through image content - visual sentiment, visual complexity, and indicators of the types of objects contained in the image – and text content - text sentiment, text complexity, and the length of the text.

*Visual Sentiment.* Visual content contains cues about affect, emotion, and sentiment or valence. For example, eyes can be described as “beautiful”, “glaring”, or “sad”. A dog could be “happy” or “angry”. These are individual components of an image, which collectively determine the image’s overall sentiment ranging from strongly negative to strongly positive. We expect that the more positive the overall image is, the higher the consumer engagement with a brand-themed post.

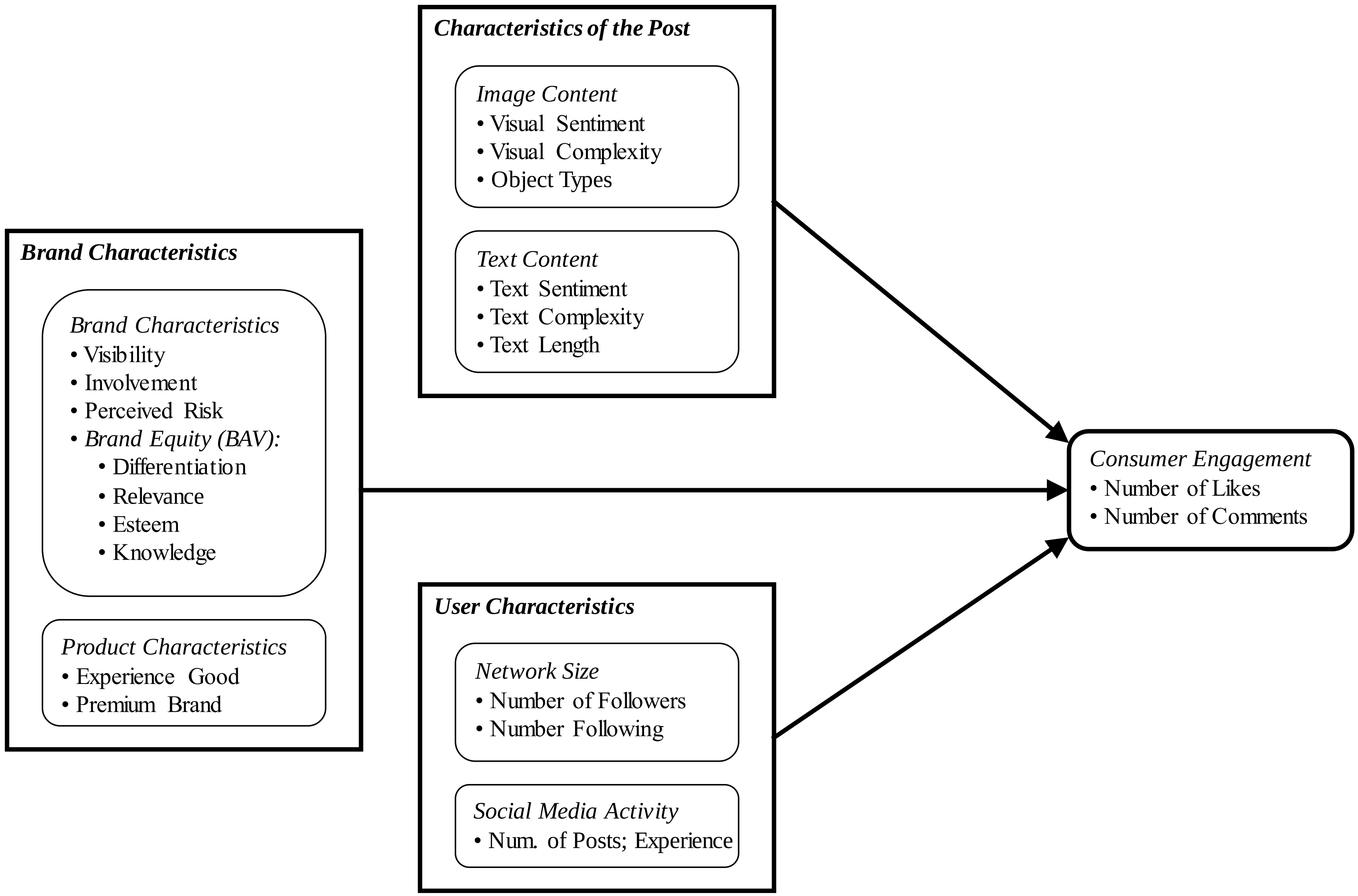
*Visual Complexity.* Visual content can be viewed as a collection of objects, both living (e.g., a girl, a cat) and/or non-living (e.g., a tree, cake). The photography industry advocates two dominant approaches for constructing an engaging photo. The first, called the isolation effect, is to take a simple photo in which the focal object stands out from everything around it. The second follows from the belief that an engaging photo contains a degree of richness and complexity that draw attention to it. Eventually, however, clutter takes hold making an image hard to process. Indeed, many social media pundits advise users to edit out unnecessary objects in a picture before posting it.

*Object Types.* We succinctly account for the various object types included in the visual content of a post. Pets (or living things in general) have long been known to generate heavy engagement volume. Furthermore, posts that include people typically involve those people and their networks, which are predisposed to engaging with the post (Saeideh et al. 2014). Thus we account for the number of people (faces) in the image.

*Text Sentiment.* Text content is commonly described by its sentiment (positive, neutral, negative). While early researchers used binary coding, measuring the degree of sentiment or its emotional divergence is now more common, capturing the total amount of valence from text. Emotionally divergent text may be associated with cognitive attention or arousal-related effects. Extant research suggests a positive relationship between the divergence aspects of text sentiment and users' viral behavior in social media or blogs (Pfitzner et al. 2012; Stieglitz and Xuan 2013). We expect that more emotionally divergent text is associated with higher consumer engagement.

*Text Complexity.* Text (or topic) complexity can be approximated by its topical diversity and topical variety. On one hand, relevant text that is clear, concise, and to the point should lead to higher consumer engagement. On the other hand, a moderate amount of complexity may also be associated with higher consumer engagement since it may better capture the richness of a post and contain elements that appeal to a wider audience. However, too much text can lead to information overload, likely indicating a lack of focus or purpose, as well as expectations for attenuated engagement with a brand-themed post.

Figure 2. Conceptual Framework



2.3.2. Brand Characteristics

We describe the hashtagged brand by its visibility, involvement, perceived risk, and brand equity, aspects that are broadly related to consumer engagement with a brand.

*Visibility.* Brands vary in their perceived visibility in the marketplace, and the extent to which they can be communicated to others. We expect a synergy between the public nature of social media and the inherent visibility of the brand. All other things being equal, a brand-themed post in which the focal brand is perceived as more visible or observable to consumers should elicit higher consumer engagement.

*Involvement.* Involvement refers to the degree of importance that consumers attach to a brand. It captures how involved a consumer is towards a brand personally, socially, and economically. All other things being equal, brand-themed posts for higher involvement brands should elicit higher consumer engagement.

*Perceived Risk.* Products (and brands) vary in the functional, financial, and emotional uncertainty associated with them (Rogers 1995). Generally, perceived risk is associated with more cautious behaviors. We expect that this extends to consumers' behaviors on social media. All other things being equal, consumers will be more cautious about engaging with a brand-themed post about brands with higher perceived risk.

*Brand Equity.* Brand equity refers to consumers' reactions to a product compared with a version of the same product that is unnamed (or fictitiously named). A brand's equity lies in the knowledge and associations that consumers tie to the brand. Young & Rubicam's (Y&R) Brand Asset Valuator (BAV) is a commonly applied four-dimensional measurement framework for brand equity: Differentiation and Relevance are related to a brand's growth potential (or Brand Vitality); Esteem and Knowledge create the power of the brand (or Brand Stature). Two dimensions are especially relevant in our context of consumer engagement with brand-themed posts as they relate to the self-presentation purpose often found in social media engagement (Jensen Schau and Gilly 2003). Differentiation refers to the defining characteristics of the brand and its distinctiveness relative to competitors. Knowledge captures consumers' awareness of the brand and an understanding of what it represents.

*Experience Good; Premium Brand.* Finally, we account for two product characteristics: whether the product is an experience good, and whether it is a premium brand. Experience goods rely on reputation, customer loyalty, and word of mouth, as these aspects are often surrogates used to make purchasing decisions. Premium brands are typically priced higher relative to the price of other brands in the category.

### 2.3.3. User Characteristics

*Network Size.* Industry pundits regularly stress two paths to earning more likes on social media: building a large network of followers, and posting regularly. Users with larger networks on social media platforms have a larger number of followers with an *a priori* affinity for engaging with their content, and thus, should elicit higher consumer engagement through their posts.

*Network Activity.* On one hand, users with more experience posting on social media (a substantial history of posting activity) have had the opportunity to learn which of their posts engage consumers and which do not. On the other hand, it is easy to imagine some users posting so excessively that the majority of their posts lack relevance and they alienate other users.

## 2.4. Data

The data were collected from Instagram.com, one of the fastest growing social media platforms in the world, during the last three months of 2017. Unlike Twitter, which entails mostly interacting with text, via *tweets*, Instagram users mostly post photos (or videos) and share them with friends or the entire Instagram community.

Our unit of analysis is a post. Hashtags with brand names (e.g., #*bmw*, #*pepsi*) are the source tags. The raw data are post-related factors, including images[1], hashtags, mentions, comments, and the number of likes, while user-related variables include number of followers, number of accounts the user is following, number of Instagram posts the user has made to date, and user names and descriptions. Figure 3 presents an example of an Instagram feed page (the source for the post-related factors), and an example of an Instagram user page (the source for the user-related factors). Variables displayed in the Instagram interface such as post-timestamps, post IDs, and description languages were extracted via the Instagram API. We filtered out posts with description languages other than English.

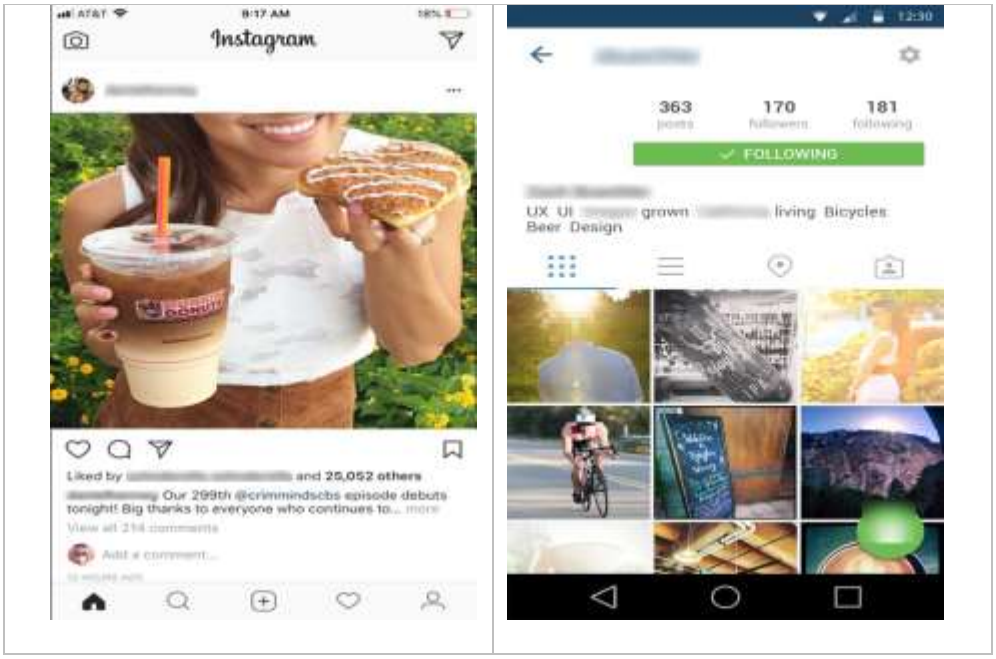
### 2.4.1. Raw Data and Sample Selection Criteria

Many consumers now include a brand name hashtag (e.g., #*bmw*, #*cocacola*) as part of their posts on Instagram.com. While hashtag usage is determined by the posting consumer, most users follow received social media etiquette and include these only when they deem them relevant. For example, posts tagged #*bmw* might include a photo of a just washed BMW automobile, a person sitting in the driver's seat of a BMW, or a photo of a BMW concept car taken at an auto show, to name a few.

We began by creating a list of brands to study. Lovett et al. (2014) describe a dataset that they published containing 136 different measures of brand characteristics for almost 700 of the leading U.S. national brands, measured in 2010. Beginning with the list of product brands (i.e., filtered based on the variable *Product*), we filtered out words that correspond to movie names (e.g., "Indiana Jones", "Shrek") and brands with common-noun names (e.g., "Brother", "Degree"). To ensure variation on perceived brand image, we focused on three distinct brand-image dimensions – Fun, Glamorous, and

[1] Instagram allows user to upload up to 10 images or videos. For this study, we extract only one image per post and exclude video posts.

Figure 3. Example Feed Page (left) and User Page (right) from the Instagram Data
(from the mobile interface)




Rugged – extracting the top 30 brands on each dimension. Each brand-image measure is the percentage of respondents in a Y&R survey, which gauged this attribute with respect to the brand. Considering only brands with at least 10 percent on one, and only one, of the three brand image dimensions we arrived at 86 brands that are the focus of our study (see Table A1 in Appendix 1).

*Instagram Data.* The Instagram data were obtained in two stages. First, we collected links to new posts via the API that Instagram uses to display post previews on its tag listing pages (e.g., #*lego*: https://www.instagram.com/explore/tags/lego/). Then, a separate process queried our collected link database for posts that had at least two days to accumulate comments and likes. We then scraped the post and user data for each link through an automated headless browser.

Searching on Instagram using the keyword #*brandname* returns posts organized under two groupings, popular posts ("Top Posts") and recent posts ("Most Recents"). To ensure our sample was representative we included both types of posts (1.57% of the posts collected were "Top Posts"). We excluded posts that were generated by firms to ensure that all posts in our dataset were created by users only. On Instagram, we manually located the official brand accounts of the brands in our data (86 brands) and found the URLs that are identical to the URLs of the brand accounts. We then removed the posts that were associated with brand accounts from the dataset.

*Brand Characteristics Data.* The brand characteristics data come from the dataset posted by Lovett et al. (2014). Three brand characteristics – Visibility, Involvement, and Perceived Risk – come from their survey administered to 4,769 respondents. Brand equity is measured as the four pillars of Y&R's Brand Asset Valuator (BAV): Differentiation, Relevance, Esteem, and Knowledge. Finally, two product characteristics that Lovett et al. (2014) determined through various secondary data sources were included: whether the brand is a premium or value brand, and whether it is an experience good.

Figure 4 depicts an overview of the raw data extraction process. Table 1 lists the constructs, variables names, descriptions, and their sources.

### 2.4.2. Variable Crafting of User Post Data

We employed various techniques to extract and craft features from both the images and the text associated with the images. These include Deep CNNs, computer vision Application Programming Interfaces (computer vision APIs), and Natural Language Processing (NLP; e.g., part-of-speech tagging). These methods are well established. For example, the visual sentiment classifier that we use (Chen et al. 2014) was first developed with a linear support vector machine (SVM), then improved with Deep CNNs a few years later. Both classifiers have been used frequently in many application papers (Bhattacharya et al. 2013; Yoon and Pavlovic 2014; Gelli et al. 2015; Jou et al. 2015; Kalayeh et al. 2015). In the following sub-sections, we elaborate on the variable crafting processes, as well as introduce the methods and describe how we applied them to our data set. Figure 5 depicts the variable crafting process for posts.

Figure 4. Variable Extraction Process

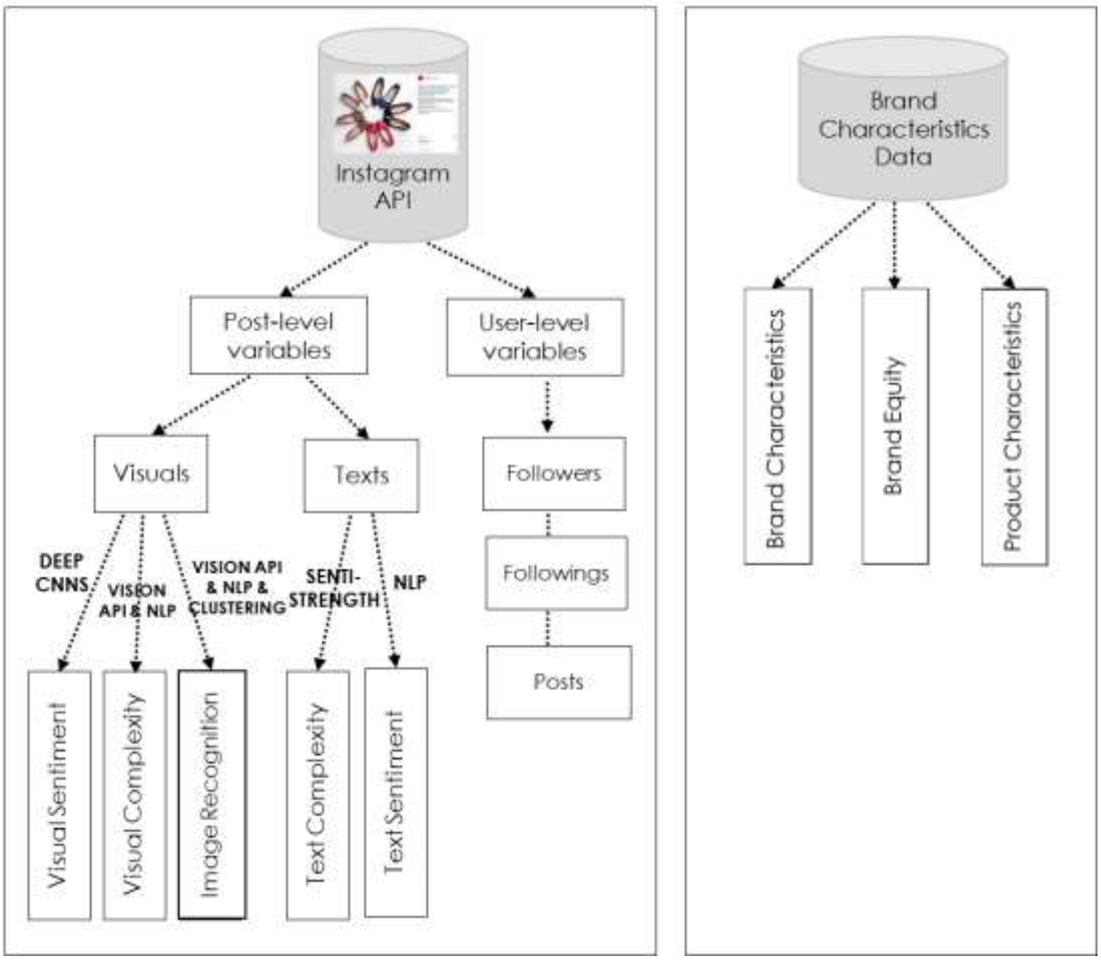
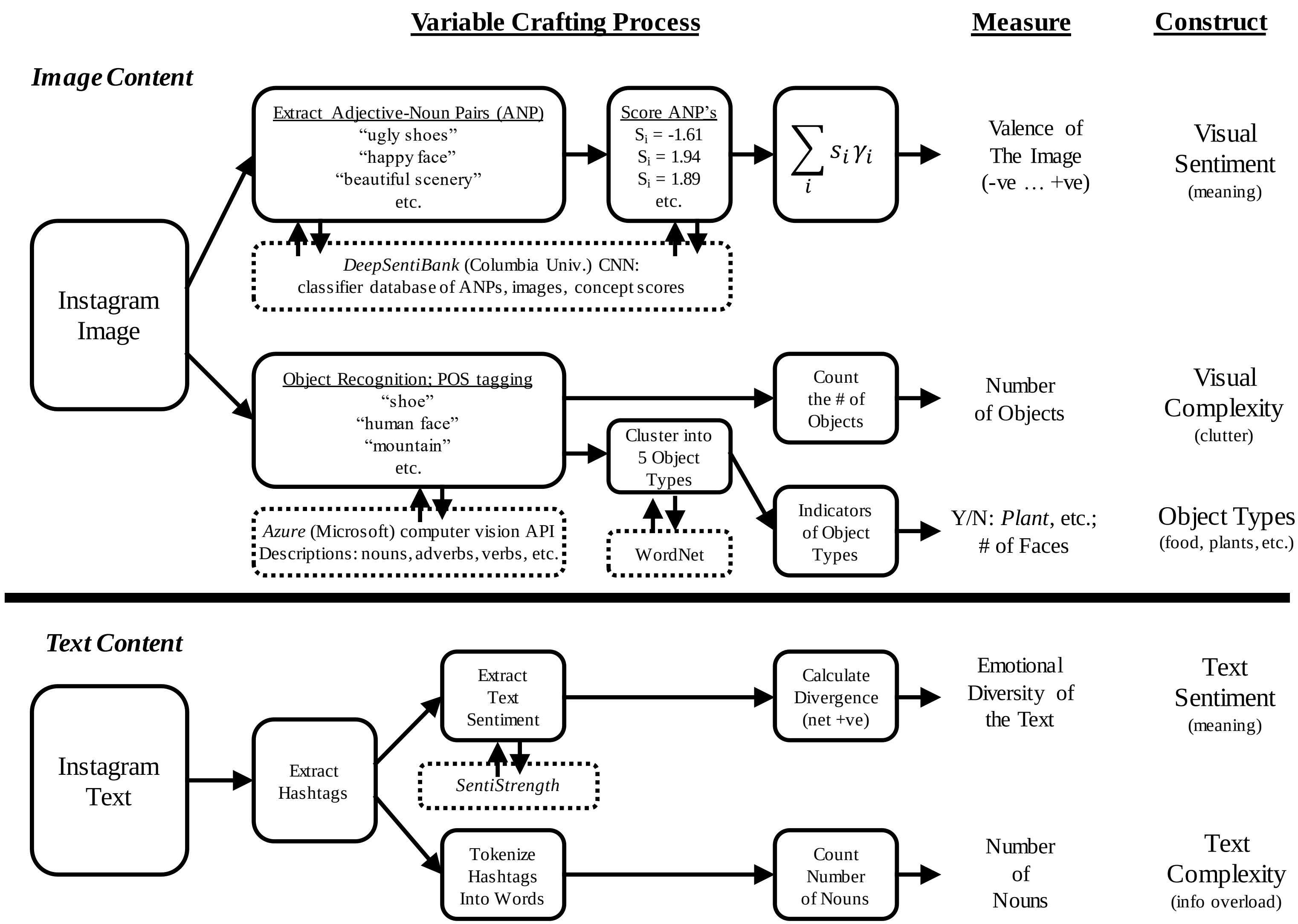


Table 1. Variable Descriptions and Sources

| Construct | Variable(s) | Definition | Source |
|---|---|---|---|
| *Image Content* | | | |
| Visual sentiment | VizSenti | Visual sentiment score of an image | Instagram API + Deep CNNs |
| Visual complexity | VizComplexity | Visual complexity score of an image | Instagram API +Computer Vision API + NLP |
| Object types | Living; Food; Plant; | Indicators of whether the image contains a living thing; a food product; a plant, respectively | Instagram API + Computer Vision API + NLP + clustering |
| Number of faces | NumFaces | Number of human faces | Instagram API + Computer Vision API |
| *Text Content* | | | |
| Text sentiment | TextSenti | Text sentiment score of an image | Instagram API + SentiStrength |
| Text complexity | TextComplexity | Text complexity score of an image | Instagram API + NLP |
| Length of the text | TextLength | Text length in characters | Instagram API + NLP |
| *Brand Characteristics* | | | |
| Visibility | Visibility | Rogers (1995) observability construct: extent to which the product is visible to others | LPS Brand Characteristics Data[1] |
| Involvement | Involvement | Importance of the purchase decision. Scale from Ratchford (1987) | LPS Brand Characteristics Data[1] |
| Perceived risk | PerceivedRisk | Functional, financial, and emotional uncertainty associated with the product. Scale from Ostlund (1974) | LPS Brand Characteristics Data[1] |
| *Brand Equity* | | | |
| Relevance | Relevance | How appropriate is the brand for you personally? | Y&R Brand Asset Valuator[1] |
| Differentiation | Differentiation | Extent to which the brand is perceived as differentiated from other brands. | Y&R Brand Asset Valuator[1] |
| Esteem | Esteem | Extent to which people hold the brand in high esteem. | Y&R Brand Asset Valuator[1] |
| Knowledge | Knowledge | Level of intimate understanding of the brand. | Y&R Brand Asset Valuator[1] |
| *Product Characteristics* | | | |
| Type of good | ExpGood | Experience good = 1 ; Search or Credence good = 0 | LPS Brand Characteristics Data[1] |
| Premium | Premium | Premium brand = 1 ; Middle or Value brand = 0 | LPS Brand Characteristics Data[1] |
| *User Characteristics* | | | |
| Network size: Followers | NumFollowers | Number of Instagram accounts that are following the user | Instagram API |
| Network size: Following | NumFollowing | Number of Instagram accounts that the user is following | Instagram API |
| Experience: Activity | PostCount | Number of Instagram posts that the user has created to date | Instagram API |

[1] From Lovett, Peres, Shachar's (2014) online supplement (dataset).

Figure 5. Variable Crafting Process for Images and Texts



### 2.4.2.1. Deep CNNs for Visual Sentiment: *DeepSentiBank*

Motivated by recent advances in CNNs to acquire a semantic representation of image-based content, we used a CNN as a classifier to detect sentiment from our image data. In particular, we applied a fine-tuned CNN network named *DeepSentiBank* (Chen et al. 2014), which is based on the Visual Sentiment Ontology (VSO)[2] to incorporate sentiment information from image posts. The initial model is *SentiBank* (Borth et al. 2013), whose sentiment detectors were trained with Linear Support Vector Machines (SVMs). *DeepSentiBank*, which is initialized with the weights trained from ImageNet (Deng et al. 2012) and fine-tuned on the *SentiBank* dataset, is the upgraded model from the previous network with detectors trained using CNN architecture (*SentiBank).*

The underlying psychological theory for the model is Plutchik's Wheel of Emotions (Plutchik 1980). The 24 emotions defined in Pluchik's theory were used to create search keywords and extract images from Flickr and YouTube, which, in turn, guided the researchers toward developing a large-scale VSO consisting of more than 3,000 adjective noun pairs (ANPs), such as "ugly shoes" or "happy face." Nearly 1 million images from Flickr were used to train the classifiers of the concepts (ANPs). *Caffe*[3] -- a deep learning framework developed by Jia et al. (2014), Berkeley AI Research, and community contributors -- was employed to train the Deep CNNs model. Their final set of classifiers contains 2,089 ANPs, with 867,919 images. (For technical details regarding the CNN architecture used for training the visual sentiment concept, please refer to Chen et al. [2014]).

This model has proven to be effective in various applications in visual attention (Fan et al. 2017), aesthetic assessment (Bhattacharya et al. 2013, Mohammad et al. 2015), and social media commenting (Chen et al. 2014). Jie et al. (2012) show a prediction of sentiment reflected in visual content. They propose a systematic, data-driven methodology to construct a large-scale sentiment

---

[2] VSO - For more information about Visual Sentiment Ontology, please refer to following website: http://visual-sentiment-ontology.appspot.com/

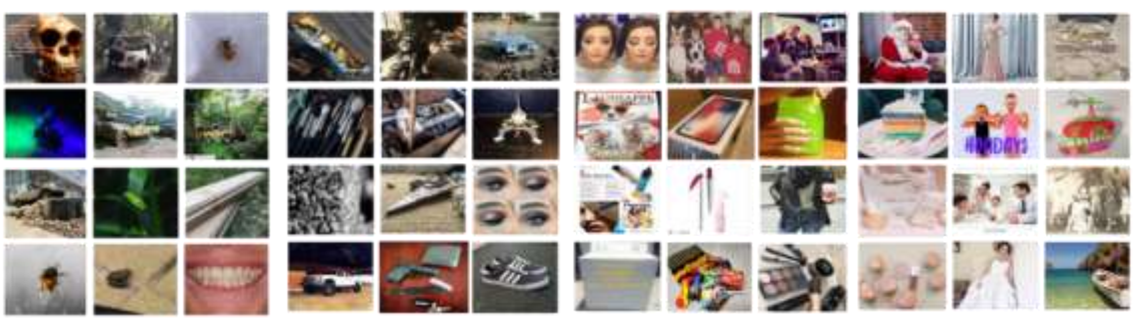
[3] Caffe: For more information regarding Caffe, please visit http://caffe.berkeleyvision.org/.

ontology built upon psychology and web crawled folksonomies using SentiBank. A number of papers employed the DeepSentiBank or the early model, SentiBank, in the context of measuring the impact of visual features. Gelli et al. (2015) adopted DeepSentiBank to extract visual emotions when predicting a popularity score of social images in social media, and Fontanini et al. (2016) applied 2089-D vector of the probabilities of visual sentiments from the DeepSentiBank network as one of the features to represent visual contents. Kalay et al. (2015) study a specific type of social images, selfies in social media, and investigate how appearance of certain objects, concepts, and attributes affect the popularity of the selfie images. In developing the system, they adopted the 2,089 visual concepts detectors from SentiBank and applied the model to their data to generate a 2,089-D vector for each image. When predicting video interestingness and capturing the trend of emotional states for the video interestingness as a temporal feature, they took the concept detectors of SentiBank model that produces 1,200-dimensional vector for each sequence of a video for certain duration. Several scholars exploit the DeepSentBank or SentiBank model to predict aesthetic aspect of images. Bhattacharya et al. (2013), for example, use the visual detector from SentiBank library to generate 1,200 dimensional Adjective Noun Pairs (ANP) when they develop an aesthetic model emphasizing psycho-visual statistics. Al-Naser et al. (2015) investigate which regions of images attract more attention when images are observed by a human participant. Employing Adjective Noun Pairs (ANP) from SentiBank and eye-tracking techniques, the authors develop a system to map the humans' attention on the images and the visual presence of the noun of the ANP. Some authors utilize the networks in social event contexts. Dewan et al. (2016) analyzed popular themes and sentiment on social images and texts using data from Facebook and applied the SentiBank model to identity image sentiment, which they employed as one of the visual features for their model.

*Visual Sentiment Measure.* In our study, we extracted two descriptors from *DeepSentiBank*: 2,089-

ranked concept scores and 4,096-dimension features (fc7[4]). For the computation of visual sentiment, we adopted the 2,089-D feature vector, in which the feature values correspond to the ANPs likelihood in the image. A sentiment measure is assigned to each ANP (e.g., "beautiful scenery") in *SentiBank*, where negative, close to zero, and positive numbers indicate negative, neutral, and positive sentiments, respectively (e.g., beautiful scenery = 1.89). We computed the visual sentiment of an image *I* as $S(I) = \sum_i s_i \gamma_i$, in which $s_i$ is a sentiment measure and $\gamma_i$ is ANPs likelihood, respectively. The final values of visual sentiment are bounded between -2 and +2, based on the design of the initial model (*SentiBank*). Figure 6 displays representative images from our Instagram data organized from low to high visual sentiment values.

Figure 6. Example Images from the Instagram Data by Visual Sentiment



Negative ←----------------------------------------→ Positive

2.4.2.2. Computer Vision API, NLP, and Clustering for Visual Complexity and Object Types

Although there is no consensus regarding the measurement of visual complexity, several approaches have been employed. Pieters et al. (2010) consider both feature complexity measured with a JPEG algorithm (Wallace 1992) and design complexity including quantity of objects, irregularity of

[4] fc7: The weights that are extracted from fully connected layer 7 in deep convolutional neural networks.

objects, dissimilarity of objects, details of objects, asymmetry of object arrangement, and irregularity of object arrangement. Some scholars focus more on certain aspects of visual complexity and employ the quantity of objects or items as a visual complexity measure. For example, Kosslyn (1975) and Palmer (1999) found that design complexity is contingent on the quantity of objects in images, and Isola et al. (2011) revealed that the number of objects in an image affect people's memorability. Luck and Vogel (1997) measured the capacity of working memory for simple vs. complex visual features. They used the number of items in the stimulus array as one dimension to measure it and found that short-term memory is a function of the number of objects, independently of the number of features. Xu and Chun (2006) studied whether neural processing capacity is a function of a fixed number of objects or it varies with increasing visual object complexity. While examining this, the authors adopted the definition from Luck and Vogel (1997) for visual object complexity. When investigating the relationship between visual short-term memory and the number of objects as well as information load, Alvarez and Cavanagh (2004) utilized the number of items as a complexity measure. In online context, Michailidou et al. (2008) used the quantity of each element that is used on the web page as one dimension to measure visual complexity of web page, when studying visual complexity and aesthetic perception of web page. Some prior research has examined the effect of specific objects on responses. For example, Lovato et al. (2013) investigated the relationship between specific objects and visual preference, and Xiao and Ding (2014) studied human faces and their effect on advertising effectiveness. Our approach to operationalizing visual complexity and object types follows those of researchers (e.g. Kossly 1975; Palmer 1999; Isola et al. 2011; Lovato et al. 2013; Xiao and Ding 2014) who take a simpler approach, focusing on the quantity of objects and specific object types. We do this for two reasons. First, our paper investigates social media image posts in a more comprehensive way than has been done to date, dealing not only with visual features but also text features from the posts. Thus we try to keep the measurements simple, but still make them consistent with metrics from the extant literature. In addition, we study images as large-scale data that requires automating their processing. Although there is a wide range of literature that studies visual complexity in various fields such as consumer behaviors, psychology, or computer vision (Berlyne 1974; Cox and Cox 1988; Donderi 2006), their approaches are not usable here, as we try to ensure that visual complexity concepts are comprehensively measured with an automated approach.

To create object-related variables from the images (visual complexity, object types and number of faces), we relied on object recognition techniques (computer vision API) that can identify a specific object in a digital image, NLP, and hierarchical clustering methods. Several models exist for object recognition in the computer vision field (e.g., Tensorflow, OpenCV, Computer Vision System Toolbox by MATLAB). Cloud services firms such as Amazon AWS, Google Cloud, or Azure by Microsoft now offer a computer vision API to analyze image content, e.g., object recognition. We used the computer vision API offered by the Microsoft cloud platform (Azure) for object recognition of our images. When we connected with the API, we specified the visual features desired: *description*, *faces, image type*, and *color*. The *description* is our main data source for coding visual complexity and object types. It is a list of words related to the image content. Initially, the *descriptions* included all types of words (e.g., nouns, adverbs, verbs), but objects or things are generally indicated by nouns. Accordingly, we extracted only nouns from the *descriptions* using part-of-speech (POS) taggers.

Figure 7 shows two examples of the POS tagging results. We used only the words tagged with "NN (noun, singular)" or "NNS (noun, plural)" to capture objects. Figure 8 presents example images from our Instagram data and their extracted nouns.

Figure 7. Examples from the Instagram Data of Part-of-Speech (POS) Tagging

- Example #1: POS tagging for *descriptions* of an image

[('*photo*' 'NNS') ('*cake*' 'VBP')('*chocolate*' 'JJ') ('*food*' 'NN') ('*different*' 'JJ') ('*woman*' 'NN') ('*top*' 'VBP') ('*various*' 'JJ') ('*standing*' 'VBG') ('*table*' 'JJ') ('*bunch*' 'JJ') ('*birthday*' 'NN') ('*filled*' 'VBD') ('*many*' 'JJ') ('*several*' 'JJ') ('*old*' 'JJ') ('*hydrant*' 'JJ') ('*man*' 'NN') ('*holding*' 'VBG') ('*display*' 'NN') ('*red*' 'JJ') ('*group*' 'NN') ('*people*' 'NNS')]

---

- Example #2: POS tagging for *descriptions* of an image

[('*red*' 'JJ') ('*laying*' 'NN') ('*sitting*' 'VBG') ('*grass*' 'NN') ('*bag*' 'NN') ('*lying*' 'VBG') ('*shoes*' 'NNS') ('*wearing*' 'VBG') ('*small*' 'JJ') ('*top*' 'JJ') ('*hat*' 'WP') ('*black*' 'JJ') ('*little*' 'JJ') ('*pair*' 'NN') ('*dog*' 'NN') ('*green*' 'JJ') ('*suitcase*' 'NN') ('*stuffed*' 'VBD') ('*luggage*' 'NN')]

*Visual Complexity Measure.* We measured the quantity of objects by counting the number of nouns and used it to operationalize visual complexity. To assess the face validity of this approach, we created a visual complexity spectrum (Figure 9) that represents the images from the smallest number of objects to the largest number (from left to right). Examining Figure 9, images representing lower visual complexity look simple and contain a few types of objects (e.g., cosmetic boxes and bottles) whereas the images presenting higher visual complexity look more cluttered and contain more objects.

*Object Type Measures.* We created variables that describe specific object types by first creating a list of unique nouns (887 words) from the *descriptions*. We then computed a semantic similarity matrix using *WordNet,* a large lexical database of English based on a hierarchical structure. All noun hierarchies ultimately go up to the root node called entity. *WordNet* is commonly used in automated text analysis and artificial intelligence applications. With the similarity matrix, we computed a distance matrix. We then employed hierarchical clustering using Ward's method to segment the nouns into similar semantic groups.

Figure 8. Examples from the Instagram Data of Images That Are Tagged with Noun *Descriptions* by the Computer Vision API

| | |
|---|---|
|  (i) |  (ii) |
| Descriptions for image (i): *indoor, bag* | Description for image (ii): *photo, food, woman, birthday, man, group, display, people* |

Figure 9. Example Images from the Instagram Data by Visual Complexity

Low High

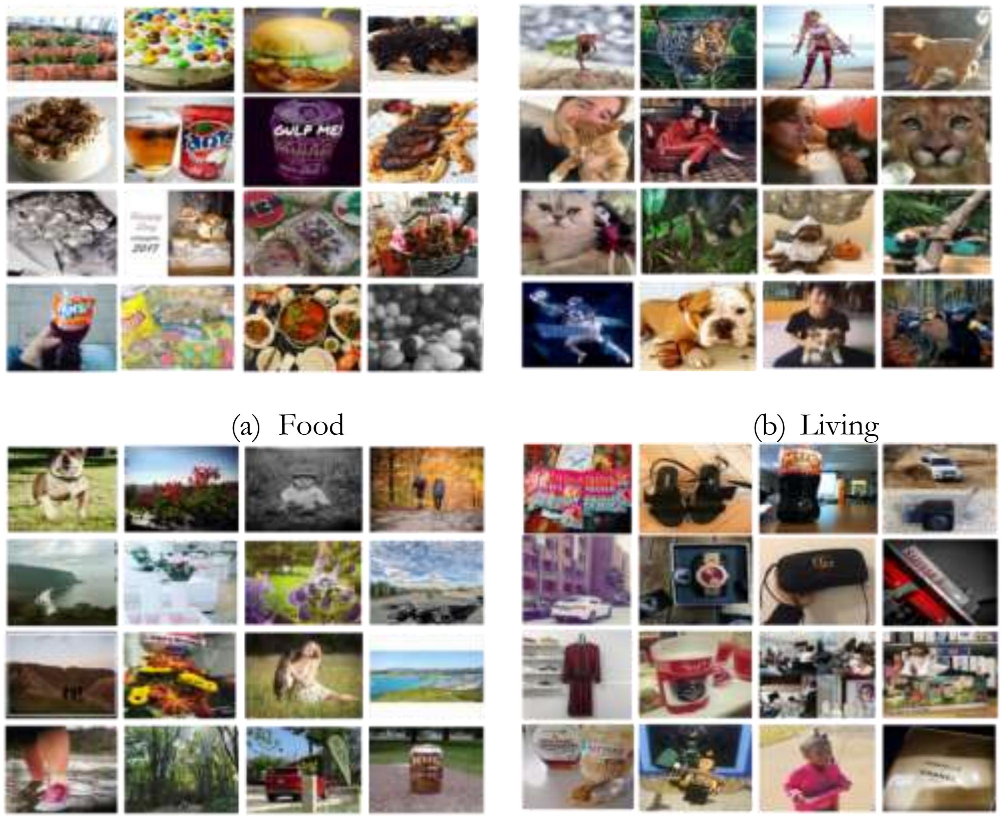
We found that five clusters group the nouns in a semantically similar way. We named the five clusters, *living things, non-living things, food and plants, scenery and events*, and *adjectives*. We discarded the *adjectives* group because adjectives usually do not entail objects or things. Finally, we hand-sorted *plants* from the *food* group. Table A2 in Appendix 3 lists examples of the words in each cluster. We created indicator variables for each of the clusters, each coded as one if the noun *descriptions* that represent an image contains any of the nouns in each cluster. For example, if the noun *descriptions* for an image contain nouns from the *food* group and nouns from the *living things* group, the image is coded as one for each from *food* and *living things* and zero for the other three clusters. Figure 10 presents examples of the images by object type. Finally, we counted the number of faces in each image, extracted from the computer vision API.

#### 2.4.2.3. Text Variables

We measured two aspects of the text data in a post: text sentiment and text complexity. We used hashtags to construct these as hashtags are popularly used by a post creator on Instagram to describe their posts.

*Text Sentiment Measure.* To craft the text sentiment variable, we employed *SentiStrength* (Thelwall et al. 2010) to capture positive and negative sentiment strength. *SentiStrength* is commonly used in text mining, and it codes not only polarity aspects of sentiment (positive, neutral and negative) but also degrees of strength. After extracting text sentiment, we computed divergence-based text sentiment using the formula: *sentiment = (positive – negative) – 2*, following Stieglitz and Xuan (2013). The divergence-based approach (i.e., level of activation) to measure emotions has long been studied in numerous papers (Berglund, Berglund and Engen 1982, Engen, Levy and Schlosberg 1958, Frijda 1969, Schlosberg 1954, Triandis and Lambert 1958) in appraising cognitive patterns in emotions as one of the key features in psychology field. For example, Reisenzein (1994) investigated the intensity

Figure 10. Example Images from the Instagram Data by Object Type



(a) Food

(b) Living

(c) Plant

(d) Non-living

aspect of emotions by applying pleasure-arousal theory (PAT) and proposed hybrid cognitive PAT of emotions. In recent years, many scholars adopt the arousal-based approach in measuring emotions in a web context. Pfitzner et al. (2012), for example, found that emotional divergence in text elements in social media affects users' information spreading behavior, where the emotional divergence is defined as the absolute difference between the positive and negative sentiment score in the text. Stieglitz and Xuan (2013) revealed that there is a positive relationship between emotion arousal and information diffusion when studying information dissemination in social networks. Berger and Milkman (2012) explored the relationship between the level of arousal in emotions and virality in social media and found a positive relationship between them.

*Text Complexity Measure.* Text (or topic) complexity can be approximated through topical diversity and topical variety (Wagner and Strohmaier 2010, Dong and Zhou 2012). Saxton et al. (2015) addressed that hashtags in social media indicate topics or themes. Wagner and Strohmaier (2010), for example, used topic diversity (topic variety) to measure properties of social awareness streams when studying latent conceptual structures and their capability to convey meaningful information. In a Twitter context, the authors used the number of unique hashtags as surrogate measure for the topic diversity.

We computed text complexity following Dong and Zhou (2012) who use the number of unique hashtags as a measure for topic diversity with Twitter data. In Instagram, compound words (e.g., *#happylife*, *#devilloveprada*) are frequently used as hashtags, and several completely different topics can be counted as one with these compound hashtags. Accordingly, instead of using the number of unique hashtags, we counted the number of unique words as our text complexity measure after tokenizing hashtags with compound words into words (e.g., *happy, life, devil, love, prada*).

### 2.4.3. Descriptive Statistics

Table 2 presents descriptive statistics for the dependent variables (*LIKES, COMMENTS*) and the explanatory variables. A brand-related post on Instagram receives, on average, 245 likes and 4 comments, though both are highly skewed and their respective median values are 34 and 1, respectively. Almost all the posts received at least one like (99.5%), while 62.3% elicited one or more comments. For visual sentiment, we find that most of the posts (88.3%) contain positive visual sentiment; only 11.7% contained negative visual sentiment. In terms of object types, 52.6% of the posts contain *living* objects, 13.5% contain *plant* objects, and 32.2% include *food* objects. Finally, Figure 11 presents a correlation plot. Overall, we find very modest correlations among the focal explanatory variables.

Figure 12 presents histograms for numbers of likes (*LIKES*) and comments (*COMMENTS*). Each is highly skewed, approximating a power law distribution. Figure 13, left, is a log-log plot of *LIKES* versus number of posts. It shows that the vast majority of posts receive a relatively small number of likes. Only 2% of posts had more than 2,500 likes, and only 15 posts had more than 50,000 likes. The log-log plot of *COMMENTS* versus number of posts (Figure 13, right) shows a similar pattern: a large number of posts have relatively few comments and only a few posts received a large number of comments. For example, 87% of posts have five or fewer comments and only 632 posts received more than 100 comments.

Next, we created density plots for visual sentiment, visual complexity, text sentiment and text complexity for the 86 brands. Those are shown in Figure 14. Each line in the plot represents a brand. The density plot for visual sentiment shows left-skewness (more posts with positive visual sentiment). For visual complexity, we find two maximum density points: one at the lowest level of visual complexity and one around the midpoint of the range of visual complexity.

In Figures 15 and 16, we report the top and bottom three brands for each brand characteristics measure and Brand Asset Valuator (BAV) measure.

Table 2. Descriptive Statistics

| Variable(s) | N | Mean | Std Dev | Min | Median | Max | Skewness |
|---|---|---|---|---|---|---|---|
| *Consumer Engagement* | | | | | | | |
| LIKES | 84,229 | 245.35 | 1,678.10 | 0 | 34 | 85,453 | 18.44 |
| COMMENTS | 84,229 | 3.85 | 11.49 | 0 | 1 | 101 | 6.51 |
| *Image Content* | | | | | | | |
| VizSentiment | 75,254 | 4.80 | 0.70 | 1.24 | 4.83 | 6.96 | -0.40 |
| VizComplexity | 84,229 | 3.04 | 1.19 | 1.26 | 3.09 | 7.00 | -0.01 |
| *Object Types:* | | | | | | | |
| Living | 84,229 | .53 | .50 | 0 | 1 | 1 | -.10 |
| Food | 84,229 | .32 | .34 | 0 | 0 | 1 | 2.14 |
| Plant | 84,229 | .13 | .47 | 0 | 0 | 1 | .76 |
| NumFaces | 84,229 | .38 | .99 | 0 | 0 | 10 | 5.07 |
| *Text Content* | | | | | | | |
| TextSentiment | 84,229 | 2.05 | 1.12 | 1.00 | 1.86 | 7.00 | .85 |
| TextComplexity | 75,254 | 2.72 | 1.20 | 1.00 | 2.59 | 7.00 | .27 |
| TextLength | 84,229 | 40.18 | 37.40 | 2 | 32 | 918 | 4.16 |
| *Brand Characteristics* | | | | | | | |
| Visibility | 81,281 | 3.13 | .45 | 1.94 | 3.16 | 3.94 | -.44 |
| Involvement | 81,281 | 3.77 | .35 | 3.09 | 3.69 | 4.32 | .18 |
| PerceivedRisk | 81,281 | 1.74 | .33 | 1.20 | 1.72 | 2.47 | .21 |
| *Brand Equity:* | | | | | | | |
| Relevance | 81,281 | 2.83 | .72 | 1.45 | 2.80 | 4.33 | .27 |
| Differentiation | 81,281 | .57 | .15 | .32 | .53 | 1.08 | .97 |
| Esteem | 81,281 | .75 | .26 | .23 | .70 | 1.43 | .46 |
| Knowledge | 81,281 | 3.85 | .69 | 1.93 | 3.93 | 4.94 | -.66 |
| *Product Characteristics* | | | | | | | |
| ExpGood | 81,281 | .66 | .47 | 0 | 1 | 1 | -.67 |
| Premium | 81,281 | .32 | .47 | 0 | 0 | 1 | .76 |
| *User Characteristics* | | | | | | | |
| NumFollowers | 83,978 | 8,075.85 | 80,429.26 | 0 | 512 | 4,250,315 | 34.46 |
| NumFollowings | 83,978 | 864.66 | 1,356.68 | 0 | 394 | 8,162 | 3.07 |
| PostCount | 83,978 | 852.80 | 2,306.22 | 1 | 306 | 114,779 | 16.05 |

*Notes.* The statistics for VizSentiment, VizComplexity, TextSentiment and TextComplexity are using the rescaled (all are on a 1-7 scale) values. For estimation, these four variables are zero centered. For estimation, NumFollowers is rescaled to 0-100 scale.

Figure 11. Correlation Plot

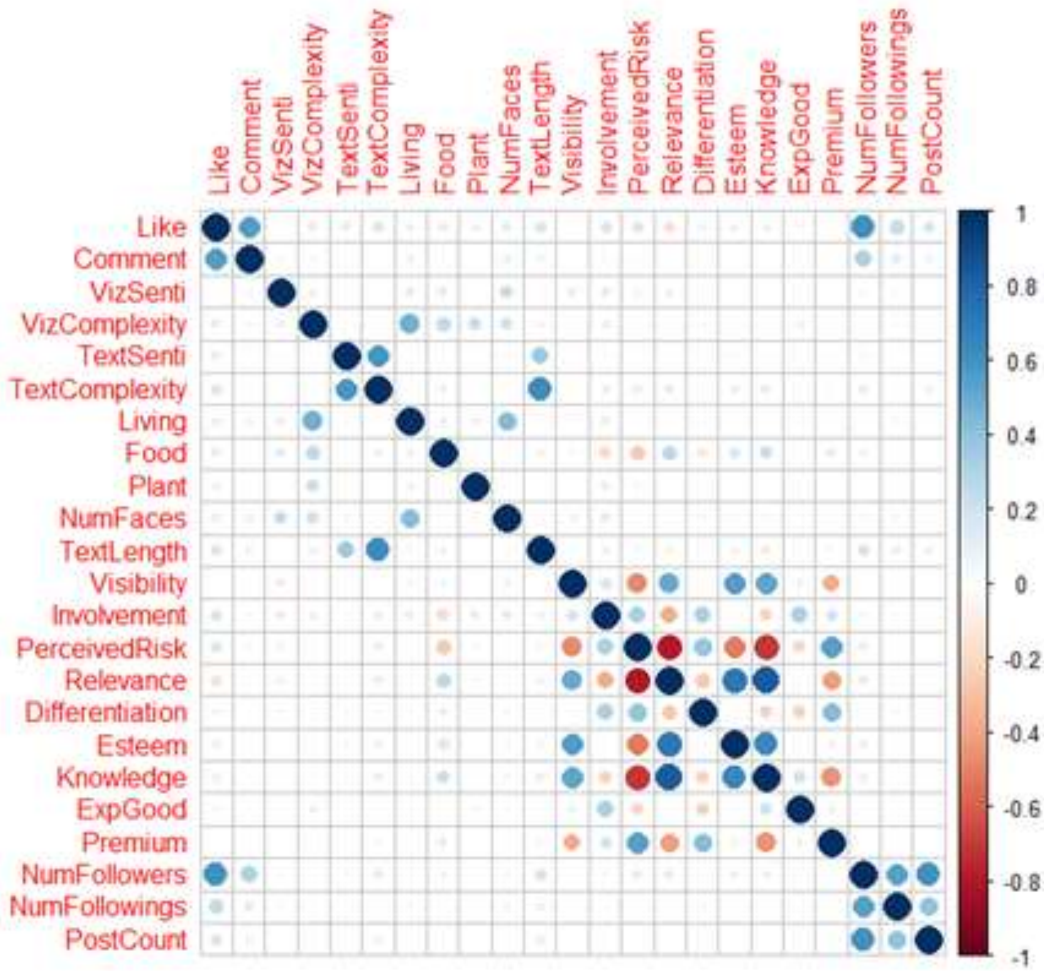
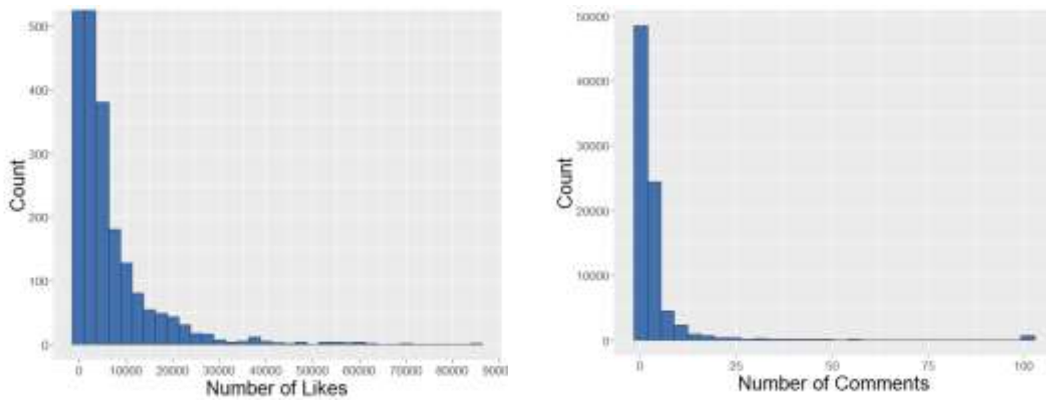


Figure 12. Histogram of Likes and Comments




*Note*. The top of the first and second bins for LIKES are cut because the histogram is highly skewed. The first bin (0-2,500 LIKES) is 82,769 and the second bin (2,500-50,000 LIKES) is 666.

Figure 13. Log-log Plot of Number of Posts

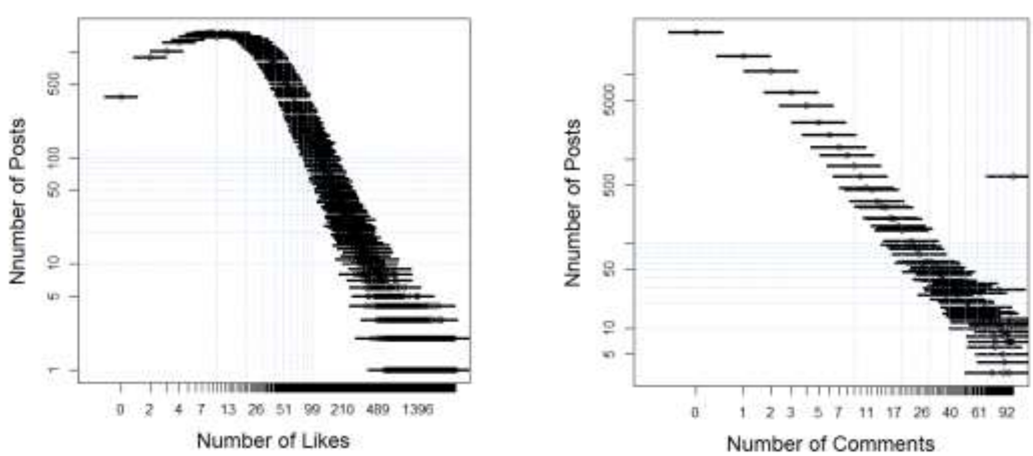
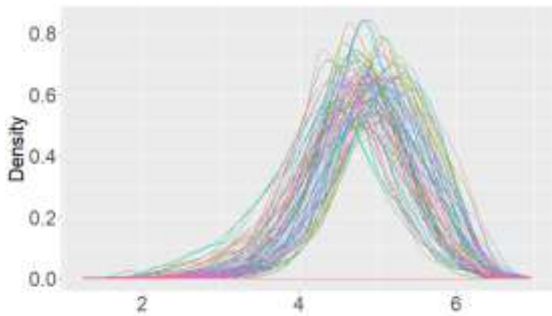
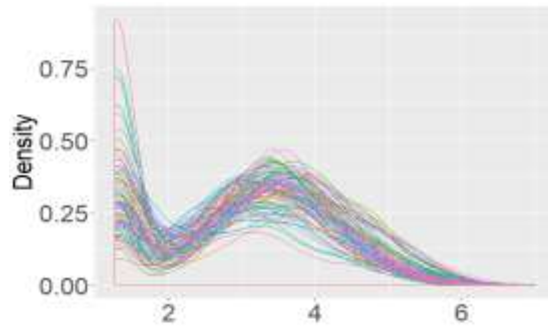


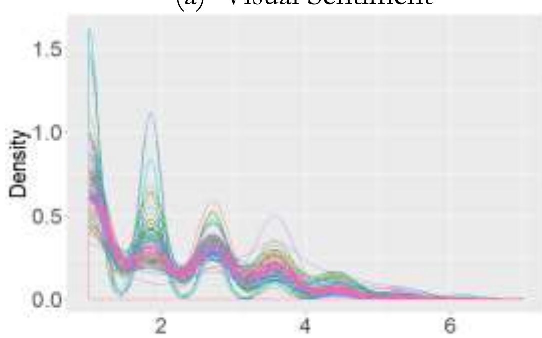
Figure 14. Density Plots at the Brand Level



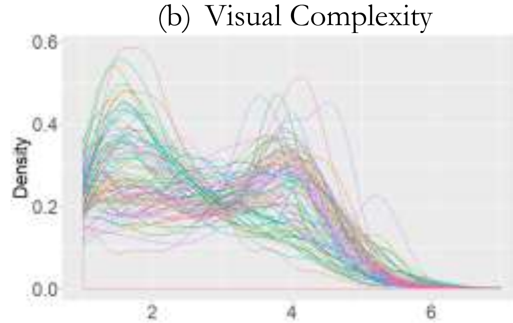

(a) Visual Sentiment




(b) Visual Complexity




(c) Text Sentiment




(d) Text Complexity

*Note.* These are the variables after re-scaling so all are on a 1-7 scale for use in the model.

Figure 15. Top 3 and Bottom 3 Brands for the Brand Characteristics Measures

| Visibility | | Involvement | | Perceived Risk | |
|---|---|---|---|---|---|
| Ford | 3.973 | Ferrari | 4.321 | Dolce and Gabbana | 2.467 |
| Jeep | 3.869 | Porsche | 4.321 | Prada | 2.5 |
| Chevrolet | 3.789 | Jaguar | 4.321 | Louis Vuitton | 2.343 |
| Dolce and Gabbana | 2.183 | Fanta | 3.088 | M&Ms | 1.209 |
| Prada | 2.154 | Mug Root Beer | 3.088 | Pepsi | 1.200 |
| Ferrari | 1.983 | Kool Aid | 3.088 | Snickers | 1.200 |

Figure 16. Top 3 and Bottom 3 Brands for the Brand Asset Valuator (BAV) Measures

| Relevance | | Differentiation | | Esteem | | Knowledge | |
|---|---|---|---|---|---|---|---|
| Hershey | 4.327 | Porsche | 1.085 | Black and Decker | 1.434 | Oreos | 4.943 |
| M&Ms | 4.165 | Wii | 0.921 | Fisher Price | 1.383 | M&Ms | 4.943 |
| Oreos | 4.100 | iPod | 0.884 | Nike | 1.234 | Hershey | 4.921 |
| Ferrari | 1.539 | Suzuki | 0.374 | Suzuki | 0.306 | Under Armour | 2.173 |
| Suzuki | 1.487 | Kool Aid | 0.364 | Garnier Fructis | 0.249 | Prada | 2.159 |
| Land Rover | 1.447 | Ajax | 0.325 | Fanta | 0.235 | Dolce and Gabbana | 1.926 |

2.5. Empirical Analysis

We estimate a negative binomial regression for *LIKES*, and a zero-inflated negative binomial regression for *COMMENTS*. In general, we model *LIKES* (and *COMMENTS*) for a user-generated brand-themed post as a function of the image content of the post; the text content of the post, characteristics of the focal brand, and characteristics of the user. We estimate the following equation for $LIKES_i$ as a dependent variable:

$$LIKES_i = \alpha_0 + \sum_{j=1}^{k} \beta_{1j} * ImageContent_{i,j} + \sum_{j=1}^{l} \beta_{2j} * TextContent_{i,j} + \sum_{j=1}^{n} \beta_{3j} * BrandCharacteristics_{i,j} + \sum_{j=1}^{s} \beta_{4j} * UserCharacteristics_{i,j} + \varepsilon_i \quad (1)$$

where $LIKES_i$ is the number of likes earned by post *i,* $\beta_{1j}$ is a vector of coefficients of the image content descriptors (visual sentiment, visual complexity, whether the image contains a living thing, food item, and/or plant, and the number of human faces), $\beta_{2j}$ is a vector of coefficients of the text content descriptors (text sentiment, text complexity, length of the text), $\beta_{3j}$ is a vector of brand (and product) characteristics (visibility; involvement; perceived risk; brand equity; whether the good is an experience good, and whether it is a premium brand), and $\beta_{4j}$ is a vector of three user characteristics (number of followers, number followings, number of posts to date). The same explanatory variables were used to model *COMMENTS*.

2.5.1. Results

Tables 3 and 4 present the results for the models of *LIKES* and *COMMENTS*, respectively. The right-far-right column is our full model. The first two columns of results are restricted versions that we estimated to check the stability of the results from the full model compared with selected restricted versions.

A third-order operationalization for visual complexity and text complexity fits the data better than both first-order and second-order operationalizations for both the *LIKES* and *COMMENTS* models[5]. To assist with interpretation, Figure 19 plots the total effect of visual complexity over its range for the *LIKES* model using the coefficients from Table 3: .18 VizComplexity + .11

[5]For the *LIKES* model, the AIC with the two complexity measures coded as third-order terms is 785,869, versus 791,003 when they are first-order only, and 788,815 when they are second-order; For the *COMMENTS* model, the AICs when the two complexity variables are coded as first-order, second-order, and third-order are 306,792, 308,370, and 306,792, respectively.

(VizComplexity)$^2$ -.04 (VizComplexity)$^3$.

The explanatory variables contain an additive effect in the ln(*LIKES*) and ln(*COMMENTS*) scales, and a multiplicative effect ($e^{\beta x}$) in the *LIKES* and *COMMENTS* scales. To facilitate interpretation, Figures 16 and 17 plot the multiplicative effect of the key explanatory variables in the *LIKES* scales for the variables defined over their ranges and their quantiles, respectively.

2.5.1.1. Consumer Engagement and Image Content

*Image Sentiment.* Higher visual sentiment is associated with more likes (.05; $p$<.0001) and more comments (.11; $p$<.0001). Over the range of our data, the multiplicative effect ranges from .84 to 1.11.

*Image Complexity.* Over the range of our data, we find that the relationship between visual complexity and number of likes is bimodal, with a strong positive effect at the minimum and just past the middle of the scale, no effect for modest values, and a strong negative effect at high levels of visual complexity. A third-order relationship fits the data best (.18 VizComplexity + .11 [VizComplexity]$^2$ - .04 [VizComplexity]$^3$).

Figure 19 Plot (a) graphically depicts the relationship between visual complexity and the number of likes in the ln(*LIKES*) scale, showing that the first inflection point occurs approximately when visual complexity is negative two and the second inflection point occurs at positive two. Since we rescaled and centered the variables, we computed them back to the original value and find that the first inflection point corresponds to four, and the second inflection point corresponds to 15.5 of the original visual complexity measure. The results when *COMMENTS* are the dependent variable are similar.

2.5.1.2. Consumer Engagement and Text Content

*Text Sentiment.* Consumer engagement is higher when a post includes more divergent emotions in its text content. We find a positive and significant relationship between text sentiment and number of likes in the ln(*LIKES*) scale (.09; $p$<.0001). The multiplicative effect of text sentiment ranges from .91 to 1.56 over the range of our data.

*Text Complexity*. We find a positive relationship at low levels of text complexity and at moderate levels of text complexity and a negative relationship at high levels of text complexity between text complexity and *LIKES* (Table 3). Plot (b) in Figure 19 graphs the estimated total effect of text complexity for *LIKES*. The first inflection point occurs approximately at -.5 of text complexity and the second inflection point occurs at two. From Plot (b), we also find that the total effect is positive or near zero until a threshold (3.25), then becomes negative after the threshold. To find the threshold, we computed the rescaled value back to the original value and find that the effect starts to decrease when a post contains 42 unique words in its hashtags. Figure 18 makes clear that text sentiment has the largest relative effect on our post characteristic descriptors.

We find similar results for text complexity effect on comments; the total text complexity effect is positive or minimal at low or moderate levels of text complexity and becomes negative at a certain point. We also computed the value back to the original scale and find that the effect starts to decrease at the point where a post contains 42 unique words and becomes negative at 54 unique words. Consumers are more engaged when a post contains a small number of unique words in its hashtags, but are not engaged when the post contains too much information.

Table 3. Consumer Engagement: Number of *LIKES*

| | *LIKES* Model 1 | *LIKES* Model 2 | *LIKES* Model 3 (full) |
|---|---|---|---|
| Intercept | 1.49***(.13) | 2.61***(.10) | 2.61***(.10) |
| *Image Content* | | | |
| VizSentiment | .16***(.01) | .04***(.01) | .05***(.01) |
| VizComplexity | .14***(.01) | .09***(.01) | .10***(.01) |
| VizlComplexity^2 | .10***(.00) | .03***(.00) | .03***(.02) |
| VizComplexity^3 | -.03***(.00) | -.02***(.00) | -.02***(.00) |
| *Object Types:* | | | |
| Living | .13***(.01) | .06***(.01) | .06***(.01) |
| Food | -.25***(.01) | -.18***(.01) | -.18***(01) |
| Plant | -.07***(.02) | .11***(.01) | .11***(.01) |
| NumFaces | .03***(.01) | .05***(.00) | .05***(.00) |
| *Text Content* | | | |
| TextSentiment | .12***(.01) | .09***(.00) | .09***(.00) |
| TextComplexity | .03*(.01) | .25***(.01) | .25***(.01) |
| TextComplexity^2 | .43***(.01) | .24***(.00) | .24***(.00) |
| TextComplexity^3 | -.15***(.00) | -.10***(.00) | -.10***(.00) |
| TextLength | .005***(.00) | .002***(.00) | .002***(.00) |
| *Brand Characteristics* | | | |
| Visibility | .31***(.02) | .19***(.01) | .11***(.01) |
| Involvement | .31***(.02) | .16***(.02) | .28***(.02) |
| PerceivedRisk | .32***(.03) | -.08***(.02) | -.13***(.03) |
| *Brand Equity:* | | | |
| Relevance | -.43***(.02) | -.18***(.01) | -.19***(.01) |
| Differentiation | .60***(.04) | .64***(.03) | .66***(.03) |
| Esteem | -.90***(.03) | -.44***(.03) | -.38***(.03) |
| Knowledge | .43***(.01) | .13**(.01) | .14***(.01) |
| *User Characteristics* | | | |
| NumFollowers[1] | | 1.62***(.00) | 1.60***(.00) |
| NumFollowing | | .00008***(.00) | .00008***(.00) |
| PostCount | | -.00004***(.00) | -.00004***(.00) |
| *Product Characteristics* | | | |
| ExpGood | | | -.16***(.01) |
| Premium | | | -.15***(.01) |
| AIC | 845,299 | 786,274 | 785,867 |
| Deviance | 93,201 | 85,296 | 85,246 |
| Alpha | 2.345 | 1.343 | 1.337 |
| N | 72,408 | 72,194 | 72,194 |

. $p<.05$; *$p<.01$; **$p<.001$; ***$p<.0001$

*Note.* [1] NumFollowers re-scaled to the range 0-100.

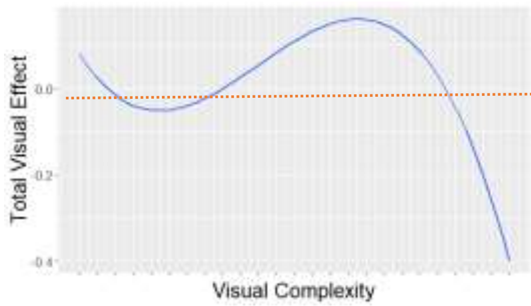
Table 4. Consumer Engagement: Number of *COMMENTS*

| | *COMMENTS* Model 1 | *COMMENTS* Model 2 | *COMMENTS* Model 3 (full) |
|---|---|---|---|
| Intercept | -1.01***(.14) | -0.03 (.14) | -0.03 (.14) |
| *Image Content* | | | |
| VizSentiment | .16***(.01) | .11***(.01) | .11***(.01) |
| VizComplexity | .03**(.01) | .02. (.10) | .02*(.01) |
| VizComplexity^2 | .04***(.01) | .004(.00) | .004(.00) |
| VizComplexity^3 | -.01***(.00) | -.007*(.00) | -.007*(.00) |
| *Object Types:* | | | |
| Living | .14***(.02) | .11***(.02) | .12***(.02) |
| Food | -.11***(.02) | -.07***(.01) | -.08***(01) |
| Plant | .002(.02) | .04*(.02) | .04*(.02) |
| NumFaces | .04***(.01) | .04***(.01) | .04***(.01) |
| *Text Content* | | | |
| TextSentiment | .12***(.01) | .12***(.01) | .12***(.01) |
| TextComplexity | .01(.01) | .02 .(.01) | .01(.01) |
| TextComplexity^2 | .30***(.01) | .21***(.01) | .21***(.01) |
| TextComplexity^3 | -.10***(.00) | -.07***(.00) | -.07***(.00) |
| TextLength | .006***(.00) | .005***(.00) | .005***(.00) |
| *Brand Characteristics* | | | |
| Visibility | .12**(.02) | .10***(.02) | .06**(.02) |
| Involvement | .17***(.03) | .02(.02) | .12***(.03) |
| PerceivedRisk | .07.(.04) | -.09**(.03) | -.18***(.04) |
| *Brand Equity:* | | | |
| Relevance | .0002(.02) | .08***(.02) | .06**(.02) |
| Differentiation | .55***(.05) | .56***(.05) | .49***(.05) |
| Esteem | -.67***(.04) | -.48***(.04) | -.52***(.04) |
| Knowledge | .13***(.02) | .01(.01) | .04**(.02) |
| *User Characteristics* | | | |
| NumFollowers[1] | | .65***(.01) | .64***(.01) |
| NumFollowing | | .00005***(.00) | .00005***(.00) |
| PostCount | | -.00006***(.00) | -.00006***(.00) |
| *Product Characteristics* | | | |
| ExpGood | | | -.14***(.02) |
| Premium | | | -.01(.02) |
| AIC | 317,524 | 306,254 | 306,176 |
| Log-likelihood | -15,870 | -15,310 | -15,300 |
| Alpha | 2.52 | 1.94 | 1.93 |
| N | 72,408 | 72,194 | 72,194 |

.*p*<.05; **p*<.01; ***p*<.001; ****p*<.0001

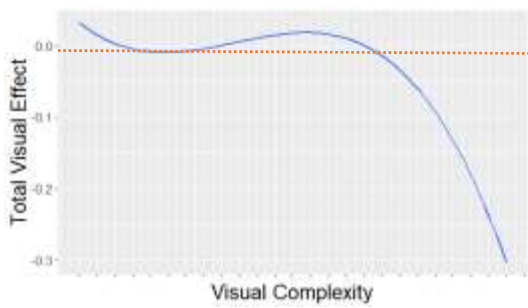
*Note.* [1] NumFollowers re-scaled to the range 0-100.

Figure 17. Visual and Text Complexity Effect on User Engagement

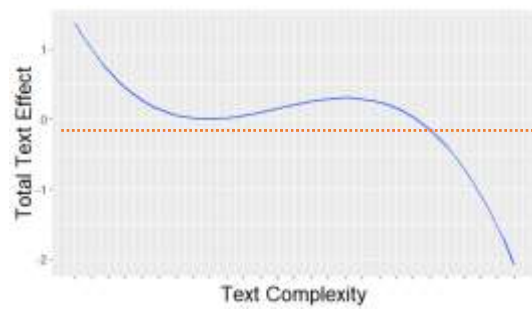



Plot (a). Visual complexity (*likes*)


Plot (b). Text complexity (*likes*)




Plot (c). Visual complexity (*comments*)




Plot (d). Text complexity (*comments*)

2.5.1.3. Consumer Engagement and Brand Characteristics

*Visibility; Involvement.* All other things being equal, brand-themed posts achieve more likes when the focal brand is more visible (.11; $p$<.0001) and has higher involvement (.28; $p$<.0001). Over the range of our data, their multiplicative effects (Figure 18) in the *LIKES* scale range from 1.24 and 2.37 at their minimum values, respectively, to 1.54 and 3.35 at their maximum values. Consumers engage more heavily with higher involvement brands offline, and this carries over to their social media activity.

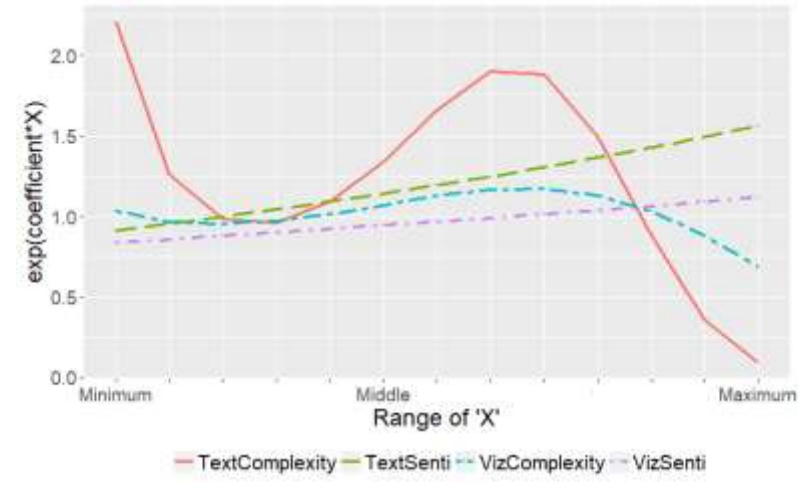
*Perceived Risk.* Brands viewed as having higher perceived risk earn fewer likes with a brand-themed user post (-.13; $p$<.0001). However, the effect over the range of our data is relatively modest compared with involvement. Over the range of our data, the multiplicative effect ranges from .86 to .73. We find the same directional results for the *COMMENTS* model.

*Brand Equity*. All other things being equal, Differentiation (.66; $p$<.0001) and Knowledge (.14; $p$<.0001) are positively associated with a brand-themed post achieving more likes. Over the breadth of our data, their multiplicative effects (Figure 18) in the *LIKES* scale range from 1.24 and 1.31 at their minimum values, respectively, to 2.05 and 2.00 at their maximum values. Relevance (-.19; $p$<.0001) and Esteem (-.38; $p$<.0001) both have attenuating effects on number of likes achieved by a brand themed post. Over the breadth of our data, their multiplicative effects (Figure 18) in the *LIKES* scale range from .76 and .92 at their minimum values, respectively, to .44 and .58 at their maximum values.

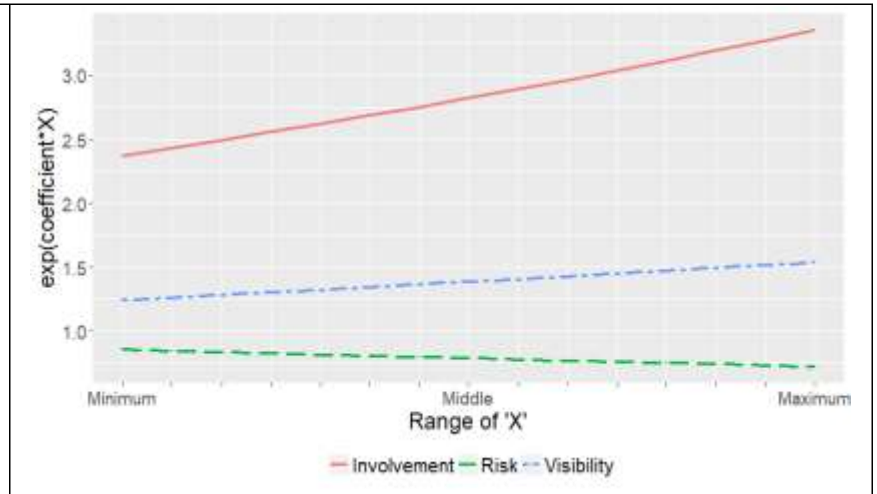
2.5.1.4. Consumer Engagement and User Characteristics

The size of the user's network has a significant effect on consumer engagement with her or his posts. All other things being equal, users with a larger number of followers earn more likes (1.60; $p$<.0001) and more comments (.64; $p$<.0001); users following a larger number of other users earn more likes (.00008; $p$<.0001) and more comments (.0005; $p$<.0001). The number of posts the user has created to date has a negative effect on the number of likes (-.00004; $p$<.0001) and number of comments (-.00006; $p$<.0001). Users with an excessive number of posts could be flippantly posting without much thought as to the relevance of posts.

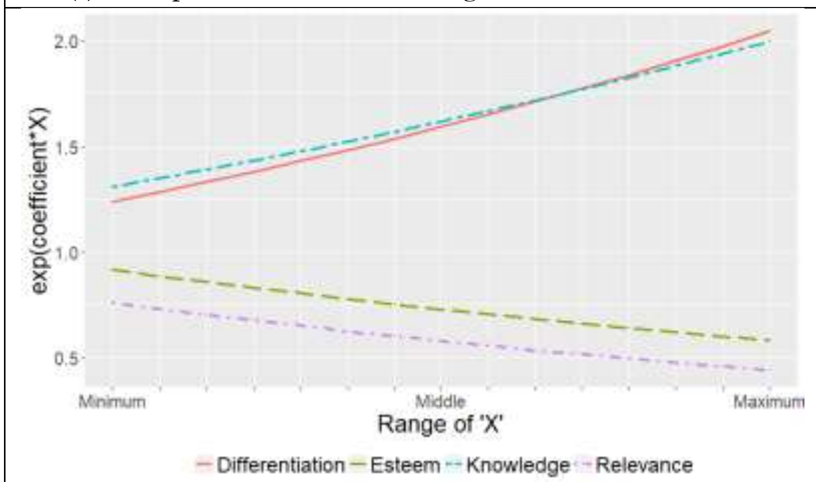
Figure 18. Multiplicative Effects on LIKES of the Metric Variables Over Their Range



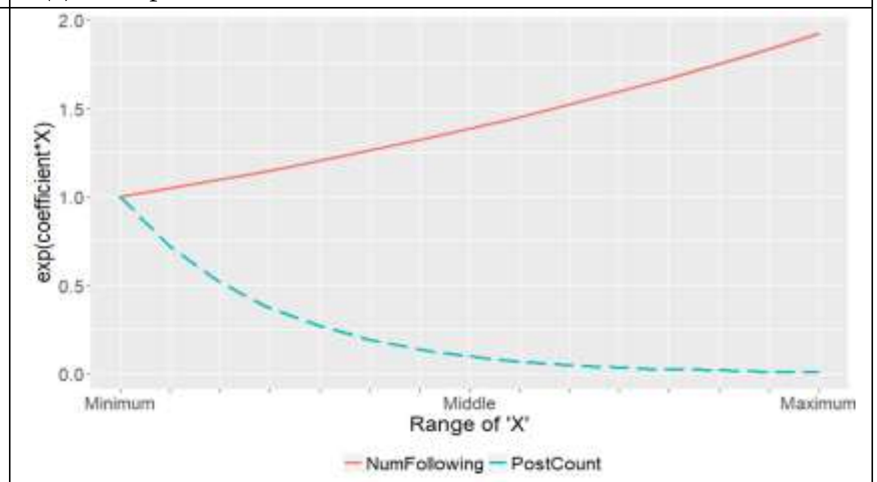

(a) Multiplicative Effects of Image and Texts on LIKES




(b) Multiplicative Effects of the Brand Characteristics on LIKES

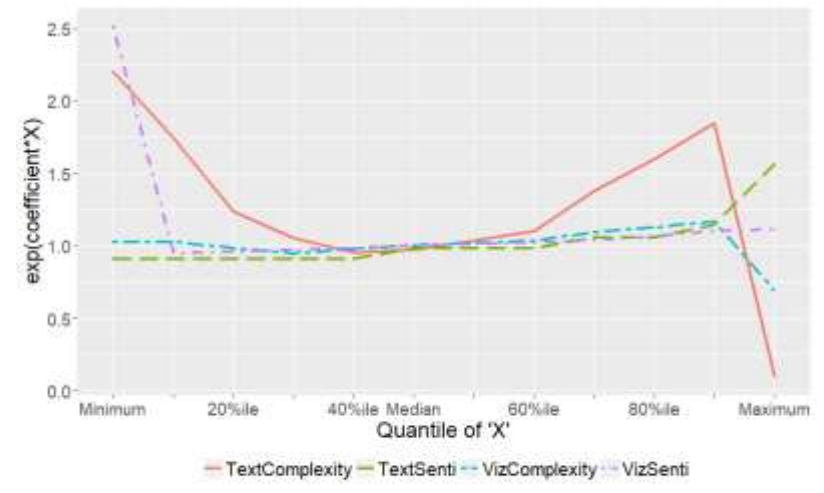



(c) Multiplicative Effects of Brand Equity on LIKES

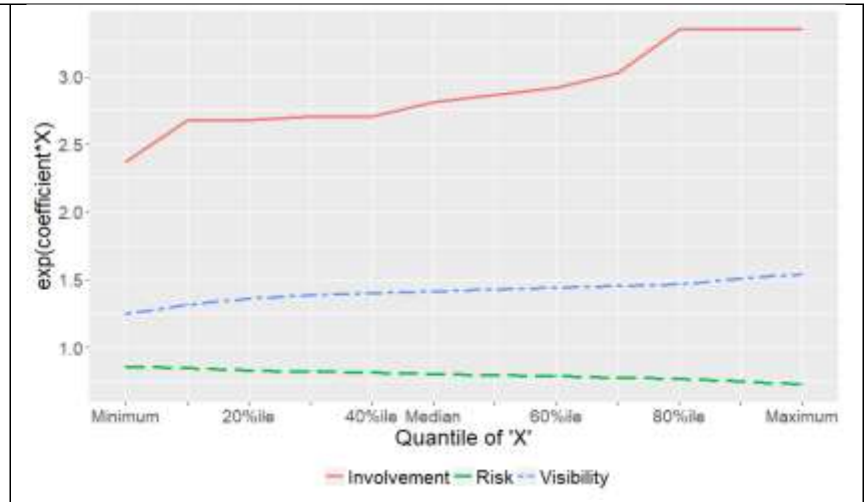



(d)Multiplicative Effects of the User Characteristics on LIKES

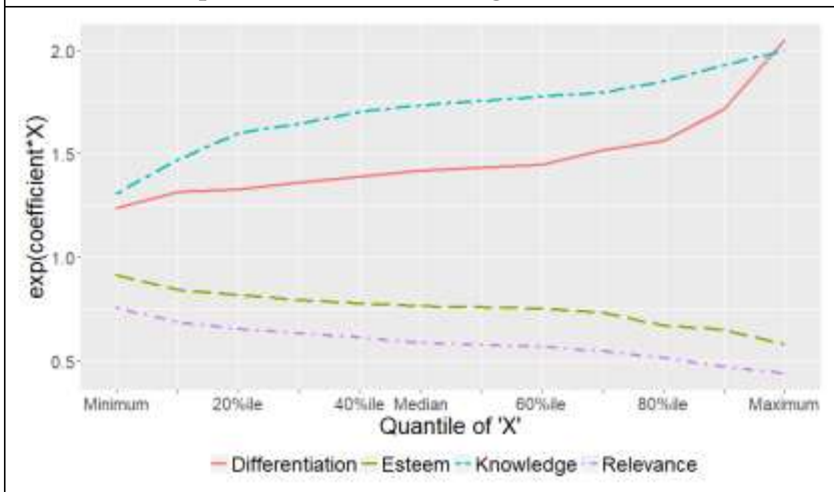
Figure 19. Multiplicative Effects on LIKES of the Metric Variables over Their Quantiles



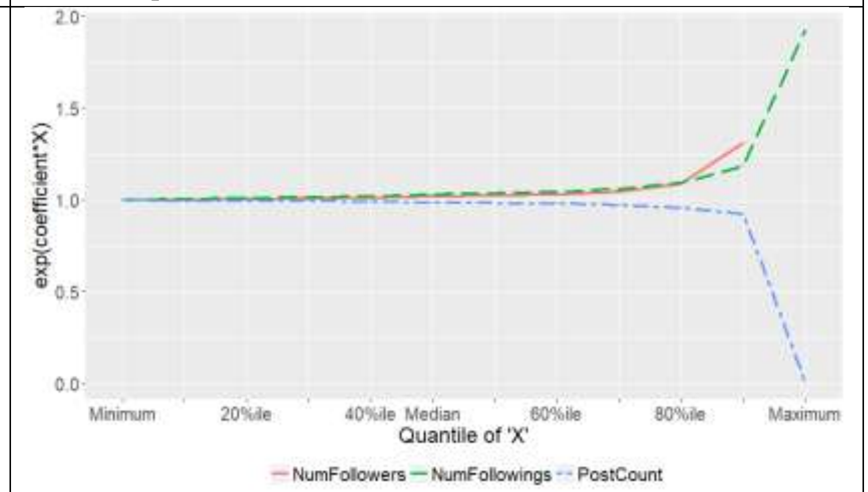

(a) Multiplicative Effects of Image and Text on LIKES




(b) Multiplicative Effects of the Brand Characteristics on LIKES




(c) Multiplicative Effects of Brand Equity on LIKES




(d) Multiplicative Effects of User Characteristics on LIKES

### 2.5.2. Simulation

To demonstrate the managerial relevance of the results and the implications of a post's visual and text content, a simulation was conducted using the model estimates from Table 3. All metric variables were set to their mean values, and all dichotomous variables were set to zero. We calculated the predicted number of likes and called this the "base." We then calculated the change in the predicted number of likes from varying image sentiment, image complexity, text sentiment, and text complexity, each at its mean and +/–2 standard deviations from its mean (i.e., a total of 3×3×3×3=81 simulations). The results are reported in Table 5. The simulation results make salient the large impact that text complexity has on earned engagement with a brand-themed post.

### 2.5.3. Robustness Checks

We conducted several robustness checks (Appendix 4). First, we considered alternative ways to measure brand equity: 1) the four BAV pillars (Relevance, Differentiation, Esteem, and Knowledge; used in this paper), 2) the two BAV summary measures (Brand Stature and Brand Strength, 3) and the single composite BAV measure, Brand Asset. The substantive conclusions hold under all three alternative operationalizations.

Second, we examined a model in which we took the log form and different functional forms (*ln*(X); i.e., nonlinear, monotone) for the explanatory variables with skewness over two and the log form for the outcome variable, *ln*(y), using ordinary least squares. The substantive conclusions do not change for the *LIKES* model (Tables A5, A6 and A7); there are some minor differences in the *COMMENTS* model.

Finally, we estimated models using only the middle 98% of the data so as to exclude posts with a very large number of likes and/or comments. The main qualitative findings are robust.

Table 5. Simulated Change in Number of *LIKES* for Changes in Image Content and Textual Content
(reference or base is all four variables at their mean value)

| | | | Image Content | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sentiment | -2 SD | -2 SD | -2 SD | Mean | Mean | Mean | +2 SD | +2 SD | +2 SD |
| | Sentiment | Complexity | -2 SD[1] | Mean | +2 SD | -2 SD[1] | Mean | +2 SD | -2 SD[1] | Mean | +2 SD |
| Text Content | -2 SD[1] | -2 SD[1] | 351 | 278 | 376 | 380 | 302 | 354 | 412 | 327 | 383 |
| | -2 SD[1] | Mean | -3 | -12 | -6 | 0 | -9 | -3 | 4 | -6 | 0 |
| | -2 SD[1] | +2 SD | 35 | 19 | 29 | 41 | 24 | 35 | 47 | 30 | 41 |
| | Mean | -2 SD[1] | 441 | 351 | 411 | 476 | 380 | 444 | 514 | 412 | 480 |
| | Mean | Mean | 7 | -3 | 4 | 11 | **Base** | 7 | 15 | 4 | 11 |
| | Mean | +2 SD | 53 | 35 | 47 | 61 | 41 | 54 | 69 | 47 | 61 |
| | +2 SD | -2 SD[1] | 550 | 441 | 513 | 594 | 476 | 554 | 640 | 514 | 598 |
| | +2 SD | Mean | 20 | 7 | 15 | 25 | 11 | 20 | 30 | 15 | 25 |
| | +2 SD | +2 SD | 76 | 53 | 68 | 85 | 61 | 77 | 95 | 69 | 86 |

*Notes.* All other metric variables set to their mean values. All dichotomous variables are set to zero.

1 Outside the range of the data.

### 2.5.4. Accounting for Commercial Posts and Multiple Brands

Some photos tagged #*brand* on Instagram may be items intended for sale or resale purposes and may have different content characteristics. For example, such photos could be simple, with only a product in the picture, and provide more directly informative content, such as the price and contact information, in the text. To assess whether commercial posts exhibit a different process leading to engagement we analyzed the data after filtering out posts with sale or resale purposes.

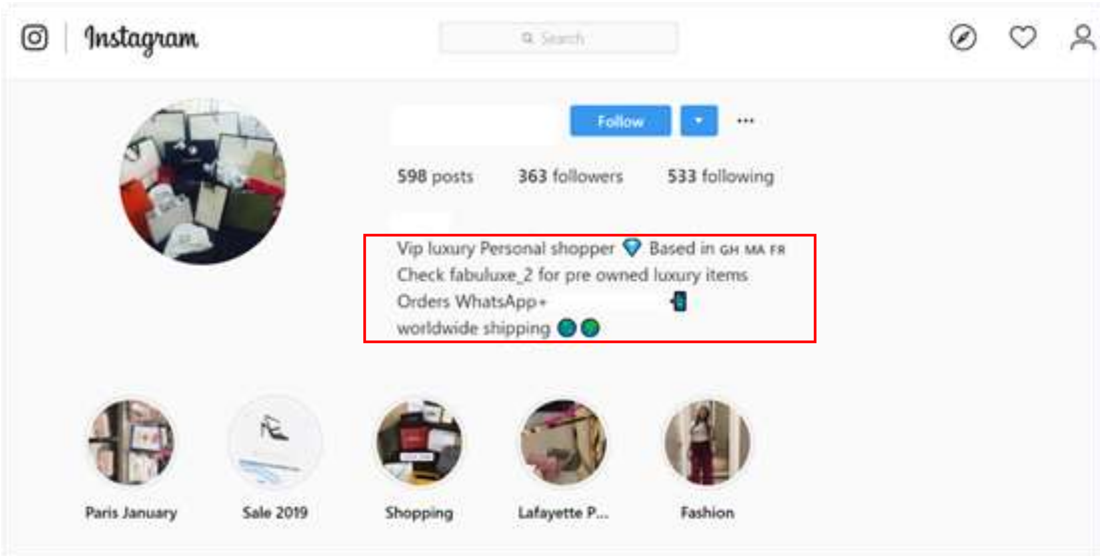
In some Instagram posts, more than one brand appears in an Instagram post. To understand how sensitive our results are to there being multiple brands in a post, we grouped the posts into those that mention only a focal brand or the focal brand plus others brands in their text.

#### 2.5.4.1. Data Cleaning Process

We started the data cleaning process by filtering out posts intended for sale and resale purposes. Instagram offers two types of interfaces: a user page and a feed page. The Instagram feed page displays post-related variables, such as engagement measures (e.g., number of likes) or comments, whereas the Instagram user page contains user-related variables, such as the user ID or descriptions of users. We identified posts with sale or resale purposes at the user level since account owners usually describe the nature of the Instagram account in the user description on the user page (see an example of a user description in the red box in Figure 20. If the Instagram account is used for sale or resale purposes, the account owner will mention the types of products that he or she sells or the shipping methods using common commercial words (e.g., order, shipping, shop, price). We assume that commercial words are an indicator of the nature of an account (i.e., commercial account vs. general account) and thus of the nature of the posts that are associated with the account.

Manually checking the individual user pages in our dataset to determine whether the user description contains commercial words is impossible due to the enormous amount of data. Instead, we randomly sampled about 1% of the entire dataset (867 posts) and manually inspected the user

Figure 20. Example of Commercial-Related Instagram Account




descriptions associated with these posts to determine if they contained any commercial words or explicitly stated that the posts are for sale or resale purposes. Figure 21 shows the list of commercial words that were found in the sample. Using regular expression and text mining libraries, we filtered out users that contain any of the words in the list in their user descriptions. Although Figure 21 does not show this, we also filtered out the users that have variations of the words in their user descriptions. For example, we removed posts associated with users that included “Shopper,” “SHOPPER,” “shoppers,” “Shoppers,” or “SHOPPERS,” even though only “shopper” was on the list. The removed posts account for approximately 9.05% (67,513 observations) of the entire set of observations.

Next, we segmented the posts into two groups: those with a focal brand and those with multiple brands including the focal brand in their hashtags. It is impossible to perform this task with automatic detection since there is no consensus regarding the list of brands that may appear in the hashtags. The most comprehensive brand list we are aware of, which contains over 168,000 brands, is found in the Stradegy database. In addition, automatic detection is impossible because brands are

Figure 21. List of Commercial-Related Words

[“pre-orders”, “shop”, “business”, “order”, “delivery”, “seller”, “shopper”, “whatsapp”, “shoppee”, “reseller”, “buy”, “sale”, “product”, “shopping”, “quality”, “customer”, “service”, “sell”, “price”, “selling”, “ship”, “deal”, “retail”, “store”, “shipping”, “refund”, “brand”, “return”, “sold”, “discount”, “buying”, “ebay”, “sales”, “buyer”]

frequently referred to in a variety of forms on social media platforms such as Instagram. For example, Instagram hashtags do not allow spaces within hashtags, but a number of brand names contain spaces (e.g., Dolce and Gabbana), and so consumers must use *#dolceandgabbana* instead of *#dolce and gabbana*. Thus, due to the vast amount of brands and variations in their names, we were not able to adopt an automatic method. Instead, we chose a manual inspection method for detecting posts that include multiple brands in their hashtags. We recruited seven students from undergraduate programs at a university in the US to inspect the hashtags and group those containing single brand or multiple brands. The students were instructed to detect sets of hashtags that include multiple brands (see Appendix 2 for more specific instruction).

2.5.4.2. Descriptive Statistics

Table 6 displays the descriptive statistics, including mean, median, standard deviation, minimum, maximum, and skewness, associated with the data set from which commercial posts were removed. The descriptive statistics are comparable to the those associated with the initial data set (Table 2), with some slight differences. For example, the text length in the original data set is longer than that in the current data set. Later, we report the scatterplots and density plots (Figures 22 and 23) and correlation matrix (Table 7).

In this section, we use two data sets. The first (Set1) is the original data set that was used in a previous section (§2), in which each post is labeled as either commercial or general. The second (Set2) is a subset of the original data set in which commercial posts were removed and individual posts were labeled as containing either multiple brands or a focal brand in their hashtags. All of the descriptive statistics that are reported in this section relate to Set2 since we reported the statistics for Set1 in a previous section (§2).

Since we are interested in the interaction effects between different types of posts (commercial vs. general) and the focal variables (visual sentiment, visual complexity, text sentiment, and text complexity) in the main analyses, we look at the marginal distributions of both numeric variables (Table 8) and categorical variables (Tables 9, 10, 11, 12, and 13) for the commercial and general posts. For a similar reason, we also report the marginal distributions of the variables for Set2 to determine whether there are differences between posts with multiple brands and those with a single brand in their hashtags (Tables 14, 15, 16, 17, 18, and 19). The model-free evidence shows that general posts are more visually complex but less textually complex than commercial posts. Interestingly, more faces appear in general posts than in commercial posts. It also seems that more diverse types of objects (e.g., food) appear in general posts than in commercial posts based on cross-tabulations.

Table 6. Descriptive Statistics: removing commercial-related posts

| Variable(s) | N | Mean | Std Dev | Min | Median | Max | Skewness |
|---|---|---|---|---|---|---|---|
| *Consumer Engagement* | | | | | | | |
| LIKES | 50,000 | 214.60 | 1,487.24 | 0 | 37 | 85,453 | 22.16 |
| COMMENTS | 50,000 | 3.68 | 10.38 | 0 | 1 | 101 | 6.88 |
| *Image Content* | | | | | | | |
| VizSenti | 50,000 | 4.79 | 0.70 | 1.24 | 4.82 | 6.96 | -.41 |
| VizComplexity | 50,000 | 3.13 | 1.17 | 1.26 | 3. | 7.00 | -.09 |
| *Object Types:* | | | | | | | |
| Living | 50,000 | .54 | .50 | 0 | 1 | 1 | -.18 |
| Food | 50,000 | .34 | .47 | 0 | 0 | 1 | .69 |
| Plant | 50,000 | .14 | .35 | 0 | 0 | 1 | 2.03 |
| NumFaces | 50,000 | .42 | 1.06 | 0 | 0 | 10 | 4.88 |
| *Text Content* | | | | | | | |
| TextSenti | 50,000 | 2.07 | 1.13 | 1.00 | 1.86 | 7.00 | .81 |
| TextComplexity | 50,000 | 2.67 | 1.18 | 1.00 | 2.50 | 7.00 | .34 |
| TextLength | 50,000 | 37.68 | 34.49 | 2.00 | 30 | 906 | 4.02 |
| *Brand Characteristics* | | | | | | | |
| Visibility | 50,000 | 3.15 | .45 | 1.98 | 3.16 | 3.94 | -.44 |
| Involvement | 50,000 | 3.77 | .37 | 3.09 | 3.69 | 4.32 | .14 |
| PerceivedRisk | 50,000 | 1.72 | .33 | 1.20 | 1.72 | 2.47 | .21 |
| *Brand Equity:* | | | | | | | |
| Relevance | 50,000 | 2.85 | .74 | 1.45 | 2.78 | 4.33 | .25 |
| Differentiation | 50,000 | .57 | .16 | .32 | .53 | 1.08 | .96 |
| Esteem | 50,000 | .75 | .26 | .23 | .70 | 1.43 | .38 |
| Knowledge | 50,000 | 3.90 | .68 | 1.93 | 3.98 | 4.94 | -.69 |
| *Product Characteristics* | | | | | | | |
| ExpGood | 50,000 | .70 | .46 | 0 | 1 | 1 | -.87 |
| Premium | 50,000 | .31 | .46 | 0 | 0 | 1 | .82 |
| *User Characteristics* | | | | | | | |
| NumFollowers | 50,000 | 6,593.83 | 75,703.70 | 0 | 460 | 4,125,705 | 40.97 |
| NumFollowings | 50,000 | 791.01 | 1,255.80 | 0 | 377 | 8,162 | 3.38 |
| PostCount | 50,000 | 758.13 | 1,884.12 | 0 | 285 | 114,779 | 13.32 |

*Notes.* The statistics for VizSentiment, VizComplexity, TextSentiment and TextComplexity are using the rescaled (all are on a 1-7 scale) values. For estimation, these four variables are zero centered. For estimation, NumFollowers is rescaled to 0-100 scale.

Figure 22. Scatter Plots

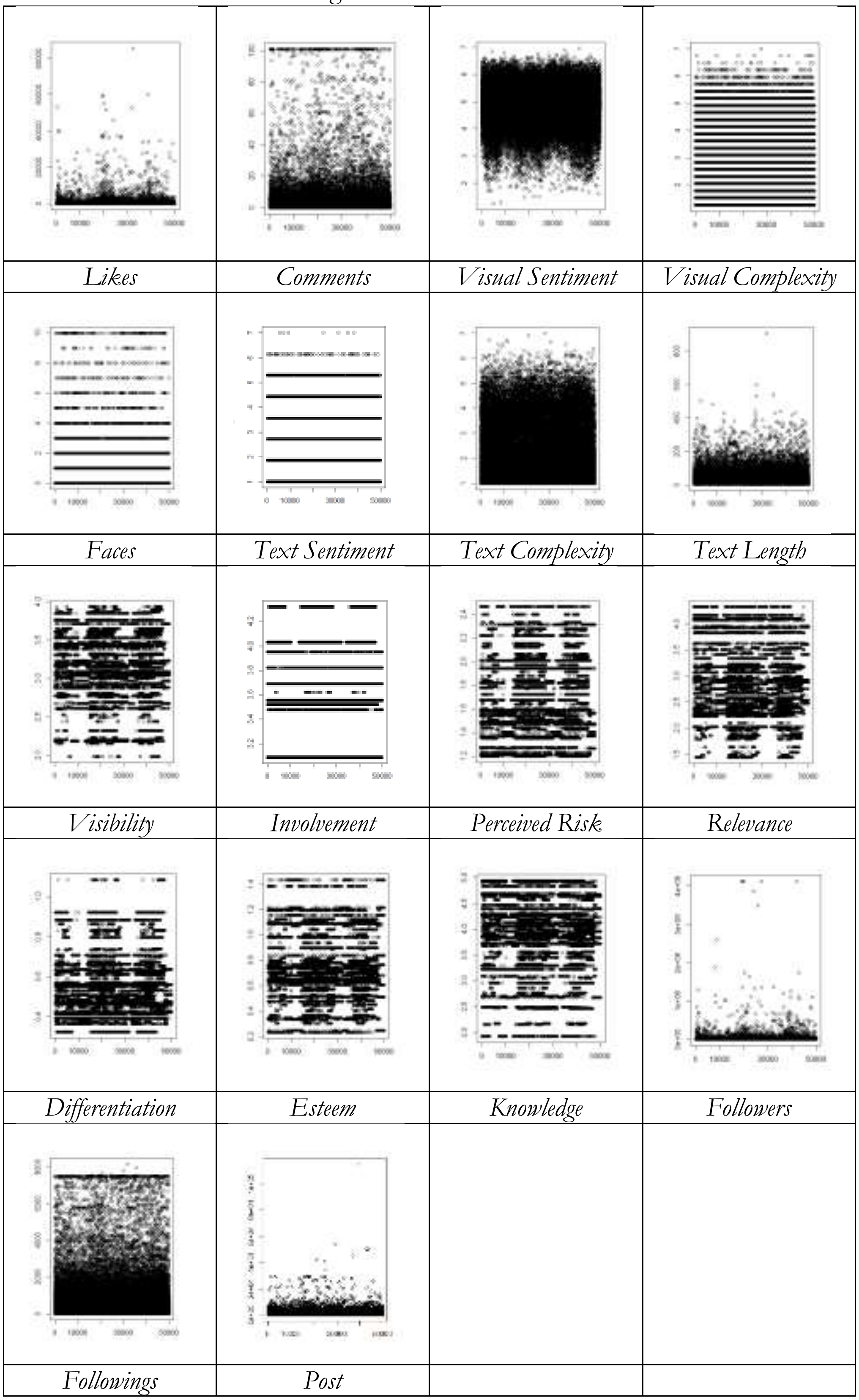


Figure 23. Density Plots

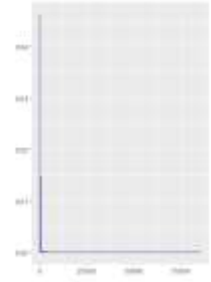
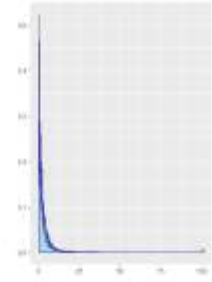
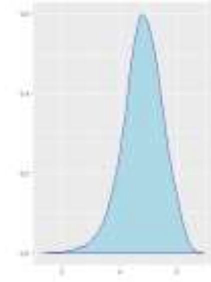
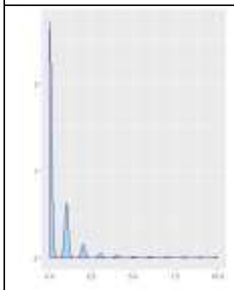
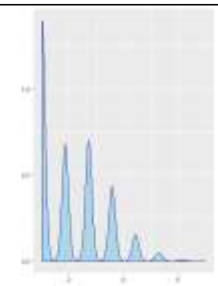
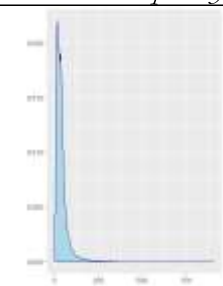
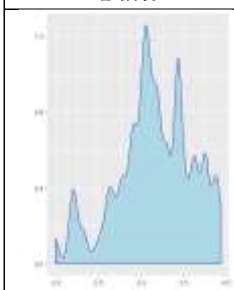
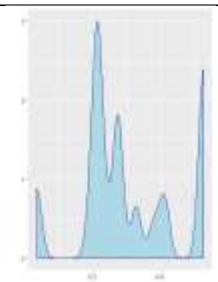
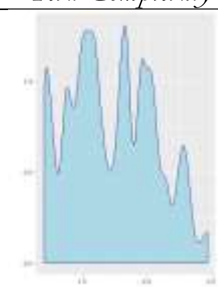
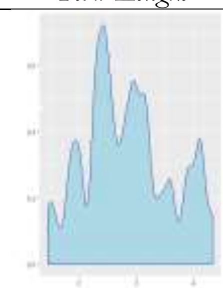
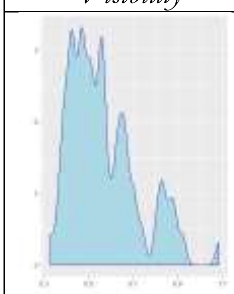
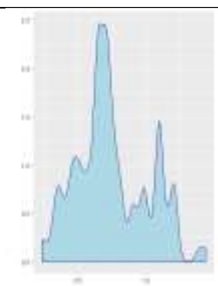
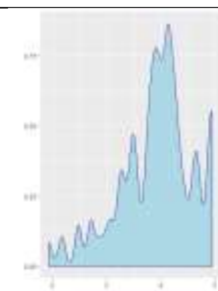
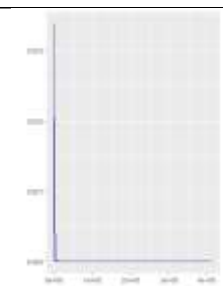
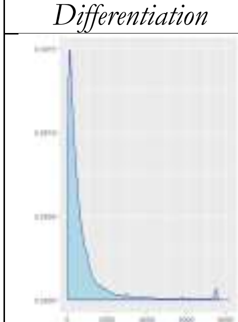
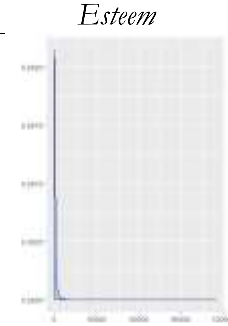
| | | | |
|---|---|---|---|
|  |  |  |  |
| *Likes* | *Comments* | *Visual Sentiment* | *Visual Complexity* |
|  |  |  |  |
| *Faces* | *Text Sentiment* | *Text Complexity* | *Text Length* |
|  |  |  |  |
| *Visibility* | *Involvement* | *Perceived Risk* | *Relevance* |
|  |  |  |  |
| *Differentiation* | *Esteem* | *Knowledge* | *Followers* |
|  |  | | |
| *Followings* | *Post* | | |

Table 7. Correlation Matrix

| | Viz Sent | Viz Complexity | Faces | Text Sent | Text Complexity | Text Length | Visibility | Involvement | Perceived risk | Relevance | Differentiation | Esteem | Knowledge | Follower | Following | Post |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VizSent | 1 | 0.1 | 0.2 | 0.03 | 0.01 | 0.03 | -0.1 | -0.2 | -0.1 | 0.1 | -0.1 | 0 | 0.01 | 0.01 | 0.03 | 0 |
| Viz Complexity | 0.1 | 1 | 0.1 | -0.03 | -0.01 | -0.05 | -0.03 | 0.1 | 0.04 | -0.01 | -0.03 | 0.02 | 0.03 | 0.02 | -0.01 | 0.01 |
| Faces | 0.2 | 0.1 | 1 | 0.01 | 0 | 0 | -0.05 | -0.1 | -0.02 | 0.02 | -0.05 | -0.02 | -0.03 | -0.01 | 0.02 | 0 |
| TextSent | 0.03 | -0.03 | 0.01 | 1 | 0.6 | 0.2 | 0 | 0 | 0 | -0.01 | -0.01 | -0.04 | -0.02 | -0.03 | 0.02 | -0.03 |
| Text Complexity | 0.01 | -0.01 | 0 | 0.6 | 1 | 0.4 | -0.02 | 0.1 | 0.1 | -0.1 | 0.05 | -0.1 | -0.1 | -0.03 | 0.1 | 0.02 |
| Text Length | 0.03 | -0.05 | 0 | 0.2 | 0.4 | 1 | -0.01 | 0.1 | 0.03 | -0.05 | 0.01 | -0.04 | -0.1 | 0.01 | 0.1 | 0.1 |
| Visibility | -0.1 | -0.03 | -0.05 | 0 | -0.02 | -0.01 | 1 | 0.1 | -0.5 | 0.5 | -0.03 | 0.5 | 0.6 | -0.02 | 0 | -0.01 |
| Involvement | -0.2 | 0.1 | -0.1 | 0 | 0.1 | 0.1 | 0.1 | 1 | 0.4 | -0.5 | 0.4 | -0.1 | -0.2 | 0.01 | 0 | 0.02 |
| Perceive risk | -0.1 | 0.04 | -0.02 | 0 | 0.1 | 0.03 | -0.5 | 0.4 | 1 | -0.8 | 0.4 | -0.5 | -0.7 | 0.03 | 0.02 | 0.04 |
| Relevance | 0.1 | -0.01 | 0.02 | -0.01 | -0.1 | -0.05 | 0.5 | -0.5 | -0.8 | 1 | -0.3 | 0.7 | 0.8 | -0.03 | -0.02 | -0.03 |
| Differentiation | -0.1 | -0.03 | -0.05 | -0.01 | 0.05 | 0.01 | -0.03 | 0.4 | 0.4 | -0.3 | 1 | 0.02 | -0.2 | 0.01 | -0.02 | 0.02 |
| Esteem | 0 | 0.02 | -0.02 | -0.04 | -0.1 | -0.04 | 0.5 | -0.1 | -0.5 | 0.7 | 0.02 | 1 | 0.6 | -0.02 | -0.02 | -0.03 |
| Knowledge | 0.01 | 0.03 | -0.03 | -0.02 | -0.1 | -0.1 | 0.6 | -0.2 | -0.7 | 0.8 | -0.2 | 0.6 | 1 | -0.02 | -0.01 | -0.03 |
| Follower | 0.01 | 0.02 | -0.01 | -0.03 | -0.03 | 0.01 | -0.02 | 0.01 | 0.03 | -0.03 | 0.01 | -0.02 | -0.02 | 1 | 0.01 | 0.1 |
| Following | 0.03 | -0.01 | 0.02 | 0.02 | 0.1 | 0.1 | 0 | 0 | 0.02 | -0.02 | -0.02 | -0.02 | -0.01 | 0.01 | 1 | 0.2 |
| Post | 0 | 0.01 | 0 | -0.03 | 0.02 | 0.1 | -0.01 | 0.02 | 0.04 | -0.03 | 0.02 | -0.03 | -0.03 | 0.1 | 0.2 | 1 |

Table 8. Marginal Distribution for Commercial Posts vs. General Posts

| Post/Variable | Visual Sentiment | Visual Complexity | Text Sentiment | Text Complexity | Visibility | Involvement |
|---|---|---|---|---|---|---|
| Commercial | 4.802 | 2.832 | 2.003 | 2.984 | 3.088 | 3.773 |
| General | 4.791 | 3.126 | 2.065 | 2.655 | 3.147 | 3.775 |
| Post/Variable | Perceived Risk | Follower | Post | Following | Face | Text Length |
| Commercial | 1.777 | 12,275.91 | 1,075.57 | 1,137.68 | 0.258 | 49.77 |
| General | 1.724 | 6,630.97 | 787.76 | 787.59 | 0.416 | 37.58 |
| Post/Variable | Relevance | Differentiation | Esteem | Knowledge | | |
| Commercial | 2.758 | 0.571 | 0.743 | 3.778 | | |
| General | 2.852 | 0.565 | 0.749 | 3.899 | | |

Table 9. Cross Tabulation between Commercial and Living

| | | Commercial | General |
|---|---|---|---|
| Non-Living | Frequency | 9,328 | 30,609 |
| | Proportion | 0.558 | 0.453 |
| Living | Frequency | 7,388 | 36,904 |
| | Proportion | 0.442 | 0.547 |

Table 10. Cross Tabulation between Commercial and Food

| | | Commercial | General |
|---|---|---|---|
| No Food | Frequency | 11,960 | 45,111 |
| | Proportion | 0.715 | 0.668 |
| Food | Frequency | 4,756 | 22,402 |
| | Proportion | 0.285 | 0.332 |

Table 11. Cross Tabulation between Commercial and Plant

| | | Commercial | General |
|---|---|---|---|
| No Plant | Frequency | 14,882 | 57,994 |
| | Proportion | 0.890 | 0.859 |
| Plant | Frequency | 1,834 | 9,519 |
| | Proportion | 0.110 | 0.141 |

Table 12. Cross Tabulation between Commercial and Premium

| | | Commercial | General |
|---|---|---|---|
| No Premium | Frequency | 9,695 | 45,452 |
| | Proportion | 0.611 | 0.694 |
| Premium | Frequency | 6,170 | 19,964 |
| | Proportion | 0.389 | 0.306 |

Table 13. Cross Tabulation between Commercial and Type of Good

| Type of good | | Commercial | General |
|---|---|---|---|
| Search or | Frequency | 6,669 | 20.949 |
| Credence | Proportion | 0.422 | 0.320 |
| Experience | Frequency | 9.166 | 44,467 |
| | Proportion | 0.578 | 0.680 |

Table 14. Marginal Distribution for a Focal Brand vs. Multi Brand

| Post/Variable | Visual Sentiment | Visual Complexity | Text Sentiment | Text Complexity | Visibility | Involvement |
|---|---|---|---|---|---|---|
| Focal | 4.785 | 3.133 | 1.926 | 2.372 | 3.169 | 3.752 |
| Multi | 4.796 | 3.116 | 2.431 | 3.417 | 3.101 | 3.831 |
| Post/Variable | Perceived Risk | Follower | Post | Following | Face | Text Length |
| Focal | 1.695 | 7,188.52 | 766.73 | 778.17 | 0.429 | 33.951 |
| Multi | 1.795 | 5,070.23 | 736.20 | 823.92 | 0.391 | 47.259 |
| Post/Variable | Relevance | Differentiation | Esteem | Knowledge | | |
| Focal | 2.907 | 0.563 | 0.762 | 3.960 | | |
| Multi | 2.694 | 0.574 | 0.700 | 3.751 | | |

Table 15. Cross Tabulation between Multi-Branded and Living

| | | Focal | Multi |
|---|---|---|---|
| Non-Living | Frequency | 18,582 | 7.629 |
| | Proportion | 0.448 | 0.458 |
| Living | Frequency | 22,878 | 9.005 |
| | Proportion | 0.552 | 0.542 |

Table 16. Cross Tabulation between Multi-Branded and Food

| | | Focal | Multi |
|---|---|---|---|
| No Food | Frequency | 26,888 | 12,067 |
| | Proportion | 0.648 | 0.725 |
| Food | Frequency | 14,572 | 4,567 |
| | Proportion | 0.352 | 0.275 |

Table 17. Cross Tabulation between Multi-Branded and Plant

| | | Focal | Multi |
|---|---|---|---|
| No Plant | Frequency | 35,522 | 14,422 |
| | Proportion | 0.856 | 0.867 |
| Plant | Frequency | 5,938 | 2,212 |
| | Proportion | 0.144 | 0.133 |

Table 18. Cross Tabulation between Multi-Branded and Premium

| | | Focal | Multi |
|---|---|---|---|
| No Premium | Frequency | 29.158 | 10,332 |
| | Proportion | 0.726 | 0.645 |
| Premium | Frequency | 10,972 | 5,664 |
| | Proportion | 0.274 | 0.355 |

Table 19. Cross Tabulation between Multi-Branded and Type of Good

| Type of good | | Focal | Multi |
|---|---|---|---|
| Search or | Frequency | 12,871 | 4,951 |
| Credence | Proportion | 0.321 | 0.309 |
| Experience | Frequency | 27,259 | 11,045 |
| | Proportion | 0.679 | 0.691 |

2.5.4.3. Empirical Strategy and Analysis Results

Here, we discuss our empirical approach. Unlike for the data set described in the previous section, we identified two different sets of posts: 1) a set of posts in which commercial and general posts are distinguished and 2) a set of posts in which posts that are tagged with multiple brands including a focal brand and those that are tagged with a focal brand are distinguished. Our analysis has two goals. First, we aimed to discover whether the results in §2.5.1 are consistent with the results described here after we incorporated the interaction effects between the new labeled variables (commercial vs. general, multiple brands vs. a focal brand) and content characteristics (visual sentiment, visual complexity, text sentiment, and text complexity). Second, we aimed to learn about the interaction effects. For example, we wanted to answer whether the effect of visual sentiment on consumer engagement was different for different types of posts.

To achieve these goals, we used the same estimation strategy that we employed in §2.5 but included the interaction effects. We estimated a negative binomial regression for *LIKES* and a zero-inflated negative binomial regression for *COMMENTS*. We built the models in four different ways for both *LIKES* and *COMMENTS*. First, we employed only the content characteristics and control variables, without any brand characteristics or interaction effects (Model 1s in Tables 20, 21, 22, and 23). Second, we incorporated the interaction effects in addition to all of the variables that were used in Model 1 (Model 2s in Tables 20, 21, 22, and 23). Third, Model 3 included all of the variables in Model 2 and brand characteristics. Finally, we estimated the models with all the content characteristics (Image Content, Text Content), brand characteristics (Brand Characteristics), brand equity (Brand Equity), interaction effects (Commercial vs. General or Multiple vs. Single) and control variables (Object Types, User Characteristics, and Product Characteristics) (Model 4s in Tables 20, 21, 22, and 23).

Overall, we estimated four different types of models (*LIKES* model with commercial interaction effect; *LIKES* model with multiple interaction effect: *COMMENT* model with commercial interaction effect; *COMMENT* model with multiple interaction effect). A third-order operationalization for complexity variables, as opposed to linear or quadratic models, was incorporated into these models since these models fit the data better than any others (e.g., a second-order model for both sentiment variables and complexity variables or a second-order model for only complexity variables). For example, in the case of the *LIKES* model, if we fit the data with second-order forms of all the visual and textual variables, the coefficient of the second-order term of visual sentiment variable is not statistically significant, so we did not need to interpret the results. Moreover, the model fit is poorer than that of the full model that we eventually employed (compare Model 4 in Table 2 with Model 2 with Table 24).

The results are presented in Tables 20, 21, 22, and 23. We find that the results are, overall, consistent with the original results, presented in Tables 3 and 4. The visual sentiment and text sentiment consistently show positive effects on consumer engagement for all models, as shown in Tables 20, 21, 22, and 23. We achieve the same results as those presented in §2.5.1: higher visual sentiment and higher text sentiment are associated with more likes. For the complexity variables, we find consistent results; the relationships between the complexity variables and consumer engagement measures (*LIKES, COMMENTS*) are bimodal, with a strong positive effect at the minimum and just past the middle of the scale, no effect for modest values, and a strong negative effect at high levels of visual complexity. Tables 20 and 22 show the estimation results for models that include different types of posts (commercial and general) for *LIKES* and *COMMENTS*, respectively (Please refer to the note for Tables 20 –23: *Note 1.* $p<.05$; *$p<.01$; **$p<.001$; ***$p<.0001$, *Note 2*. NumFollowers re-scaled to the range 0-100).

Table 20. Interaction Model with Commercial-Related Posts: *LIKES*

| **LIKES** | Model1 | Model2 | Model3 | Model4 |
|---|---|---|---|---|
| Intercept | 3.638***(105.34) | 3.678***(98.15) | 1.501***(18.55) | 2.231***(20.99) |
| *Image Content* | | | | |
| VizSentiment | 0.059 (0.92) | 0.025***(3.48) | 0.059***(8.35) | 0.065***(9.19) |
| VizSentiment^2 | | | | |
| VizComplexity | 0.125***(15.18) | 0.081***(9.82) | 0.046***(5.59) | 0.055***(6.66) |
| VizlComplexity^2 | 0.029***(8.45) | 0.030***(8.56) | 0.041***(11.61) | 0.031***(9.01) |
| VizComplexity^3 | -0.019***(-8.17) | -0.017***(-7.47) | -0.019***(-8.41) | -0.016***(-7.22) |
| *Object Types:* | | | | |
| Living | 0.005 (0.44) | 0.003 (0.24) | 0.061***(5.55) | 0.057***(5.26) |
| Food | -0.287***(-28.14) | -0.281***(-27.91) | -0.195***(-18.73) | -0.179***(-17.21) |
| Plant | 0.126***(9.59) | 0.125***(9.54) | 0.091***(7.03) | 0.105***(8.07) |
| NumFaces | 0.038***(8.34) | 0.036***(7.94) | 0.052***(11.49) | 0.049***(10.73) |
| *Text Content* | | | | |
| TextSentiment | 0.079***(16.79) | 0.036***(6.77) | 0.048***(9.04) | 0.047***(8.96) |
| TextSentiment^2 | | | | |
| TextComplexity | 0.286***(37.29) | 0.351***(44.08) | 0.316***(39.59) | 0.309***(38.91) |
| TextComplexity^2 | 0.246***(61.81) | 0.247***(62.53) | 0.244***(61.91) | 0.237***(60.46) |
| TextComplexity^3 | -0.109***(-46.92) | -0.111***(-47.85) | -0.104***(-45.44) | -0.103***(-44.95) |
| TextLength | 0.002***(14.47) | 0.022***(17.07) | 0.002***(17.36) | 0.002***(17.79) |
| *Brand Characteristics* | | | | |
| Visibility | | | 0.037**(2.79) | 0.100***(7.17) |
| Involvement | | | 0.407***(23.82) | 0.283***(14.78) |
| PerceivedRisk | | | 0.209***(9.38) | -0.104***(-3.91) |
| *Brand Equity:* | | | | |
| Relevance | | | | -0.191***(-13.98) |
| Differentiation | | | | 0.601***(17.42) |
| Esteem | | | | -0.357***(-12.16) |
| Knowledge | | | | 0.139***(12.11) |
| *User Characteristics* | | | | |
| NumFollowers[1] | 1.666***(690.55) | 1.674***(699.73) | 1.630***(686.00) | 1.612***(681.84) |
| NumFollowing | 0.00009***(25.74) | 0.00009**(27.47) | 0.00009***(26.93) | 0.00009***(27.08) |
| PostCount | -0.00004***(-17.90) | -0.00004***(-18.17) | -0.00004***(-18.93) | -0.00004***(-18.36) |
| *Product Characteristics* | | | | |
| ExpGood | -0.04***(-3.74) | -0.042***(-4.46) | -0.123***(-11.39) | -0.167***(-14.32) |
| Premium | -0.069***(-7.27) | -0.048***(-5.13) | -0.184***(-16.58) | -0.134***(-10.15) |
| *Commercial vs. General* | | | | |
| General | | 0.011 (0.13) | 0.013 (0.16) | 0.039 (0.49) |
| General*VizSenment | | 0.125***(7.78) | 0.126***(7.85) | 0.124***(7.81) |
| General*VizComplexity | | -0.169***(-18.39) | -0.170**(-18.57) | -0.151***(-17.61) |
| General*TextSentiment | | -0.147***(-12.71) | -0.145***(-12.62) | -0.135***(-11.76) |
| General*TextComplexity | | 0.252***(24.29) | 0.247***(23.98) | 0.241***(23.55) |
| AIC | 788,323 | 786,569 | 785,232 | 784,210 |
| N | 72,194 | 72,194 | 72,194 | 72,194 |

Table 21. Interaction Model with Multi-Branded Posts: *LIKES*

| **LIKES** | Model1 | Model2 | Model3 | Model4 |
|---|---|---|---|---|
| Intercept | 3.653***(91.08) | 3.643***(88.81) | 1.628***(17.11) | 2.513***(20.01) |
| *Image Content* | | | | |
| VizSentiment | 0.023**(3.12) | 0.034***(3.87) | 0.065***(7.42) | 0.073***(8.29) |
| VizSentiment^2 | | | | |
| VizComplexity | 0.093***(9.77) | 0.092***(9.35) | 0.059***(5.93) | 0.067***(6.76) |
| VizlComplexity^2 | 0.044***(10.56) | 0.044***(10.58) | 0.056***(13.41) | 0.047***(11.31) |
| VizComplexity^3 | -0.022***(-8.21) | -0.022***(-8.12) | -0.024***(-8.87) | -0.021***(-8.01) |
| *Object Types:* | | | | |
| Living | 0.023.(1.81) | 0.025.(1.94) | 0.083**(6.47) | 0.077***(6.04) |
| Food | -0.289***(-24.52) | -0.290***(-24.61) | -0.205***(-16.78) | 0.186***(-15.22) |
| Plant | 0.091***(6.03) | 0.089***(5.92) | 0.060***(4.02) | 0.069***(4.61) |
| NumFaces | 0.027***(5.24) | 0.026***(5.07) | 0.042****(8.28) | 0.038***(7.53) |
| *Text Content* | | | | |
| TextSentiment | 0.046***(8.39) | 0.030***(4.41) | 0.039***(5.72) | 0.042***(6.11) |
| TextSentiment^2 | | | | |
| TextComplexity | 0.347***(38.22) | 0.373***(36.49) | 0.347***(34.02) | 0.338***(33.26) |
| TextComplexity^2 | 0.249***(53.74) | 0.253***(50.26) | 0.252***(50.18) | 0.244***(48.92) |
| TextComplexity^3 | -0.112***(-41.10) | -0.113***(-41.10) | -0.108***(-39.35) | -0.107***(-39.13) |
| TextLength | 0.002***(13.91) | 0.002***(13.98) | 0.002***(13.96) | 0.002***(14.64) |
| *Brand Characteristics* | | | | |
| Visibility | | | 0.041**(2.67) | 0.105***(6.41) |
| Involvement | | | 0.359***(18.06) | 0.240***(10.94) |
| PerceivedRisk | | | 0.218***(8.23) | -0.117***(-3.61) |
| *Brand Equity:* | | | | |
| Relevance | | | | -0.223***(-13.91) |
| Differentiation | | | | 0.498***(12.35) |
| Esteem | | | | -0.236***(-6.72) |
| Knowledge | | | | 0.116***(8.45) |
| *User Characteristics* | | | | |
| NumFollowers[1] | 1.808***(627.86) | 1.810***(628.63) | 1.770***(618.54) | 1.760***(618.13) |
| NumFollowing | 0.0001***(25.71) | 0.0001***(25.68) | 0.0001***(25.64) | 0.0001***(26.27) |
| PostCount | -0.00003***(-9.14) | -0.00003***(-9.11) | -0.00003***(-10.07) | -0.00003***(-9.78) |
| *Product Characteristics* | | | | |
| ExpGood | -0.111***(-9.87) | -0.109***(-9.72) | -0.179***(-14.02) | -0.221***(-15.90) |
| Premium | -0.050***(-4.32) | -0.046***(-4.07) | -0.179***(-13.67) | -0.138***(-8.82) |
| *Multiple vs. Focal* | | | | |
| Multiple | | 0.074 (0.91) | 0.009 (0.11) | 0.057 (0.70) |
| Multiple*VizSentiment | | -0.039*(-2.47) | -0.032*(-1.99) | -0.036* (-2.22) |
| Multiple*VizComplexity | | 0.002 (0.16) | -0.003 (-0.26) | -0.0006 (-0.06) |
| Multiple*TextSentment | | 0.047***(4.07) | 0.051***(4.41) | 0.043***(3.72) |
| Multiple*TextComplexity | | -0.060***(-4.49) | -0.065***(-4.87) | -0.061***(-4.59) |
| AIC | 547,869 | 547,815 | 546,957 | 546,327 |
| N | 50,000 | 50,000 | 50,000 | 50,000 |

Table 22. Interaction Model with Commercial-Related Posts: *COMMENTS*

| **COMMENTS** | Model1 | Model2 | Model3 | Model4 |
|---|---|---|---|---|
| Intercept | -0.187***(-3.99) | -0.329**(-3.11) | -0.836***(-5.71) | -0.898***(-5.01) |
| *Image Content* | | | | |
| VizSentiment | 0.101***(11.53) | 0.048*(2.30) | 0.058**(2.73) | 0.061**(2.89) |
| VizSentiment^2 | | | | |
| VizComplexity | 0.024*(2.14) | 0.125***(8.01) | 0.119***(7.62) | 0.122***(7.74) |
| VizlComplexity^2 | 0.007 (1.51) | 0.009.(1.87) | 0.011*(2.16) | 0.006 (1.30) |
| VizComplexity^3 | -0.008*(-2.48) | -0.007*(-2.33) | -0.008*(-2.51) | -0.007*(-2.09) |
| *Object Types:* | | | | |
| Living | 0.099***(6.44) | 0.094***(6.13) | 0.111***(7.08) | 0.109***(6.97) |
| Food | -0.083***(-5.95) | -0.079***(-5.71) | -0.068***(-4.69) | -0.072***(-4.92) |
| Plant | 0.045*(2.51) | 0.041*(2.29) | 0.034.(1.88) | 0.037*(2.05) |
| NumFaces | 0.041***(5.83) | 0.036***(5.17) | 0.039***(5.69) | 0.039***(.5.56) |
| *Text Content* | | | | |
| TextSentiment | 0.119***(17.97) | 0.198***(13.00) | 0.203***(13.25) | 0.197***(12.66) |
| TextSentiment^2 | | | | |
| TextComplexity | 0.025*(2.39) | -0.128***(-8.06) | -0.134***(-8.36) | -0.137***(-8.55) |
| TextComplexity^2 | 0.214***(42.59) | 0.217***(42.47) | 0.217***(42,45) | 0.214***(40.94) |
| TextComplexity^3 | -0.073***(-24.05) | -0.073***(-24.49) | -0.072***(-24.14) | -0.070***(-23.30) |
| TextLength | 0.005***(26.19) | 0.006***(27.83) | 0.006*(27.63) | 0.004***(25.86) |
| *Brand Characteristics* | | | | |
| Visibility | | | 0.021 (1.14) | 0.062**(3.15) |
| Involvement | | | 0.114***(4.68) | 0.118***(4.29) |
| PerceivedRisk | | | -0.021 (-0.66) | -0.172***(-4.48) |
| *Brand Equity:* | | | | |
| Relevance | | | | 0.057**(2.85) |
| Differentiation | | | | 0.452***(9.29) |
| Esteem | | | | -0.495***(-11.71) |
| Knowledge | | | | 0.038*(2.39) |
| *User Characteristics* | | | | |
| NumFollowers[1] | 0.660***(47.48) | 0.668***(48.54) | 0.662***(48.18) | 0.631**(48.42) |
| NumFollowing | 0.00005***(11.17) | 0.00006***(12.10) | 0.00006***(12.08) | 0.00006***(11.65) |
| PostCount | -0.00007***(-11.15) | -0.00007***(-10.90) | -0.00006***(-10.94) | -0.00006***(-9.93) |
| *Product Characteristics* | | | | |
| ExpGood | -0.097***(-7.51) | -0.112***(-8.71) | -0.139***(-9.17) | -0.159***(-9.72) |
| Premium | -0.066***(-5.07) | -0.047***(-3.67) | -0.051**(-3.27) | 0.010 (0.55) |
| *Commercial vs. General* | | | | |
| General | | 0.226.(1.96) | 0.228*(1.98) | 0.225.(1.96) |
| General*VizSentiment | | 0.061**(2.68) | 0.062**(2.69) | 0.058*92.51) |
| General*VizComplexity | | -0.134***(-10.36) | -0.134***(-10.35) | -0.132***(-10.16) |
| General*TextSentiment | | -0.119***(-7.04) | -0.119***(-7.04) | -0.115***(-6.69) |
| General*TextComplexity | | 0.206***(13.89) | 0.202***(13.96) | 0.199***(13.57) |
| AIC | 306,461 | 305,603 | 305,561 | 305,368 |
| N | 72,194 | 72,194 | 72,194 | 72,194 |

Table 23. Interaction Model with Multi-Branded Posts: *COMMENTS*

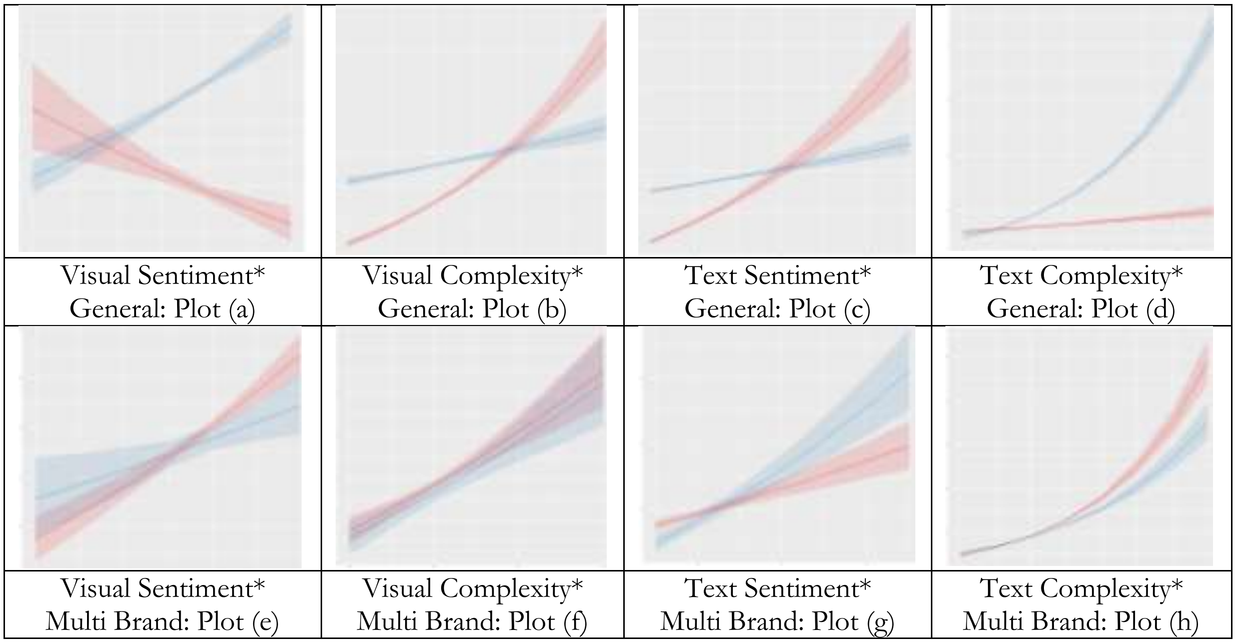
| **COMMENTS** | Model1 | Model2 | Model3 | Model4 |
|---|---|---|---|---|
| Intercept | -0.090 (-1.64) | -0.115.(-1.79) | -0.258.(-1.95) | -0.290.(-1.65) |
| *Image Content* | | | | |
| VizSentiment | 0.105***(10.35) | 0.114***(9.78) | 0.178 (0.96) | 0.116***(9.68) |
| VizSentiment^2 | | | | |
| VizComplexity | -0.001 (-0.09) | 0.008 (0.56) | 0.007 (0.51) | 0.011 (0.82) |
| VizlComplexity^2 | 0.015**(2.68) | 0.015**(2.66) | 0.016**(2.78) | 0.013*(2.25) |
| VizComplexity^3 | -0.008*(-2.51) | -0.009*(-2.40) | -0.009*(-2.53) | -0.008*(-2.29) |
| *Object Types:* | | | | |
| Living | 0.101***(5.66) | 0.101***(5.64) | 0.106***(5.78) | 0.102***(5.59) |
| Food | -0.079***(-4.89) | -0.083***(-5.12) | -0.081***(-4.75) | -0.086 (0.71) |
| Plant | 0.018 (0.88) | 0.015 (0.77) | 0.013 (0.64) | 0.014 (0.71) |
| NumFaces | 0.029***(3.73) | 0.028***(3.59) | 0.029***(3.84) | 0.029***(3.71) |
| *Text Content* | | | | |
| TextSentiment | 0.088***(11.42) | 0.076***(7.92) | 0.078***(8.10) | 0.078***(8.08) |
| TextSentiment^2 | | | | |
| TextComplexity | 0.069***(5.52) | 0.099***(7.05) | 0.099***(7.00) | 0.099***(6.98) |
| TextComplexity^2 | 0.208***(34.48) | 0.215***(32.32) | 0.217***(32.43) | 0.215***(32.17) |
| TextComplexity^3 | -0.074***(-20.89) | -0.076***(-21.20) | -0.076***(-21.14) | -0.076***(-21.17) |
| TextLength | 0.006***(25.23) | 0.006***(25.21) | 0.006***(25.09) | 0.006***(24.72) |
| *Brand Characteristics* | | | | |
| Visibility | | | 0.009 (0.40) | 0.037 (1.63) |
| Involvement | | | 0.029 (1.03) | 0.036 (1.16) |
| PerceivedRisk | | | -0.005 (-0.14) | -0.119**(-2.61) |
| *Brand Equity:* | | | | |
| Relevance | | | | 0.058*(2.51) |
| Differentiation | | | | 0.367***(6.45) |
| Esteem | | | | -0.364***(-7.27) |
| Knowledge | | | | 0.005 (0.26) |
| *User Characteristics* | | | | |
| NumFollowers[1] | 0.664***(36.79) | 0.663***(36.61) | 0.661***(36.85) | 0.654***(36.27) |
| NumFollowing | 0.00008***(13.20) | 0.00008***(13.10) | 0.00007***(13.08) | 0.00008***(13.25) |
| PostCount | -0.00005***(-6.92) | -0.00005***(-6.85) | -0.00005***(-6.92) | -0.00005***(-6.75) |
| *Product Characteristics* | | | | |
| ExpGood | -0.179***(-11.57) | -0.175***(-11.27) | -0.181***(-10.07) | -0.183***(9.41) |
| Premium | -0.004*(-2.54) | -0.035*(-2.26) | -0.036*(-1.98) | -0.0007 (-0.03) |
| *Multiple vs. Focal* | | | | |
| Multiple | | 0.106 (0.95) | 0.107 (0.96) | 0.094 (0.84) |
| Multiple*VizSentiment | | -0.004.(-1.67) | -0.037.(-1.67) | -0.034 (-1.54) |
| Multiple*VizComplexity | | -0.003*(-2.39) | -0.031*(-2.40) | -0.029*(-2.27) |
| Multiple*TextSentiment | | 0.032*(1.99) | 0.031.(1.96) | 0.029.(1.83) |
| Multiple*TextComplexity | | -0.076***(-3.89) | -0.077***(-3.90) | -0.076***(-3.85) |
| AIC | 214,880 | 214,846 | 214,835 | 214,753 |
| N | 50,000 | 50,000 | 50,000 | 50,000 |

Most of the interaction effects in the *LIKES* and *COMMENTS* models are significant. This means that the effects of the content characteristics differ for commercial and general posts and for the posts with multiple brands and a focal brand mentioned in hashtags. Figure 24 displays the effects for the *LIKES* model. Regarding the interaction effect between visual sentiment and general vs. commercial posts, we find completely opposite results for the two types of posts. Specifically, we find that, as visual sentiment increases, the effect decreases for commercial posts. On the other hand, the interaction effect for general posts is consistent with that in the main models; there is a positive relation between visual sentiment and consumer engagement for general posts (see Plot (a) in Figure 24). We find that, for commercial posts, as visual complexity increases, so does consumer engagement. A similar pattern is observed for the general posts, although the effect is not nearly as strong (Plot (b)).

There is a similar upward trend concerning the interaction effect between text sentiment and general vs. commercial posts (Plot (c)). Regarding the effect of text complexity, we find a huge difference between general posts and commercial posts, although they both show increasing trends (Plot (d)). Specifically, as text complexity increases, so does the level of consumer engagement with general posts. The effect has a slightly upward trend compared to commercial posts. This means that consumers more sensitively respond to text diversity in posts that are generated by general Instagram users but are not sensitive to commercial posts.

Plots (f). (g), and (h) in Figure 24 present the interaction effect between content characteristics and posts with multiple brands and those with single brands in their hashtags in the full *LIKES* model (Model 4 in Table 20). We can visualize the finding that, for posts mentioning a focal brand in their hashtags, as visual sentiment increases, so does the number of likes. We see a similar pattern for posts mentioning multiple brands, though the effect is not nearly as strong. For visual complexity, the effects are not different for posts with a focal brand or multiple brands in their hashtags. The effect of text sentiment is stronger for posts mentioning multiple brands, while the effect of text complexity is stronger for those mentioning a focal brand.

Figure 24. Interaction Effect for *LIKES*




*Notes 1.* From Plot (a) to Plot (d), the red line represents commercial posts and the blue line general posts; from Plot (e) to Plot (h), the red line represents single branded posts and blue line multiple branded posts.

*Notes 2.* Plot (a): range of x-axis: 1-7 / range of y-axis: 50-110; Plot (b): range of x-axis: -2-4 / range of y-axis: 40-170; Plot (c): range of x-axis: 1-7 / range of y-axis: 50-250; Plot (d): range of x-axis: -2-5 / range of y-axis: 50-480; Plot (e): range of x-axis: 1-7 / range of y-axis: 60-110; Plot (f): range of x-axis: -2-4 / range of y-axis: 70-125; Plot (g): range of x-axis: 1-7 / range of y-axis: 75-145; Plot (h): range of x-axis: -2-5 / range of y-axis: 50-550

2.5.4.4. Robustness Checks

Several robustness checks were conducted only for the *LIKES* models (Tables 24 and 25). We first consider first-order and second-order forms of the four focal variables (visual sentiment, visual complexity, text sentiment, and text complexity). Tables 24 and 25 present the results, which indicate that the substantive conclusions for the linear and quadratic forms (Linear (Model1) and Quadratic (Model2) in Tables 24 and 25) are similar to those for the cubed model (Cube (Model3) in Tables 24 and 25) (Please refer to the note for Tables 20 –23: *Note 1.* $p<.05$; *$p<.01$; **$p<.001$; ***$p<.0001$, *Note*

*2*. NumFollowers re-scaled to the range 0-100).

Second, we examine the same *LIKES* models with a hold-out sample. We split the data into a training set (70%) and a test set (30%) for each the two types of data and then perform prediction. Tables 26, 27, 28, and 29 report the prediction results using the root mean square error (RMSE) and mean absolute percentage error (MAPE) for the *LIKES* models. We conduct prediction in two different ways. Tables 26 and 27 display the RMSE and MAPE of Model 1s to Model 4s in Tables 20 and 21 respectively. Tables 28 and 29 present the RMSE and MAPE for the linear, quadratic, and cube models in Tables 24 and 25, respectively. The MAPE indicates that the best model is Model 4 with third-order operationalization and complete sets of features for both Set 1 and Set 2, while RMSE barely differentiates between the models. For the results regarding the linear, quadratic and cube models (Tables 28 and 29), the cubed model shows the best prediction accuracy in terms of MAPE with Set 2, while the linear model is best with Set 1, contradicting the main results presented in Table 24.

Table 24. Robustness Checks: Accounting for a Commercial Post

| **LIKES** | Linear (Model1) | Quadratic (Model2) | Cube (Model3) |
|---|---|---|---|
| Intercept | 2.639***(24.46) | 1.950***(11.75) | 2.231***(20.99) |
| *Image Content* | | | |
| VizSentiment | 0.085***(11.78) | 0.130*(2.39) | 0.065***(9.19) |
| VizSentiment^2 | | -0.006 (-0.95) | |
| VizComplexity | 0.051 (1.02) | 0.006 (1.16) | 0.055***(6.66) |
| VizlComplexity^2 | | 0.021***(7.06) | 0.031***(9.01) |
| VizComplexity^3 | | | -0.016***(-7.22) |
| *Object Types:* | | | |
| Living | 0.014 (1.36) | 0.054***(4.89) | 0.057***(5.26) |
| Food | -0.221***(-21.15) | -0.194***(-18.54) | -0.179***(-17.21) |
| Plant | 0.106***(7.99) | 0.104***(7.98) | 0.105***(8.07) |
| NumFaces | 0.042***(9.12) | 0.047***(10.29) | 0.049***(10.73) |
| *Text Content* | | | |
| TextSentiment | 0.035***(6.62) | -0.201***(-11.22) | 0.047***(8.96) |
| TextSentiment^2 | | 0.046***(14.25) | |
| TextComplexity | 0.055***(10.21) | 0.050***(9.04) | 0.309***(38.91) |
| TextComplexity^2 | | 0.138***(42.56) | 0.237***(60.46) |
| TextComplexity^3 | | | -0.103***(-44.95) |
| TextLength | 0.003***(20.34) | 0.003***(20.22) | 0.002***(17.79) |
| *Brand Characteristics* | | | |
| Visibility | 0.132***(9.23) | 0.123***(8.71) | 0.100***(7.17) |
| Involvement | 0.265***(13.59) | 0.327***(16.94) | 0.283***(14.78) |
| PerceivedRisk | -0.110***(-4.03) | -0.032 (-1.20) | -0.104***(-3.91) |
| *Brand Equity:* | | | |
| Relevance | -0.201***(-15.05) | -0.184***(-13.36) | -0.191***(-13.98) |
| Differentiation | 0.687***(19.48) | 0.628***(18.04) | 0.601***(17.42) |
| Esteem | -0.423***(-14.13) | -0.388***(-13.11) | -0.357***(-12.16) |
| Knowledge | 0.129***(11.01) | 0.164***(14.14) | 0.139***(12.11) |
| *User Characteristics* | | | |
| NumFollowers[1] | 1.721***(712.61) | 1.648***(690.53) | 1.612***(681.84) |
| NumFollowing | 0.00008***(24.27) | 0.00009****(26.68) | 0.00009***(27.08) |
| PostCount | -0.00005***(-21.67) | -0.00004****(-19.55) | -0.00004***(-18.36) |
| *Product Characteristics* | | | |
| ExpGood | -0.197***(-16.49) | -0.178***(-15.05) | -0.167***(-14.32) |
| Premium | -0.139***(-10.36) | -0.129***(-0.75) | -0.134***(-10.15) |
| *Commercial vs. General* | | | |
| General | 0.102 (1.26) | 0.056***(0.69) | 0.039 (0.49) |
| General*VizSentiment | 0.128***(7.86) | 0.128***(7.95) | 0.124***(-7.81) |
| General*VizComplexity | -0.184***(-19.81) | -0.158***(-17.16) | -0.151***(-17.61) |
| General*TextSentiment | -0.121***(-10.33) | -0.142***(-12.29) | -0.135***(-11.76) |
| General*TextComplexity | 0.217***(20.78) | 0.246***(23.76) | 0.241***(23.55) |
| AIC | 788,504 | 786,107 | 784,210 |
| N | 72,194 | 72,194 | 72,194 |

Table 25. Robustness Checks: Accounting for Multiple Brands in a Post

| **LIKES** | Linear (Model1) | Quadratic (Model2) | Cube (Model3) |
|---|---|---|---|
| Intercept | 3.031***(23.78) | 2.419***(12.42) | 2.231***(20.99) |
| *Image Content* | | | 2.513***(20.01) |
| VizSentiment | 0.038***(10.49) | 0.059 (0.93) | 0.073***(8.29) |
| VizSentiment^2 | | 0.003 (0.51) | |
| VizComplexity | -0.003 (-0.49) | -0.005 (-0.84) | 0.067***(6.76) |
| VizlComplexity^2 | | 0.031***(8.74) | 0.047***(11.31) |
| VizComplexity^3 | | | -0.021***(-8.01) |
| *Object Types:* | | | |
| Living | 0.0244.(1.96) | 0.079***(6.11) | 0.077***(6.04) |
| Food | -0.231***(-18.91) | -0.199***(-16.16) | 0.186***(-15.22) |
| Plant | 0.068***(4.46) | 0.069***(4.53) | 0.069***(4.61) |
| NumFaces | 0.032***(6.07) | 0.038***(7.31) | 0.038***(7.53) |
| *Text Content* | | | |
| TextSentiment | 0.022**(3.23) | -0.262***(-12.54) | 0.042***(6.11) |
| TextSentiment^2 | | 0.057***(15.01) | |
| TextComplexity | 0.036***(5.03) | 0.052***(7.29) | 0.338***(33.26) |
| TextComplexity^2 | | 0.134***(32.00) | 0.244***(48.92) |
| TextComplexity^3 | | | -0.107***(-39.13) |
| TextLength | 0.003***(16.70) | 0.003***(16.96) | 0.002***(14.64) |
| *Brand Characteristics* | | | |
| Visibility | 0.134***(8.06) | 0.125***(7.56) | 0.105***(6.41) |
| Involvement | 0.212***(9.48) | 0.285***(12.87) | 0.240***(10.94) |
| PerceivedRisk | -0.137***(-4.12) | -0.038 (-1.14) | -0.117***(-3.61) |
| *Brand Equity:* | | | |
| Relevance | -0.244***(-14.86) | -0.214***(-13.21) | -0.223***(-13.91) |
| Differentiation | 0.572***(13.89) | 0.510***(12.52) | 0.498***(12.35) |
| Esteem | -0.309***(-8.63) | -0.263***(-7.40) | -0.236***(-6.72) |
| Knowledge | 0.111***(7.91) | 0.148***(10.72) | 0.116***(8.45) |
| *User Characteristics* | | | |
| NumFollowers[1] | 1.881***(646.31) | 1.799***(625.41) | 1.760***(618.13) |
| NumFollowing | 0.0001***(24.19) | 0.0001***(26.36) | 0.0001***(26.27) |
| PostCount | -0.00004***(-12.99) | -0.00003***(-10.56) | -0.00003***(-9.78) |
| *Product Characteristics* | | | |
| ExpGood | -0.257***(-18.08) | -0.228***(-16.24) | -0.221***(-15.90) |
| Premium | -0.140***(-8.78) | -0.130***(-8.24) | -0.138***(-8.82) |
| *Multiple vs. Focal* | | | |
| Multiple | -0.060 (-0.73) | 0.111 (1.36) | 0.057 (0.70) |
| Multiple*VizSentiment | -0.055***(-3.35) | -0.050**(-3.09) | -0.036* (-2.22) |
| Multiple*VizComplexity | 0.0029**(3.05) | 0.015 (1.59) | -0.0006 (-0.06) |
| Multiple*TextSentiment | 0.065***(5.51) | 0.040***(3.44) | 0.043***(3.72) |
| Multiple*TextComplexity | 0.173***(0.013) | 0.024.(1.83) | -0.061***(-4.59) |
| AIC | 549,269 | 547,696 | 546,327 |
| N | 50,000 | 50,000 | 50,000 |

Table 26. Prediction Based on Table 20

| | Model1 | Model2 | Model3 | Model4 |
|---|---|---|---|---|
| RMSE | 1428.54 | 1428.54 | 1428.57 | 1428.59 |
| MAPE | 267.26 | 264.40 | 263.01 | 260.79 |

Table 27. Prediction Based on Table 21

| | Model1 | Model2 | Model3 | Model4 |
|---|---|---|---|---|
| RMSE | 1379.44 | 1379.44 | 1379.46 | 1379.48 |
| MAPE | 203.28 | 203.31 | 202.04 | 200.79 |

Table 28: Prediction Based on Table 24

| | Linear | Quadratic | Cube |
|---|---|---|---|
| RMSE | 1428.50 | 1428.56 | 1428.59 |
| MAPE | 255.15 | 259.04 | 260.79 |

Table 29: Prediction Based on Table 25

| | Linear | Quadratic | Cube |
|---|---|---|---|
| RMSE | 1379.38 | 1379.45 | 1379.48 |
| MAPE | 201.73 | 201.31 | 200.79 |

## 2.6. Managerial Implications and Conclusions

The results from this study support that consumer engagement with brand-themed image-based user posts (Instagram) is affected by the visual and text characteristics of the post, characteristics of the focal brand, and characteristics of the user, and for the most part, in a manner we predicted. Some brand-themed posts are made because of a user's desire to express their relationship with a brand. Others may be the result of efforts by marketers that include experiential marketing, collaborating with general users, and influencer campaigns.

We find that more positive visual sentiment is associated with higher consumer engagement. Visual sentiment, which to date has received limited attention by marketing researchers, is critical to understanding consumer engagement with brand-themed posts in social media. A positive mood is associated with more positive evaluations of products (Schwarz and Clore 1983; Gorn et al. 1993). Feelings about advertisements are associated with consumer preferences for products, and positive moods nearly always drive consumer preferences over negative or neutral moods (Cho and Schwarz 2006). We find a similar effect in the context of engagement with a brand-themed post.

We find that visual complexity, operationalized as the number of objects in an image, positively affects consumer engagement at its low or moderate levels, with an optimal point that drives the most consumer engagement somewhere in the middle of the scale, then quickly becomes negative in its effect after a certain threshold. Visual complexity has been frequently studied in the contexts of advertising and web interface design. Several of those studies suggest that an optimal level of visual complexity drives the most positive responses (Geissler et al. 2006; Reinecke et al. 2013). We find a similar relationship. Consumers are engaged the most around the midpoint of visual complexity and become distracted by too much visual information.

Consistent with popular discussions, we find that including living objects such as people or pets in user-generated posts earns higher engagement.

Emotionally charged text content in user-generated brand-themed posts is more likely to drive consumer engagement compared with neutral content. We find that text sentiment, operationalized as the total amount of sentiment (positive or negative), is positively related to consumer engagement with a user-generated brand-themed image-based post. Emotionally charged text may activate cognitive attention or the arousal effect shown to impact viral behaviors (Berger and Milkman 2012; Pfitzner et al. 2012).

We find a strong effect for text complexity on consumer engagement with a user-generated brand-themed post. Similar to the effect of visual complexity, we find that simple text (low text complexity) is positively associated with consumer engagement, there is a second peak around the middle of the scale, and a strong negative effect for too much text information. Information overload has been widely studied in consumer contexts, and numerous studies have found a negative relationship between the amount of information and responses (e.g., Malhotra et al. 1982; Dolinsky and Feinberg 1986; Chen et al. 2008; Townsend and Kahn 2014) mainly due to an individual's limited capability to process information (Newell and Simon 1972; Payne 1976).

Including *#brandname* in a social media post is an emerging way that consumers engage with brands. We find that brands can benefit from encouraging this form of expression by users. We find that more visible brands and higher involvement brands earn more consumer engagement with user image-based posts on social media. These findings are consistent with Sprott et al. (2009) who find that brand visibility increases consumers' attitudes towards products and Lovett et al. (2013) who reveal that brand visibility as a component of social drivers stimulates word-of-mouth. Different from Lovett et al. (2013), we find a positive relationship between brand involvement and the response variable (consumer engagement), which is consistent with their expected hypothesis (i.e., a positive relationship between brand involvement and word-of-mouth).

This article makes several contributions to the extant marketing and related literatures. First, we discover that visual sentiment affects consumer response, which has seen only limited attention in marketing literature. And, we find this using large-scale observational data. Prior studies characterized visual sentiment through only a few aspects of images (e.g., facial expressions), and the analyses typically examined a small number of observations in a laboratory setting. Employing a technique that has been developed recently in the computer science literature (Deep CNNs), we can extract visual sentiment from large-scale, real-world data and connected it with a critical social media metric (consumer engagement), which firms are increasingly incorporating into their social media strategies.

Second, this article empirically accounts for visual complexity, text sentiment based on emotional divergence, and text complexity. All have received relatively little attention in marketing. Our results with large-scale field data illustrate an interesting *S*-shaped relationship that implies the existence of an optimal point around the midpoint of our visual complexity scale to drive the most consumer engagement and a threshold where consumer engagement rapidly decreases, presumably due to information overload on visual complexity and text complexity. We also find a positive relationship between consumer engagement and higher levels of text sentiment that may be the result of more cognitive attention or higher emotional arousal.

Third, to the best of our knowledge, this is the first study to investigate how brand characteristics affect consumer-engagement behaviors to user-generated social media posts. A number of studies have linked brand characteristics and various consumer behaviors, but little attention has been paid to the effect of brand properties on consumer engagement, an emerging behavioral metric in social media.

Finally, we develop a procedure to measure visual complexity using a computer vision API based on object-detection techniques and natural language processing (NLP). Although we take a relatively simple approach to operationalizing visual complexity (quantity of objects), this approach made it feasible to deal with large-scale image data and use it in our model.

We hope that our study will encourage future research possibly involving new data sets. We have studied a single social media platform (Instagram). Differences in the structural aspects of some platforms may need to be accounted for. However, the underlying results from this study that consumer engagement is affected by characteristics of the post, brand, and user may still be present and important to understand for the same reasons we have highlighted in this paper. Describing visual content is an emerging area of focus in many fields. Future studies may adopt a richer description of visual content than we have proposed. Finally, the context in which the post is viewed may be an important determinant of engagement. This has implications for data collection.

# Chapter 3

## Suspicious Online Product Reviews and Brand Advertising Effort

### 3.1. Introduction

Online product reviews are a major source of information for consumers making purchase decisions. Some 66% of shoppers who research online say they read customer-generated reviews on websites (*Wall Street Journal* 2016). Consumers use product reviews as a medium to collect information regarding product quality and performance (Forman, Ghose, and Wiesenfeld 2008). And, opinions posted online that are based on consumers' experiences with products or services have an impact on future consumer decisions (Chevalier and Mayzlin 2006; Gong et al. 2017; Stonedahl, Rand, and Wilensky 2010; Xu et al. 2014; Zhu and Zhang 2010). From a firm's perspective, user-generated content regarding its products or services is critical because of its financial effect on sales (Babić et al. 2016; Floyd et al. 2014; Ho-Dac, Carson, and Moore 2013; Zhu and Zhang 2010).

According to a survey by BrightLocal (2017), 85% of consumers state that they trust online reviews as much as personal recommendations.[6] However, the survey also revealed that 79% of consumers believe they have read a fake review in the last year. In the modern competitive business environment, firms may be tempted to perpetrate review fraud by creating positive reviews for their own products and denigrating reviews for competitors' products especially when they believe that their manipulated reviews are undetectable by customers in the partially anonymous e-world.

Figure 25 shows examples of suspicious online product reviews from our data. Both are highly

---

[6] 1,031 US-based consumers from BrightLocal's Local Consumer Panel, surveyed October 2017.

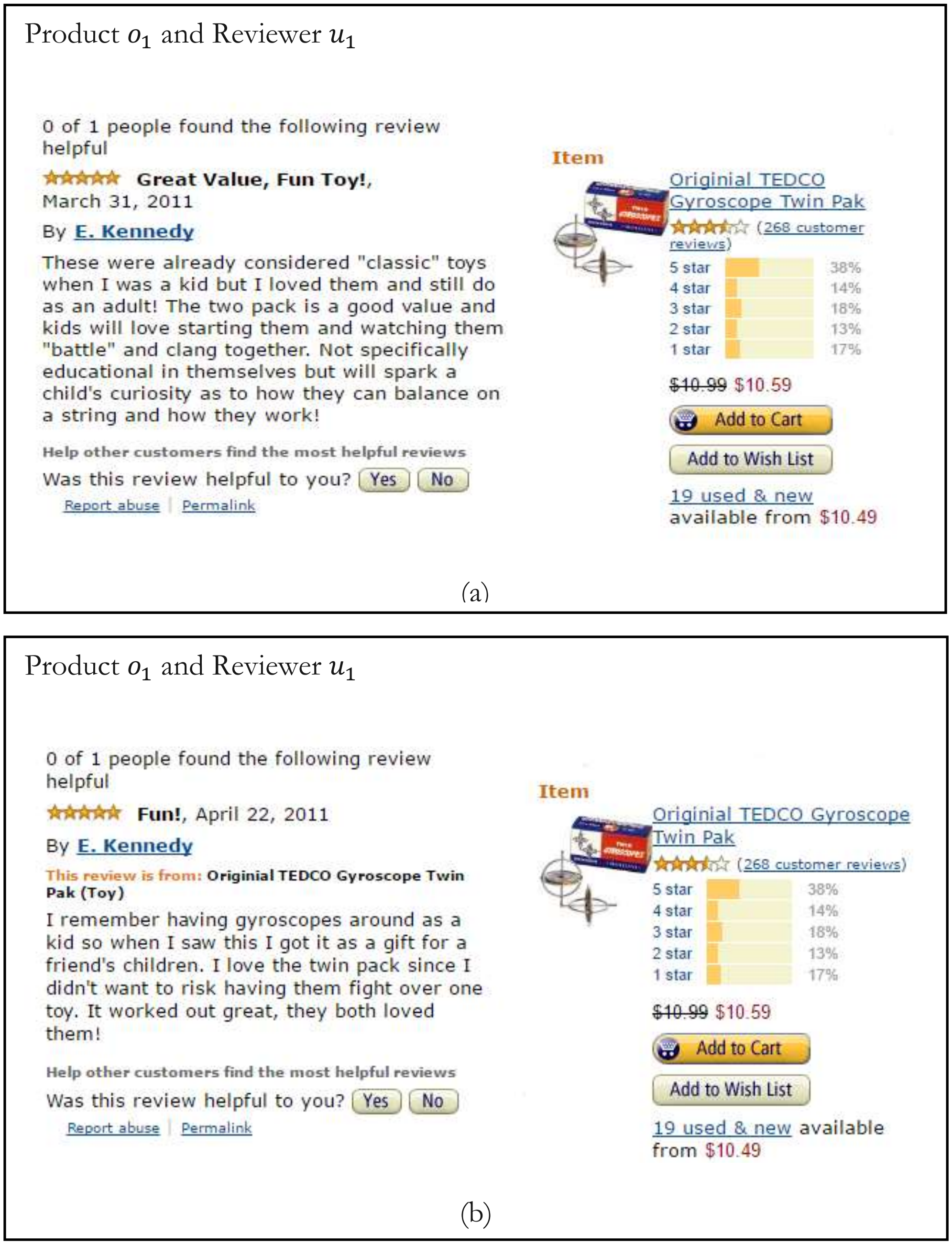
positive, with 5-star ratings that deviate from the product's overall 3.5-star rating. The most suspicious part of these examples is that the two reviews are written over three weeks apart by the same user (reviewer) about the same product, with seemingly different content and distinctive topics.

Review fraud is neither a hypothetical nor a rare practice. It is estimated that 10%–15% of reviews on an e-commerce website mirror earlier reviews and are potentially influenced by spamming practices (Gilbert and Karahalios 2010). Approximately 5% of reviews on a retailer's website without a verified purchase feature are created by customers who do not have a record of purchasing the product (Anderson and Simester 2014).

Potential harm from review manipulation affects both sellers and opinion-sharing platforms. Customers who purchase a product based on promotional reviews, and then find a gap between the product's expected and actual performance, may make suboptimal choices in the future (Mayzlin, Dover, and Chevalier 2014). The presence of fake reviews may hurt the reputation (and thus the revenue) of the recommendation systems of opinion-sharing platforms. Once customers mistrust reviews on a platform, they may underweight most of the reviews on that platform even though the majority of the reviews are created by genuine customers (Munzel 2015).

Social influence contributes to a bias in reviews that may be induced by manipulative practices. Social influence bias occurs when past ratings affect an individual's rating behavior (Muchnik, Aral, and Taylor 2013). Ratings can be influenced by historically positive reviews, resulting in the *J*-shaped distribution of online ratings found in many opinion sharing platforms (Aral 2013). Ho et al. (2017) find that a consumer is more likely to post a review and that the bias in the rating is more substantial when disconfirmation—discrepancy between the expected and experienced assessment of the same product—is larger.

Figure 25. Examples of Suspicious Reviews




Lappas, Sabnis, and Valkanas (2016) investigate the vulnerability of businesses to spam reviews, finding that even a small number of fake or suspicious reviews can significantly affect the visible status of a business, which is its position in the e-commerce platform. Recognizing the detrimental effects on their businesses, companies have sought to battle fake reviews. Amazon filed lawsuits against more than 1,000 alleged fake reviewers in April 2016, and Yelp began flagging stores with suspicious activity in 2012 through the use of Consumer Alerts.

Despite the potential harmful effect of review manipulation on a firm's value and business outcomes, and the importance of customer opinions on firms' finances, relatively few studies in marketing or related literature have investigated them. Researchers in computer science and information systems have developed various spam detection algorithms (see Crawford et al. 2015; Dewang and Singh 2018; Dixit and Agrawal 2013; Heydari et al. 2015 for comprehensive reviews), but few studies employ economic frames to delve into the malpractice. Authors who have advanced detection techniques or contributed to finding spam review characteristics typically exploit rating behavior or linguistic features such as sentiment polarity (e.g., positive, neutral, or negative), text similarity, and part-of-speech n-grams for classifying reviews as spam. Although linguistic features are substantially effective in distinguishing manipulated reviews, these methods are not comprehensive. Techniques based on linguistic features ignore the semantic aspects of reviews' contents. In spite of the superior predictive power of semantic aspects found in other studies (Linshi 2014; McAuley and Leskovec 2013), the hidden semantic structure in review text has rarely been exploited (Archak, Ghose, and Ipeirotis 2011).

Further, limited knowledge exists regarding the underlying mechanisms of opportunistic behavior surrounding manipulative reviews. It can be assumed that people have the goal of promoting their own products or businesses while denigrating others, but the factors that drive these malicious activities have rarely been explored in spite of the importance for brand management and quality control on review platforms.

We seek to eliminate these gaps in the literature. We analyze characteristics of suspicious reviews by focusing on semantic features such as emotionality and topic distributions, which go beyond basic sentiment polarity or text similarity. We then investigate weak brand status as a condition under which firms or individuals may be tempted to commit review fraud.

## 3.2. Background

A small number of papers have documented economic incentives for deceptive practices related to online reviews. Mayzlin et al. (2014) found that independent hotels tend to have more promotional (positive fake) reviews when compared to branded chain hotels. Luca and Zervas (2016) found that local and independent restaurants commit review manipulation more often than national chain outlets, which they argue is due to their weaker reputations. Luca (2011) discovered that chain restaurants are less affected by customer ratings on Yelp than independent stores, arguing that chain restaurants rely on different methods to establish their reputations such as branding and advertising campaigns. In this article we focus on products (as opposed to services) by linking brand advertising to the relative incidence of reviews that consumers may view as suspicious. Our interest is not in reviews that are known to be manipulative, something we cannot establish for certain, but rather in reviews that consumers are likely to suspect are not genuine.

Figure 26 summarizes our four-step research plan. Step 1 involves preprocessing the raw data to create a dataset suitable for analysis. We cannot directly know which reviews are spams (the ground truth). Accordingly, in Step 2 we develop a labeling procedure to classify reviews as suspicious versus genuine. Due to the large number of reviews, we use human evaluators to classify a small number of the reviews and then employ a machine-learning algorithm to classify the remaining reviews.

After building the labeled dataset, we develop a model (Step 3) that describes a review as suspicious (or not) by the semantic aspects of its text in addition to numeric features such as its star rating and word count. Our focus is on 1) how the emotional aspects of a review's content may indicate a suspicious review, and 2) how accounting for the underlying semantic structures of the text improves the forecasting power of the model.

In Step 4, we test the hypothesis that lower brand advertising effort is associated with more promotional (positive) reviews that consumers suspect may not be authentic. Using regression discontinuity analysis, a quasi-experimental approach, we discover a strong drive for suspicious promotional (positive) reviews when brand advertising effort is low. Negative suspicious reviews, on the other hand, are not related to brand advertising effort.

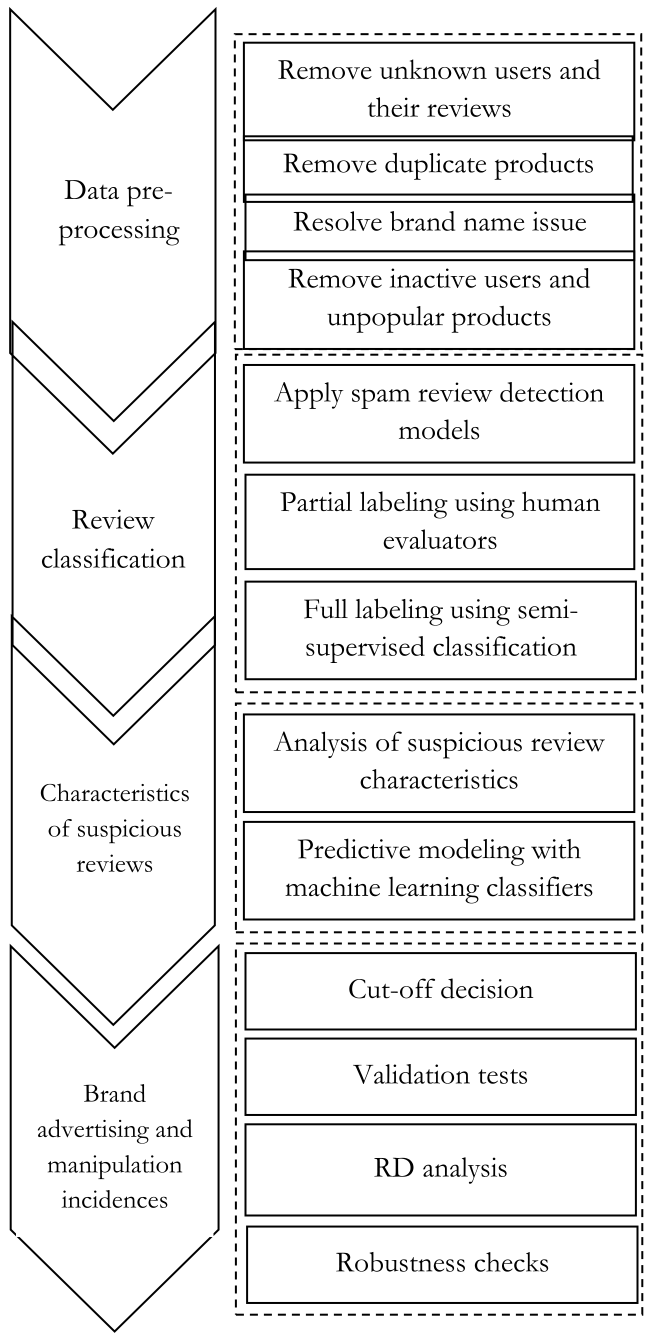
We make several contributions to the related literature. First, we seek to better understand the motivation for brands to contribute manipulative reviews, finding that suspicious reviews are more likely for brands that advertise less. Second, we extend prior research to provide a richer description of the characteristics of suspicious reviews, adding semantic and emotional aspects. Finally, we build a full labeling procedure for suspicious reviews using a semi-supervised learning method.

### 3.3. Create the Analysis Sample

Amazon review data is the main data source for our analyses for several reasons. First, Amazon.com is the leading online retailer in the US and has accumulated numerous reviews on products. Second, the rich body of text comments in Amazon reviews allows us to capture their semantic features, which is an essential component of our analyses. Third, we can observe multiple reviews by a given user (reviewer) and thus can adopt a reviewer-centric approach, which is more efficient at detecting spam than a review-centric approach (Dixit and Agrawal 2013). Finally, we can append brand information to each review, which allows us to merge the product-level review data with brand-level advertising expenditures.

The Amazon review data were obtained from the Stanford Network Analysis Project (SNAP) group at Stanford University (McAuley and Leskovec 2013). This data spans a period of 18 years to March 2013. We selected 16 physical product categories: Arts, Beauty, Cell, Phones and Accessories, Clothing and Accessories, Electronics, Gourmet and Foods, Health, Home and Kitchen, Industrial and Scientific, Jewelry, Shoes, Software, Sport and Outdoor, Tools and Home Improvement, Toys

Figure 26. Analysis Plan



and Games, and Watches. Exclusions were largely hedonic goods categories.[7] There are approximately 5.8 million reviews across the 16 categories. The data have product, review, and user attributes, including a unique product identifier (ASIN), product title, price, unique user identifier, user (reviewer) name, helpfulness score, review score, review time, review summary, and review text.

We also obtained Amazon metadata from the same group at Stanford University. The metadata includes ASIN, product title, price, brand, and co-purchasing links. Since the brand information is indexed with a unique product identifier, we can link the brand information to the product-level review data. The brand information was also used to facilitate merging brand-level advertising expenditures from Kantar Media's Stradegy database.

### 3.3.1. Initial Preprocessing

We performed two initial preprocessing steps. First, we removed unknown users and their reviews. Second, we identified sets of duplicate products in the dataset. Amazon.com maintains duplicate product listings, which are essentially the same products with some minor variations such as colors or sizes, and replicates the same reviews across the entire set of products. Using the identical reviews and unique identifiers, we removed the duplicate products except for one representative product per set. These steps dropped about 7% of the reviews, leaving approximately 5.5 million reviews, 2.6 million unique user accounts, and 565,745 unique products. A user (reviewer) posts about 2.08 reviews on average (std. dev. = 6.94), and each product has approximately 9.63 reviews on average (std. dev. = 50.73).

### 3.3.2. Merging Advertising Expenditure Data

We manually edited the brand names in the Amazon metadata to correct differences in spelling

---

[7] The full SNAP dataset contains 34 million reviews over 28 product categories. Almost 80% of the reviews are from three hedonic goods categories: books, music, and videos.

and case compared with their equivalent in the Kantar data. We merged the Amazon metadata that includes brand information with the review data. The metadata do not cover all the unique product identifiers in the Amazon review data, and the merging process resulted in approximately one-fourth of the data, or 1.5 million reviews and 107,428 unique products remaining.

The relationship between the number of reviews and the number of users (reviewers), and between the number of products, each follows the characteristic power law distribution ($y = ax^k$) found in online user review data: the logarithm of number of users and number of products as a function of the logarithm of number of reviews is a straight line. A relatively small number of users post a large number of reviews: 18 users each wrote more than 100 reviews. Most users post very few reviews: 82% posted only one review. Seventy-nine percent of the products have 10 or fewer reviews, and only 33 products out of 107,428 have more than 1,000 reviews. The star ratings follow the characteristic J-shape found in online user reviews (see Figure 27). The majority of reviews (60%) have the highest star rating of 5. The least frequent star rating (7.6%) is 2. Most products have a high average rating: 47.3% of products have an average rating between 4.5 and 5. Most users award very high scores, giving an average star rating between 4.5 and 5. Only 11.4% of products and 8.7% of users have an average star rating between 3 and 3.5. Figures 28 and 29 plot the frequency distributions by number of reviews.

### 3.3.3. Final Preprocessing

Our final preprocessing step removed inactive users and unpopular products. We dropped all users and all products with fewer than three reviews. In the resulting sample each user (reviewer) on average contributed nine reviews (std. dev. = 22).

Figure 27. Percentage of Reviews by Star Rating

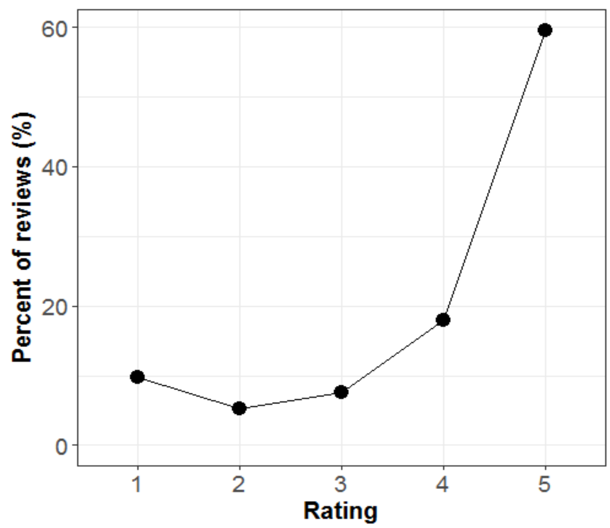



Figure 28. Number of Users Vs. Number of Reviews: Log-Log Plot

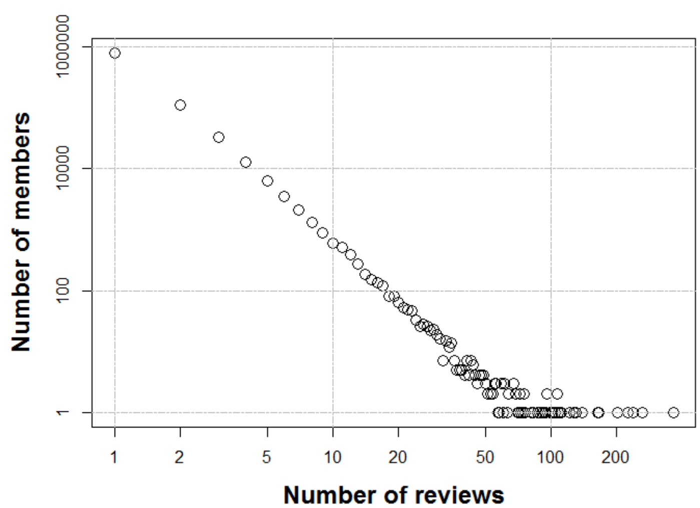
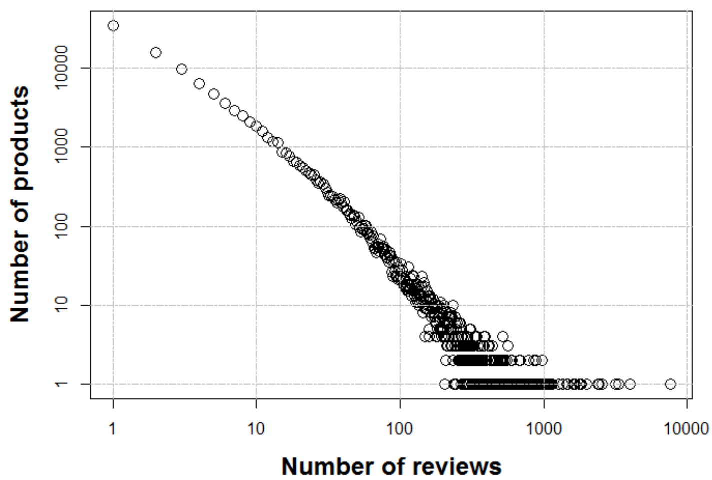


Figure 29. Number of Products Vs. Number of Reviews: Log-Log Plot




3.4. Classify and Label Reviews as Suspicious (Or Not)

There might be four different types of taxonomy for fake reviews: 1) truly genuine reviews that are created by verified purchasers, 2) reviews that are generated by robots, 3) reviews that are generated by non-verified purchasers, and 4) reviews that are generated by consumers who are verified to purchase products but have certain purposes such as promoting or denigrating the purchased products. In this study, we classify in a binary way (suspicious vs. genuine reviews).

Ideally human evaluators would judge the entire set of reviews, but this is not feasible due to the number of observations. Instead, human evaluators assess a subset of the reviews as suspicious (or not), which forms a 'ground truth' for a machine-learning algorithm to classify the remaining reviews.

The relatively low incidence of suspicious reviews presents a challenge if we are to create the training set using a randomly selected sample from all reviews. As such we first use an established spam detection algorithm (an automated procedure with no human evaluation) to identify reviews that are potentially suspicious. This informs a stratified sampling procedure used to select a small number of reviews to be evaluated and labeled by human investigators. We then label the remaining reviews using a semi-supervised method that employs both the labeled and unlabeled data to build a classifier for the unlabeled portion of observations.

### 3.4.1. Selecting a Training Set for Use by Human Evaluators

We first classified each review as potentially spam or genuine using Lim et al.'s (2010) spam detection algorithm, which, like our study, was developed with Amazon review data in a consumer opinion context. Assuming that spammers have promotional or defamatory purposes in perpetrating malicious reviews, the method relies on the patterns of review content and ratings to capture five different spamming behaviors: Targeting Product (TP) uses the number of reviews written by a reviewer on a product as well as the rating and text similarities of the reviews; Targeting Group (TG) is based on promoting or denigrating a set of products sharing common attributes within a short span of time; General Rating Deviation (GD) uses the deviation of each rating from the average rating of the rated product; Early Rating Deviation (ED) weights the GD score based on the order of the reviews on the rated product; and Combined Score (CS) is the weighted sum of the four prior terms (.65, .25, .05, and .05 for TP, TG, GD, and ED, respectively). TP and TG are highly weighted because it is possible for a potential spammer to create manipulative and genuine reviews with the same account. Appendix 5 describes how we applied the five behaviors to our data.

To create the training set for human evaluation, we selected the 20 top-ranked and the 20 bottom-ranked users (reviewers) from each of the four spam detection methods (not CS). This resulted in 107 reviewers and 820 reviews. We randomly ordered the users to avoid any systematic relationship between user order and their spammer scores. Then, we randomly sorted the reviews.

### 3.4.2. Coding the Training Set as Suspicious (or Not) Using Human Evaluators

Six human evaluators were recruited from a leading US university, screened for people who regularly read reviews on e-commerce websites. Three evaluators screened users 1 to 54; the other three screened users 55 to 107. They scored each review on a 5-point Likert scale: 1 = Non-spammer, 2 = Slightly suspicious, 3 = Somewhat suspicious, 4 = Highly suspicious, and 5 = Spammer. We coded the first two levels as non-suspicious and the remaining levels as suspicious. Fleiss's kappa, a measure of inter-evaluator consistency for multiple categorical items, is .64 (substantial agreement) for the first group of evaluators and .43 (moderate agreement) for the second group. We assigned a label to each review using majority voting. This classified 67 users as spammers and 40 users as non-spammers, and identified 758 reviews as suspicious and 62 reviews as genuine.[8]

Using unhelpfulness scores as the benchmark[9], we assessed the performance of the human evaluators using Normalized Discount Cumulative Gain (NDCG). NDCG is a normalization of the Discount Cumulative Gain (DCG), a weighted sum of the degree of relevancy of ranked items where the weight is a decreasing function of rank. The DCG function accumulates the gain from the top of the result list to the bottom, with the gain of each result discounted at lower ranks.[10] We compared NDCG scores for the top 50 ranked as spammers based on their Combined Score with benchmark scores measured using unhelpfulness for the same 50 users. With two exceptions (49 and 50), the NDCG scores based on Combined Score are consistently higher across the 50 top ranked users, and produce better rank orders than the baseline model using unhelpfulness scores. Details are in Figure

---

[8] This will become the input to a semi-supervised classifier. It is common to have imbalanced classification as input to semi-supervised learning (Li et al. 2011).

[9] We computed unhelpfulness votes by subtracting the helpfulness votes from the total votes for each review. We then computed an unhelpfulness score for a user by averaging unhelpfulness votes by total votes (unhelpfulness = 1: only unhelpful votes from the user; unhelpfulness = 0: only helpful votes).

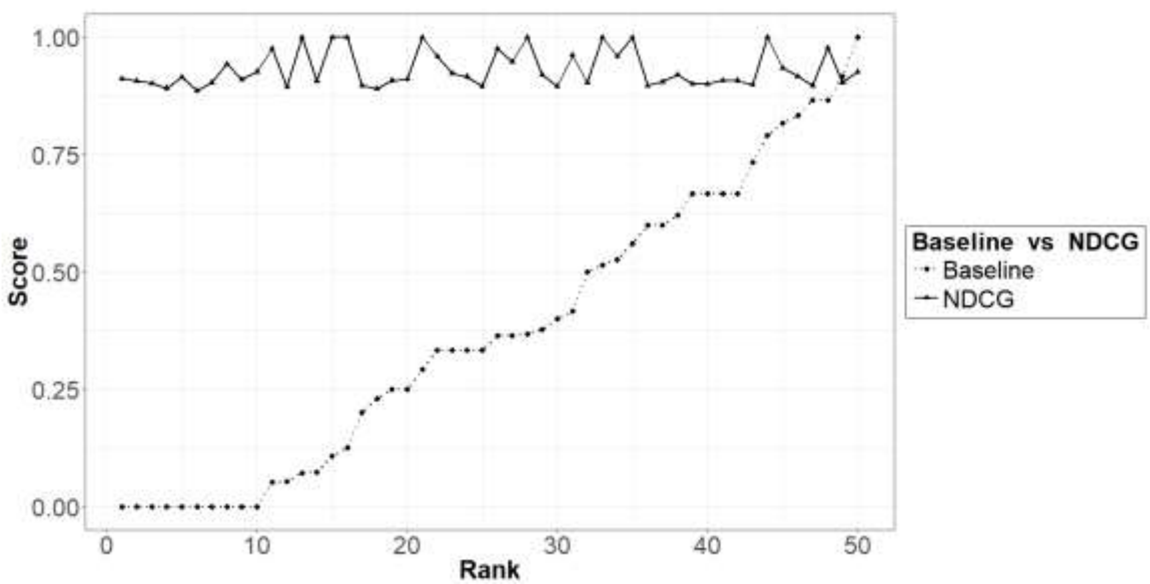
[10] The DCG of the top 50 reviewers ($u_{i1}, u_{i2}, u_{i3}, \dots, u_{i50}$) based on Combined Score is defined as follows: DCG = $\sum_{p=1}^{50} \frac{2^{f(i_p)} - 1}{log_2(1+p)}$, where $f(i_p)$ is the number of votes $u_{ip}$ received ($f(i_p) \in [0,3]$). NDCG is a normalization of DCG by the DCG of the ideal rank order of the entities. NDCG = $\frac{DCG}{DCG\ of\ ideal\ ordering}$, where the DCG of the ideal ordering is a monotonically decreasing sort of the relevance judgments provided by the human evaluator. Ideal ordering in our study would be, for example, 3, 3, 3, 2, 2, 1, 1, 0.

30 (sorted by unhelpfulness score).

### 3.4.3. Coding the Full Dataset as Suspicious (or Not) Using Semi-Supervised Classification

Next, we labeled the remaining 269,610 reviews. Semi-supervised learning methods can be used for a classification task with partially labeled data by building models from a small portion of labeled data along with unlabeled data (Zhu et al. 2015). Semi-supervised classification assumes that the underlying structure in the labeled data is also the dominant pattern in the unlabeled data, so applying the semi-supervised learning method is beneficial in constructing the classifier (Zhu 2007).

Figure 30. NDCG Results




We built our semi-supervised classifier using a model-based classification method. The method utilizes an EM algorithm to obtain maximum likelihood estimates of the model parameters and classifications for the unlabeled observations that are updated iteratively until convergence (Russell, Cribbin, and Murphy 2012). It performs well, particularly in cases where few labeled instances are available. The model was fit using the five spamming detection methods (Targeting Product, Targeting Group, General Deviation, Early Deviation, and Combined Score) described earlier as independent variables.

To build the semi-supervised classifier, we first split the labeled data into a calibration set and a test set using three candidate splits: 80:20, 60:40, and 50:50. Next, several models with different covariance structures involving possible constraints placed on the volume, shape, and orientation were tested in both the calibration phase and the test phase (Russell, Cribbin, and Murphy 2012). We selected the model with the covariance structure that had the lowest error in the calibration phase. The selected covariance structure was then used to fit the full training data. We fit the selected model to the test data to observe the performance of the classifier.

3.4.4. Results from Semi-Supervised Classification and Comparison with Supervised Classifiers

All three semi-supervised classifiers (80:20, 60:40, and 50:50) provide reasonably high and consistent predictive power. The model with 80% calibration and 20% test sets had the best performance (misclassification rate = 2.44%). We selected this model for our further analyses.

To further assess its predictive power, we conducted analyses using fully supervised methods (Support Vector Machine [SVM], Random Forest, and Logistic model) with the manually labeled data. We find that the semi-supervised classifiers are superior or comparable to the supervised classifiers. Random Forest and a semi-supervised classifier with the 80:20 model achieved the best (misclassification rate = 1.03%) and second-best (misclassification rate = 2.44%) performance. SVM achieved the worst performance (14.55%). We conclude that the semi-supervised classifier is suitable for our task. Summary statistics from the confusion matrix, including specificity and recall, are presented in Table 30.

Based on the selected model, the entire process of semi-supervised classification results in 8,073 suspicious reviews and 262,357 genuine reviews; potentially manipulative reviews account for approximately 3% of all product reviews. This is a considerably lower incidence than found in the hospitality industry where upwards of 20% of ratings have been found to be suspicious (Schuckert, Liu, and Law 2016).

Table 30. Confusion Matrix from The 80 Calibration: 20 Test Data Split

| True | Predicted | |
|---|---|---|
| (n = 164) | Genuine | Suspicious |
| Genuine | 10 (TN) | 1 (FP) |
| Suspicious | 3 (FN) | 150 (TP) |
| Overall Accuracy | 97.6% | |
| Precision | 99.3% | |
| Recall | 98.0% | |
| Specificity | 90.9% | |

*Notes.* "TN" True Negative: "FP" False Positive: "FN" False Negative:"TP" True Positive. Overall accuracy is computed as (TP + TN) / (TN + FN + FP + TP);
Precision is TP / (TP + FP); Recall is TP / (TP + FN); and, Specificity is TN / (TN + FP).

## 3.5. Characterize Suspicious Reviews Using Semantic Features

### 3.5.1. Semantic Characteristics of Suspicious Reviews

Although the role of emotions has been widely studied (Berger and Milkman 2012; Lovett, Peres, and Shachar 2013; Yin, Bond, and Zhang 2014), less is known about their effect in a spam review context. We focus on the semantic characteristics of reviews using language features such as emotionality and topic models to move beyond a mere valence approach. High arousal content is more viral (Berger and Milkman 2012), which may contribute as an external (efficiency) motivation for a brand manager to write a fake review. Further, Filieri's (2016) in-depth interviews revealed that a review's emotional content is a dominant cue used by consumers to gauge a review's trustworthiness.

To extract the semantic variables, we employ an automated text analysis system to quantify the emotionality of the customer opinions on products, and another automated system to uncover the unobserved underlying topics in a review. This allows us to understand what types of emotions (i.e., extreme emotions or mixed emotions) are associated with suspicious versus genuine reviews, and what language features lead to superior models. We tested various machine-learning classifiers (Logistic model, SVM, Random Forest, and Deep Learning) to find if semantic features enhance the predictive power of the model and which method performs best.

After preprocessing the text data by cleaning and stemming, we extracted eight primary emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) by employing a text mining system with a dictionary-based sentiment approach that automatically extracts emotionality from the text. The term frequency (TF) of an emotion is measured as its incidence divided by the document length (the total number of terms) for normalization.[11] Descriptive statistics for the data and their sources are provided in Table 31.

We estimated the following logistic regression model describing the incidence of suspicious reviews:

(1) $$s_{ijt} = p_j + X_{ijt}\beta_{ijt}^{x} + Z_{ijt}\beta_{ijt}^{z} + \varepsilon_{ijt.}$$

Where, $s_{ijt}$ indicates whether the review written by user i on product j at time t is a suspicious review or not (a binary indicator); $p_j$ is a product fixed effect; $X_{ijt}$ is a vector of numeric variables including (log of) number of reviews per user, word count per review, and star rating (reference level = 3); and, $Z_{ijt}$ is a vector of semantic variables, including (log of normalized) anger, anticipation, disgust, fear,

[11] $tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$, where $n_{i,j}$ is the number of occurrences of the considered term $(t_i)$ in document $d_j$, and the denominator is the sum of the number of occurrences of all terms in document $d_j$.

joy, sadness, surprise, and trust. Table 32 presents the results. After discussing the results, we apply several machine-learning classifiers to the data to examine if the analysis strategy that consolidates the information from textual comments of product reviews into models performs better than a baseline model, and to find the classifier that performs best with our models.

We first investigated the relationship between the numeric features of the reviews and the users (i.e., star ratings, number of reviews per user, and word count per review) and a review being classified as suspicious (or not). We expect suspicious reviews to be more extreme in terms of star ratings (see Figure 31) and shorter in length (Mukherjee et al. 2013; Shojaee et al. 2013). We also expect a positive relationship for the user's review count; some review manipulations may occur on a massive scale, perpetrated by machines as well as humans (see Figure 32 for model-free evidence).

The Baseline Model (column 1) in Table 32 includes only the numeric variables: (log of) number of reviews per user, word count per review, and star rating (reference level = 3). As expected, reviews created by users with a higher review count and reviews with extreme star ratings (1,2, or 5) are more likely to be suspicious. Consistent with Luca and Zervas's (2016) finding from Yelp reviews, lengthier reviews are less likely to be suspicious. We also examined a specification that also included a quadratic term for Word Count per Review (to test for a possible U-shaped relationship), but this did not fit the data as well (higher AIC).

Text Model A (column 2) extends the Baseline Model to incorporate emotionality. Reviews that include the emotions fear or joy are more likely to be suspicious reviews. According to social psychology theory (Plutchik and Kellerman 1980), fear and joy are bipolar emotions that possess clear positive or negative valence. Fear is a primary emotion that is covered by negative secondary emotions such as horror and nervousness, and joy is a primary emotion masked by secondary emotions of cheerfulness, zest, contentment, optimism, and relief. A one standard deviation increase in "joy" results in a 13.6% increase in the odds of a review being viewed as suspicious.

Customer reviews containing emotions that are characterized by extreme valence are more likely to be suspicious reviews. We interpret this result as partial evidence in favor of the argument that spammers with promotional or defamatory purposes exaggerate their opinions about products by using very positive or very negative words more frequently than genuine reviewers.

The results from Text Model A also suggest that suspicious reviews are less likely to use words that evoke neutral or mixed emotions that can have any valence, such as surprise or anticipation (Davis, Palladino, and Christopherson 2012; Fontaine et al. 2007; Kopec 2006). A one standard deviation increase in "surprise" decreases the odds of a review being suspicious by 5%.

Since spammers usually aim to promote or denigrate products, they may have less incentive to utilize opinion words with neutral emotional properties. The coefficient size and the negative direction of sadness, which is a low-arousal or deactivating emotion (Berger and Milkman 2012), is approximately identical to that of anticipation. These effects are even clearer in Text Model B: Full Text (column 3).

Text Model B: Full Text extends Text Model A by including as control variables 30 topic distribution values extracted from Latent Dirichlet Allocation (LDA). Since LDA requires a specified number of topics as an input, we heuristically determined 30 topics using a within-group, sum-of-squares plot. LDA is a topic model generated from input text documents in natural language processing. We consider the textural copy associated with each review a text document, and each review text is thus viewed as a mixture of topics. We use a probability distribution based on both frequency and weight of words that is associated with the generated topics as input values for our models.

Using the results from Text Model B: Full Text, Figure 33 summarizes the percentage change in the odds of a review being viewed as suspicious for a one standard deviation increase above the mean for each of the eight emotional characteristics.

Table 31. Descriptive Statistics

| Variable | Data Source | N | Mean | Std. dev. | Min | Max | Skewness |
|---|---|---|---|---|---|---|---|
| Number of reviews per user | Amazon review data | 270,418 | 9.11 | 21.80 | 3 | 362 | 10.02 |
| Word count per review | Amazon review data | 270,418 | 178 | 158 | 1 | 1878 | .18 |
| Star rating | Amazon review data | 270,418 | 4.26 | 1.17 | 1 | 5 | -1.61 |
| Anger | Extracted from a dictionary-based sentiment analysis | 270,418 | .008 | .01 | 0 | .25 | 2.41 |
| Anticipation | Extracted from a dictionary-based sentiment analysis | 270,418 | .02 | .02 | 0 | .50 | 1.88 |
| Disgust | Extracted from a dictionary-based sentiment analysis | 270,418 | .005 | .01 | 0 | .20 | 3.11 |
| Fear | Extracted from a dictionary-based sentiment analysis | 270,418 | .009 | .01 | 0 | .26 | 2.23 |
| Joy | Extracted from a dictionary-based sentiment analysis | 270,418 | .02 | .02 | 0 | .50 | 2.00 |
| Sadness | Extracted from a dictionary-based sentiment analysis | 270,418 | .009 | .01 | 0 | .20 | 2.11 |
| Surprise | Extracted from a dictionary-based sentiment analysis | 270,418 | .01 | .02 | 0 | .50 | 2.34 |
| Trust | Extracted from a dictionary-based sentiment analysis | 270,418 | .03 | .02 | 0 | .50 | 1.72 |

Table 32. Characteristics of Suspicious Reviews

| Variables | (1)<br>Baseline Model | (2)<br>Text Model A | (3)<br>Text Model B: Full Text |
|---|---|---|---|
| <u>Numeric Features</u> | | | |
| ln(*No. of Reviews per User*) | .551*** (38.94) | .539*** (38.27) | .478*** (47.22) |
| Word Count per Review | -.002*** (-5.89) | -.001* (-2.50) | -.001* (-2.53) |
| Star rating = 1 | 1.991*** (26.31) | 1.878*** (25.40) | 1.570*** (23.38) |
| Star rating = 2 | 1.151*** (13.60) | 1.062*** (12.82) | .939*** (12.39) |
| Star rating = 3 (reference level) | - | - | - |
| Star rating = 4 | .002 (.03) | -.107 (-1.46) | -.059 (-.82) |
| Star rating = 5 | 1.105*** (16.88) | .983*** (15.48) | .812*** (13.72) |
| <u>Semantic Characteristics</u> | | | |
| ln(*Anger*) | - | .035 (.92) | .085 (.26) |
| ln(*Anticipation*) | - | -.058# (-1.73) | -.104*** (-3.48) |
| ln(*Disgust*) | - | .061 (1.56) | .034 (.99) |
| ln(*Fear*) | - | .089* (2.32) | .089** (2.63) |
| ln(*Joy*) | - | .189*** (5.14) | .223*** (6.79) |
| ln(*Sadness*) | - | -.058 (-1.57) | -.103** (-3.16) |
| ln(*Surprise*) | - | -.077* (-2.19) | -.115*** (-3.71) |
| ln(*Trust*) | - | .037 (1.10) | .058# (1.95) |
| Topics 1,…,30 included | No | No | Yes |
| Product fixed effects included | Yes | Yes | Yes |
| n | 270,403 | 270,403 | 270,403 |
| AIC | 61,840 | 61,828 | 68,068 |

#*p*<.05; **p*<.01; ***p*<.001; ****p*<.0001

Notes. The dependent variable is a binary indicator of whether a review is viewed as suspicious or genuine. z-statistics in parentheses.

Figure 31. Distribution of Star Ratings for Suspicious and Genuine Reviews

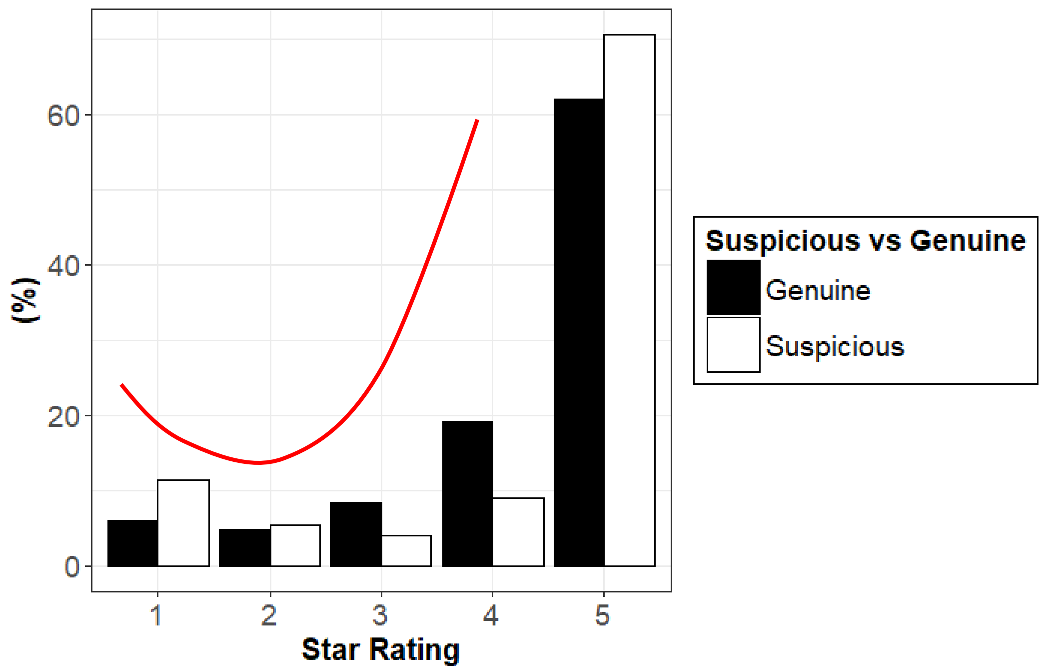



Figure 32. Incidence of Suspicious Reviews Versus Number of Reviews Per User

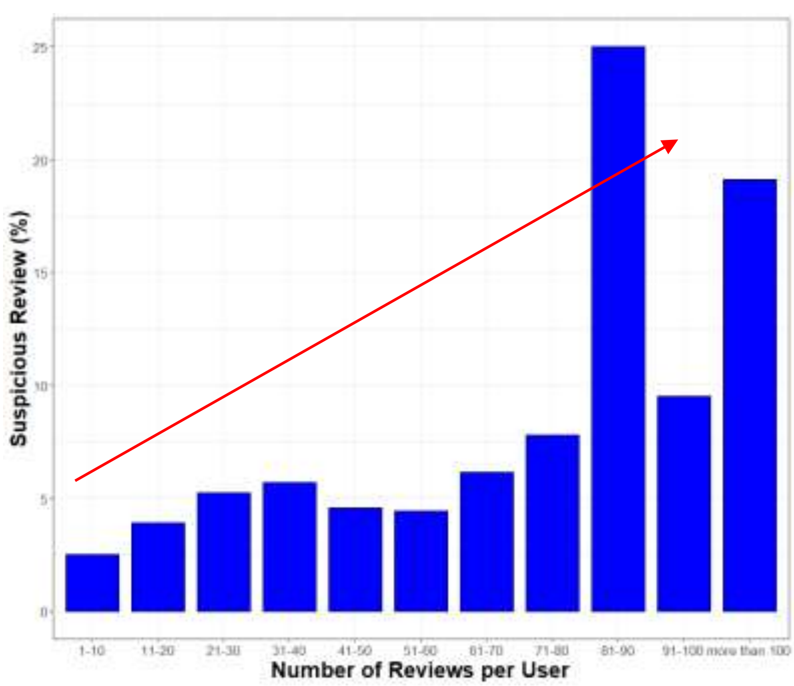
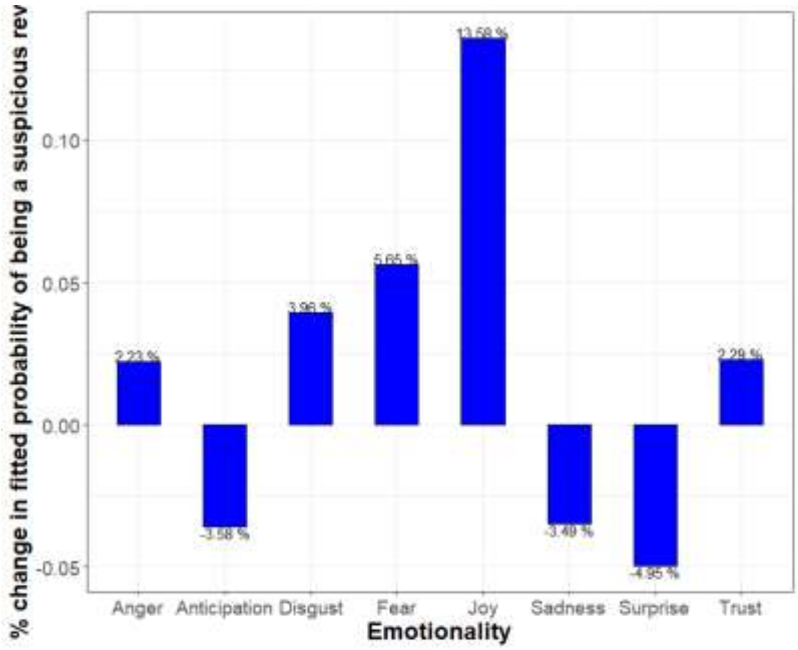


Figure 33. Percentage Change in the Odds of an Online Product Review Being Suspicious for a One Standard Deviation Increase Above the Mean for Each Emotional Characteristic




### 3.5.2. Robustness Check with Holdout Sample

This section examines the logistic model in §3.5.1. with a holdout sample. The same variables were used with 80% as a training set and 20% as a test set. The results for the logistic model with training set are reported in Table 33. They are consistent with the initial models. The *Number of Reviews per Reviewer* and negative (1- or 2-) or positive (3- or 4-) *Star Ratings* all have a significant impact on a review being suspicious. As for semantic variables from the text, extreme types of emotions (*Disgust, Joy*) again have a strong and positive association with the probability of a review being more suspicious. The negative coefficients for *Anticipation, Sadness,* and *Surprise* suggest that all other variables being equal, a suspicious review is less likely to contain these types of neutral emotions.

Table 33. Model Result with Training Set

| Variables | Model with Training Set |
| --- | --- |
| ln(*No. of Reviews per Reviewer*) | .488*** (37.57) |
| Word Count per Review | -.004 (-.614) |
| | |
| Star rating (reference level = 3) | |
| Star rating = 1 | 1.512***(20.32) |
| Star rating = 2 | .943***(11.27) |
| Star rating = 4 | -.070 (-.93) |
| Star rating = 5 | .756***(11.54) |
| | |
| ln(*Anger*) | .008 (-.21) |
| ln(*Anticipation*) | -.136***(-4.07) |
| ln(*Disgust*) | .119**(3.11) |
| ln(*Fear*) | .046 (1.24) |
| ln(*Joy*) | .034***(9.56) |
| ln(*Sadness*) | -.148***(-4.08) |
| ln(*Surprise*) | -0.68*(-1.96) |
| ln(*Trust*) | -.031 (.93) |
| | |
| N | 216,323 |
| AIC | 54,757 |

Notes. The dependent variable is a binary indicator of whether a review is viewed as a suspicious or genuine review. The model with training set include the probability values of 30 topics from LDA (i.e., Topic1, Topic2, … , Topic 30). The z-statistics are in parentheses.

#### 3.5.2.1. Diagnostic Analysis

The difference between the null deviance and the residual deviance implies how the model performs against the null model (a model with only the intercept). Table 34 reports the drop in residual deviance as each variable is added to the model. Adding *Number of Reviewer per Reviewer*, *Word Count, Reviewer Rate, and some of the emotions (Disgust, Joy, Sadness, Surprise)* significantly reduces the residual deviance. The large and insignificant variables (*Anger, Fear*, and *Trust*) show that the model without the variables explains the similar amount of variation. Since there is no exact equivalent to the $R^2$ of linear regression, the McFadden $R^2$ index is used to assess the model fit.

#### 3.5.2.2. Predicted Power: Accuracy and ROC Curve

The discussion above focused on within sample fit. Now we assess out-of-sample fit; how the model performs when predicting *y* on a new set of data (test data). The decision boundary is 0.5, although different thresholds can be better for some applications. The accuracy on the test set is 0.969, which indicates very good predictive performance. Finally, Figure 34 is the plotted *ROC curve*. The *AUC* (area under the curve), a typical performance measurement for a binary classifier[12]., is 0.699.

### 3.6. Predictive Modeling with Alternative Machine-Learning Classifiers

We examined different machine-learning algorithms with and without the semantic features from the text comments. In doing so, we compared the Baseline Model without language features with models with these features and find that the language processing technique improves the predictive power of our classification problem. We also assess whether our results are robust across

---

[12] The ROC is a curve generated by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings while the AUC is the area under the ROC curve. As a rule of thumb, a model with good predictive ability should have an AUC closer to 1 (1 is ideal) than to 0.5.

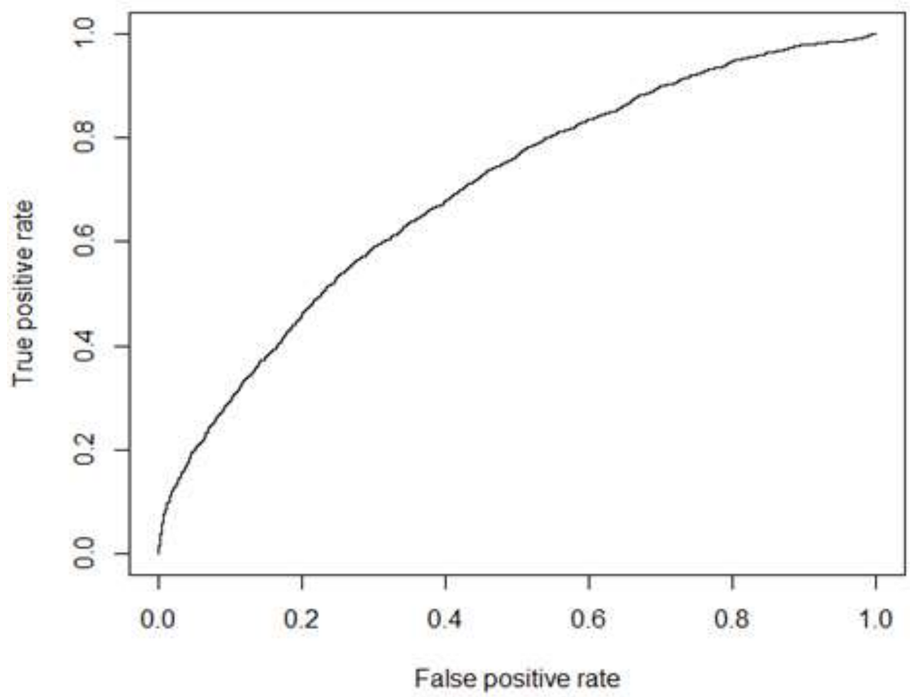
Table 34. Diagnostics 1

| | DF | Deviance | Resid. Df | Resid. Dev | Pr (>Chi) |
|---|---|---|---|---|---|
| NULL | | | 216,322 | 57,713 | |
| ln(*No. of Reviews per Reviewer*) | 1 | 1312.12 | 216,321 | 56,401 | .000 *** |
| Word Count per Review | 1 | 35.31 | 216,320 | 56,365 | .000 *** |
| Reviewer Rate | 4 | 980.38 | 216,316 | 55,385 | .000*** |
| ln(*Anger*) | 1 | 0.93 | 216,315 | 55,384 | .334 |
| ln(*Anticipation*) | 1 | 3.25 | 216,314 | 55,381 | .071 |
| ln(*Disgust*) | 1 | 14.45 | 216,313 | 55,366 | .000 *** |
| ln(*Fear*) | 1 | 2.35 | 216,312 | 55,364 | .126 |
| ln(*Joy*) | 1 | 143.02 | 216,311 | 55,221 | .000 *** |
| ln(*Sadness*) | 1 | 20.08 | 216,310 | 55,201 | .000 *** |
| ln(*Surprise*) | 1 | 6.80 | 216,309 | 55,194 | .009** |
| ln(*Trust*) | 1 | 1.61 | 216,308 | 55,192 | .205 |

Table 35. Diagnostics 2

| llh | llhNull | G2 | McFadden | r2ML | r2CU |
|---|---|---|---|---|---|
| -2,733.86 | -2,885.64 | 303.56 | 0.053 | 0.014 | 0.059 |

Figure 34. ROC Curve




different machine-learning classifiers and determine which algorithm performs best with our models.

We assessed four standard algorithms: Logistic classification, SVM, Random Forest, and Deep Learning. For each classifier, we estimated four models: the Baseline Model that includes the three numeric variables; Text Model 1 that incorporates basic sentiment values (i.e., a strong sentiment score and an ordinary sentiment score)[13] from the text data into the Baseline Model; Text Model 2 which incorporates emotionality from the text data into Text Model 1; and, Full Text Model which also

---

[13] We first extract strong (or ordinary) positive (or negative) sentiment terms from each review (4 terms) and then compute strong (or ordinary) sentiment scores (2 scores) for each review following (Hu et al. 2012): Strong sentiment score = $\frac{str\_pos_i + str\_neg_i}{senti\_tot_i}$ , and Ordinary sentiment score = $\frac{ord\_pos_i + ord\_neg_i}{senti\_tot_i}$,

where $str\_pos_i$ is a strong positive sentiment score, $str\_neg_i$ is a strong negative sentiment score, $ord\_pos_i$ is an ordinary positive sentiment score, $ord\_neg_i$ is an ordinary negative sentiment score, and $senti\_tot_i$ is a total sentiment score for a review $i$.

combines 30 topic distribution values from LDA into Text Model 2. Before running the models, we rescaled the range of the independent variables to [0, 1] using *Z*-score normalization.[14]

Table 36 reports the results of 10-fold cross-validation on the four models (Baseline Model, Text Model 1, Text Model 2, and Full Text Model) for the four different classifier types. Model fit, measured as the 10-fold cross-validated area under the curve (AUC) of receiver operating characteristics (ROC) curve, trends upward as we add more language features. However, the model fits decrease for SVM, Random Forest, and Deep Learning when we add only basic sentiment values to the Baseline Model. When we delve into the topic distribution values from LDA (Full Text Model), the results from the models for all four classifiers greatly improve. For example, the Full Text Model for Random Forest increases the cross-validated AUC from Text Model 2 without the topic distribution values from .613 to .678. As for classifiers, Logistic and Deep Learning perform better than the other two classifiers. The AUCs of the first three models (except the Full Text Model) of the Logistic classifier are higher than those of the comparable models for the other three classifiers, and the Full Text Model using the Deep Learning classifier demonstrates superior performance to all the other models, with a 21% improvement in AUC (.639 to .776) from the Baseline Model.

Overall, we find that language features improve the predictive power of the models, and that the probability values from topic models in particular improve the models to the greatest degree, which is consistent with prior research in other contexts (e.g., Netzer et al. 2016). The Full Text Models are enhanced from the Baseline Models by 6.2%, 10.6%, 4%, and 21.4% for Logistic, SVM, Random Forest, and Deep Learning, respectively. SVM tends to do the worst and Logistic and Deep Learning tend to perform the best, although there are slight variations depending on the model.

[14] (Z-score normalization) = $X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$.

Table 36. 10-Fold Cross Validation Area Under the Curve (AUC) for Different Models and Classifiers

| Model | Classifier | 10-Fold Cross Validated AUC |
|---|---|---|
| Baseline | Logistic | .664 |
| Text Model 1: Baseline+Sentiment | Logistic | .663 |
| Text Model 2: Baseline+Sentiment+Emotions | Logistic | .670 |
| Full Text Model: Baseline+Sentiment+Emotions+Topics | Logistic | .705 |
| Baseline | SVM | .527 |
| Text Model 1: Baseline+Sentiment | SVM | .508 |
| Text Model 2: Baseline+Sentiment+Emotions | SVM | .516 |
| Full Text Model: Baseline+Sentiment+Emotions+Topics | SVM | .583 |
| Baseline | Random forest | .652 |
| Text Model 1: Baseline+Sentiment | Random forest | .624 |
| Text Model 2: Baseline+Sentiment+Emotions | Random forest | .613 |
| Full Text Model: Baseline+Sentiment+Emotions+Topics | Random forest | .678 |
| Baseline | Deep learning | .639 |
| Text Model 1: Baseline+Sentiment | Deep learning | .627 |
| Text Model 2: Baseline+Sentiment+Emotions | Deep learning | .659 |
| Full Text Model: Baseline+Sentiment+Emotions+Topics | Deep learning | .776 |

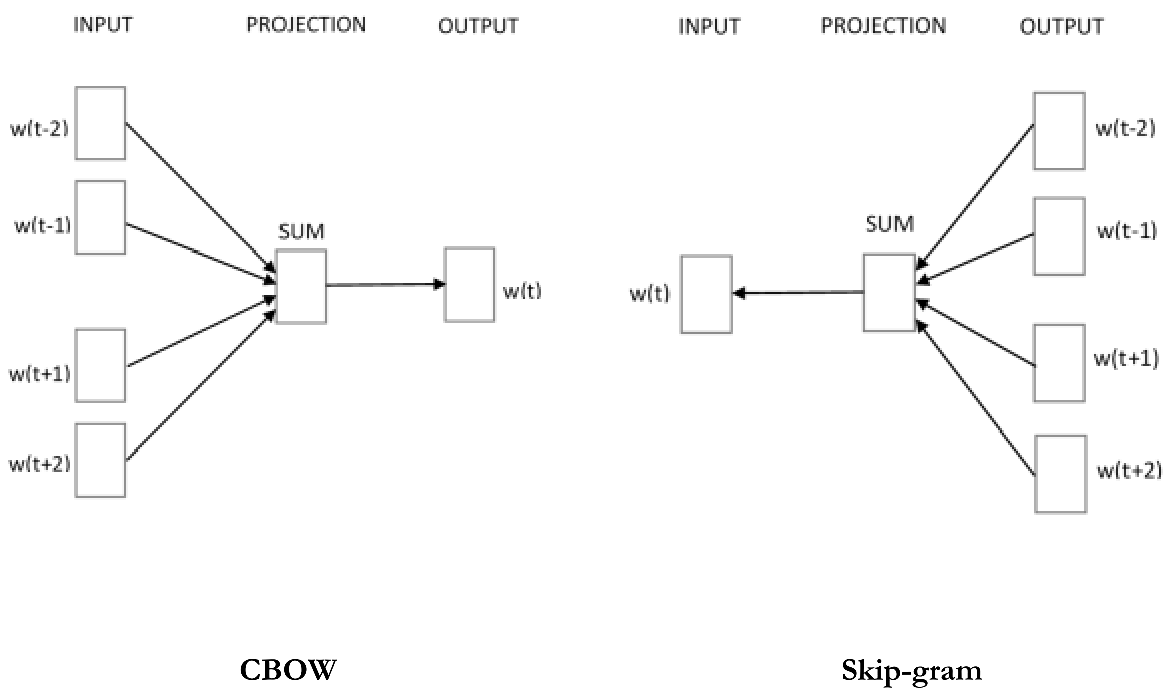
Notes. The Baseline Model includes ln(*Number of Reviewers per Reviewer*), *Word Count per Review*, and *Star Rating*.
Text Model 1 adds *Strong Sentiment Score* and *Ordinary Sentiment Score* to the Baseline Model.
Text Model 2 adds the emotion variables (i.e., ln(normalized *Anger*), ln(normalized *Anticipation*), ln(normalized *Disgust*), ln(normalized *Fear*), ln(normalized *Joy*), ln(normalized *Sadness*), ln(normalized *Surprise*), and ln(normalized *Trust*) to Text Model 1.
The Full Text Model adds the probability values of 30 topics from LDA (i.e., Topic1, Topic2… Topic 30) to Text Model 2.

## 3.7. Explore Word2vec Model

Word2vec is a popular technique used to learn word embeddings based on deep learning. It was developed and released in 2013 by a team of researchers at Google led by Tomas Mikolov. Unlike the traditional text mining models based on bag-of-words (BoW), which describe the occurrence of words within a document and disregard grammar and word order, the Word2vec model uses a vector space model in which words that share a common context in the corpus of text are located close to one another in the vector space. Word2vec employs either a continuous bag-of-words (CBOW) model, which predicts the current word based on the surrounding words, or a continuous skip-gram model, which predicts the surrounding words based on the current word (Figure 35).

Word2vec has shown superior performance in a variety of contexts for both supervised and unsupervised learning. This section explores the Word2vec model in the supervised learning context (i.e., text classification). In particular, the paper concentrates on the sentiment classification of the textual part of Amazon reviews. Sentiment classification is a special text classification task in which texts are classified based on their sentimental polarities or emotional properties (Pang et al. 2002), such as positive or negative, favorable or unfavorable, and anger or joy. Sentiment classification has been employed in many applications in a variety of contexts, such as business intelligence, recommender systems (Terveen et al. 1997; Tatemura 2000), and reputation monitoring systems. Due to the importance of the task, it has been widely explored with diverse techniques to achieve better performance. In particular, machine learning techniques (e.g., Naïve Bayes, support vector machine, random forest) have been widely used for the prediction task associated with sentiment classification. This section explores the Word2vec model and machine learning techniques (i.e., Naïve Bayes and random forest) to examine and compare the predictive effectiveness of different types of models—a deep learning model (i.e., Word2vec) based on a word embedding approach and machine learning models based on traditional BoW vector representations—in a text classification context.

Figure 35. CBOW and Skip-gram




3.7.1. Labelling Texts

Since the classification task requires labeled data, each Amazon review (i.e., document) must be labeled before classification. To do so, we employ SentiStrength (Thelwall et al. 2010), which estimates the strength of positive and negative sentiments and achieves human-level accuracy in social web contexts. SentiStrength reports two types of sentiments, negative and positive, and rates them on a scale ranging from not negative (or positive) to extremely negative (or extremely positive) for each document. SentiStrength also reports binary (positive/negative) results. We use the binary information to label individual documents. To create balanced data, we randomly sample 90,000 reviews, 45,000 of which have a negative sentiment and 45,000 of which have a positive sentiment.

### 3.7.2. Fine-Tuning Learned Word Embeddings from Word2vec

We train a one-dimensional convolutional neural network (CNN) to classify a document as either positive or negative. The one-dimensional CNN learns convolution filters that work on sentences a few words at a time and max pools the results to create a vector that represents the most critical ideas presented in the document. We perform sentiment analysis using Word2vec, a deep learning model, and train the sentiment classifier with Keras, a high-level neural network API that is written in Python. The natural language toolkit (NLTK) in Python is used to clean, tokenize, and parse the texts.

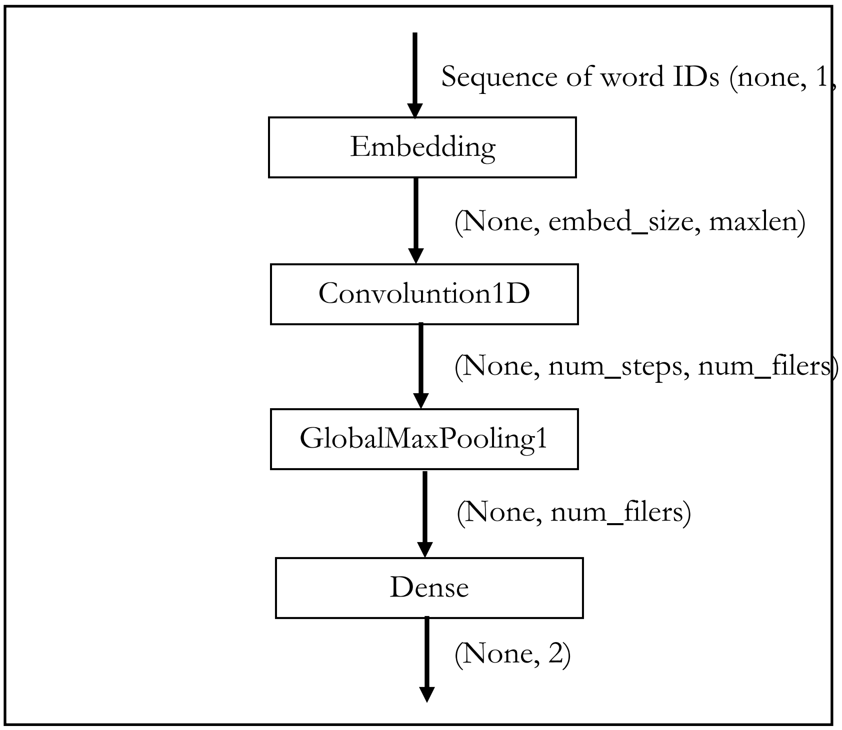
In the one-dimensional CNN network, the sequence of word indices is fed into an array of embedding layers, and the embedding layers are initialized to random values by default. The 1D convolutional layer convolves the output of the embedding layer in 256 different ways. A global max pooling layer then pools the features into a single pooled word. This (256) vector is then fed into a dense layer. A pair of probabilities, one corresponding to positive sentiment and another to negative sentiment, are returned by softmax activation. Figure 36 displays the CNN network.

We use a fine-tuning word embedding model, Word2vec, based on a pre-trained model. This model was trained with about 10 billion words from Google News articles and contains a vocabulary of 3 million words. The weight of the embedding layer of the network was initialized with the embedding weight matrix, which was built with the pre-trained model. For prediction purposes, the data was split into 70/30 training and test sets. Finally, the model was compiled with the categorical cross-entropy loss function since the outcome variable is binary (positive or negative).

### 3.7.3. Naïve Bayes

Naïve Bayes is a simple probabilistic classifier used in machine learning that is based on Bayes' theorem. It has a strong assumption that all predictors are strictly independent. Naïve Bayes first converts the data into a frequency table and then creates a likelihood table by finding the probabilities corresponding to each class. Next, it calculates the posterior probability for each class. The outcome of the prediction is the class with the highest posterior probability.

Figure 36. One-Dimensional CNN Architecture




### 3.7.4. Random Forest

Random forest is a supervised learning method that is used for classification, regression, and other tasks related to a multitude of decision trees. It uses ensemble methods that incorporate a multiple learning model to achieve better predictive results by growing many uncorrelated decision trees and determining the best possible results. Random forests produce an average of the multiple decision trees that are trained on different parts of the same training set in order to diminish the variance between multiple trees.

### 3.7.5. Results

We conduct experiments to evaluate the sentiment classification of Amazon reviews by employing several different machine learning classifiers and a deep learning model (Word2vec). The accuracy of classification by Word2vec, Naïve Bayes, and the random forest method is shown in Figure 37. Table 37 shows a confusion matrix that includes the precision, recall, and F1-scores of the two machine learning methods. For the two machine learning classifiers, we evaluate the models based on both BoW and TF-IDF. In terms of accuracy, the Word2vec model shows the best performance (93%) in comparison to Naïve Bayes (81%) and random forest (89%). The models with TF-IDF do not show improved performance compared to models with BoW. Figure 38 reports the accuracy and loss of the training and validation sets. Based on the plot of accuracy, 10-epoch training is enough to train the model, as the trend in accuracy for both datasets is stable between 4 and 6 epochs. In the plot of loss, the two plots for the training and test sets consistently depart from each other, implying again that the model was sufficient trained with 10 epochs.

Figure 37. Accuracy of Word2vec vs. Machine Learning

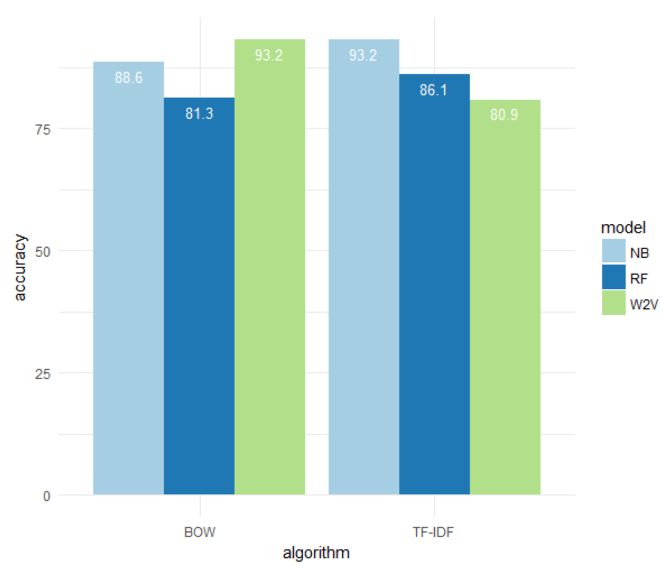
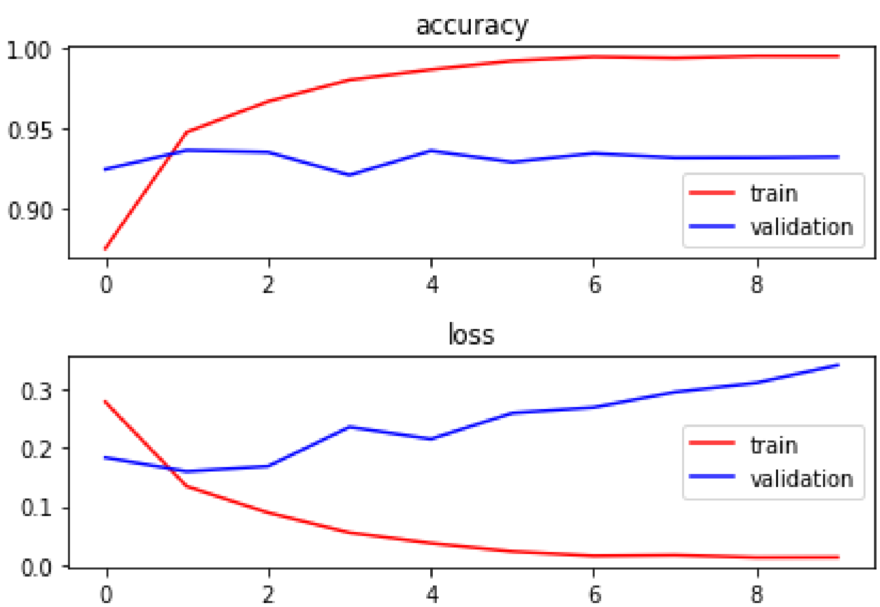


Table 37. Confusion Matrix of Word2vec vs. Machine Learning

| Model | | Accuracy | | Precision | | Recall | |
|---|---|---|---|---|---|---|---|
| Word2Vec | | 0.93 | | | | | |
| Naïve Bayes | BOW | 0.81 | BOW | 0.84 | BOW | 0.77 | BOW |
| | TF-IDF | 0.81 | TF-IDF | 0.85 | TF-IDF | 0.76 | TF-IDF |
| Random Forest | BOW | 0.89 | BOW | 0.90 | BOW | 0.87 | BOW |
| | TF-IDF | 0.86 | TF-IDF | 0.88 | TF-IDF | 0.84 | TF-IDF |

Figure 38. Accuracy of Word2vec Model Over Number of Epochs



### 3.7.6. Conclusions

The current section investigates the deep learning model (Word2vec) and machine learning methods for a text classification task. We compare BoW models, which rely on the number of pre-computed co-occurrences, and Word2vec, which takes a large corpus of text as input and predicts a word vector given its surrounding context. Instead of training a unique model from scratch, we employ a fine-tune learned embedding model based on a pre-trained Google News model due to the limited amount of available data. In terms of relative performance, Word2vec tends to perform the best (93%) and Naïve Bayes tends to perform the worst (81%), implying that the deep learning model relying on surrounding context achieved better performance rather than traditional models based on the number of lexicon-based co-occurrences.

## 3.8. Merge Suspicious Reviews and Brand Advertising Effort

To date research on social media and online word-of-mouth has largely focused on outcomes and not on the motivations of user activities (Toubia and Stephen 2013). The motivation behind manipulative reviews has not been studied extensively, although it seems clear that the goal of review manipulation is promoting or denigrating products or services. We hypothesize that a weaker brand, operationalized as having lower advertising effort, has more suspicious reviews that are promotional (positive) in nature. We argue that this is because a seller with a weak brand has a greater financial incentive for review fraud due to a less established reputation and more limited budget for advertising.

We merge brand advertising effort in dollars with the incidence of suspicious reviews aggregated from the product-level to the brand-level. Since the two datasets have different observation windows, we align them with each other by restricting them to January 2012 to March 2013.

Endogeneity may be a concern when assessing whether there is a relationship between brand advertising effort and the incidence of suspicious reviews. For example, products with higher prices may have a lower incidence of review fraud because the level of consumer involvement is usually higher, and customers will therefore scrutinize the reviews more carefully. The potential correlation between a product's price and review fraud may amplify the causal influence of advertising effort. Demand may also create a selection bias. Low-demand products may have more spam reviews than high-demand products, because the retailers of low-selling products may have a greater financial incentive for review manipulation that promotes their products. It would be ideal to design a random assignment study that eliminates the effects from a confounding variable. However, this is usually not practical or feasible due to constraints such as cost.

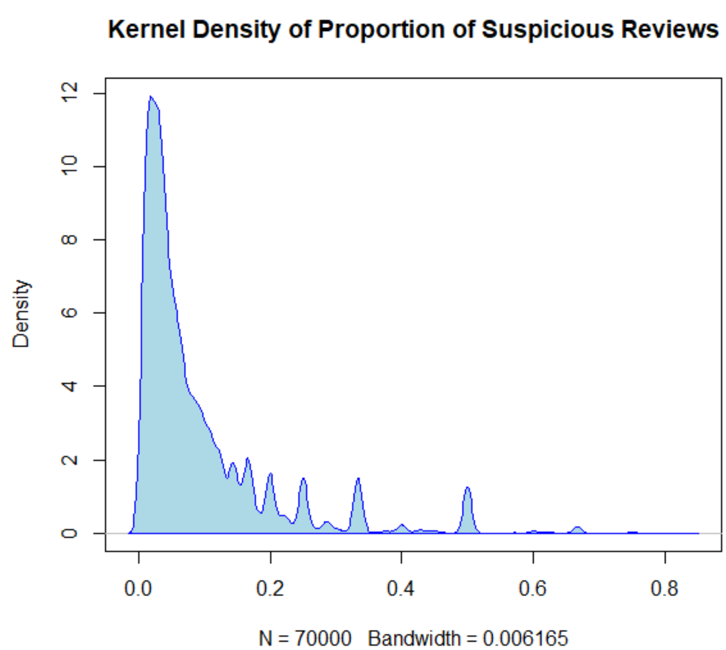
We employ a quasi-experimental design, regression discontinuity (RD), which compares advertising expenditure scores that are treated by a predetermined intervention (cutoff point) with untreated scores. The intuition behind the method is that we do not expect any systematic difference between the units that are barely above or barely below the threshold and thus can measure the unbiased estimates of the relevant effect. RD compares the outcomes whose score (also known as a running variable) is just above or below a known cutoff, excluding the possibility of a systematic difference between products with brand advertising effort just above and below the cutoff. The threshold determines assignment to the treatment or the control group. This approach assumes that the treatment assignment, based on the cutoff, generates a discontinuity in the probability of the treatment receipt at the point (Jacob et al. 2012). In our context, a treatment is given to those observations with advertising expenditure exceeding the cutoff and is withheld from the observations whose value is below the cutoff.

To differentiate positive and negative reviews, we estimate specifications with the ratio of the number of positive suspicious reviews (i.e., suspicious reviews with 4 stars or 5 stars) to the total number of reviews per product as the dependent variable for suspicious reviews with a promotional purpose. The independent variable is the total advertising expenditure for a brand from January 2012 to March 2013. Two covariates, mean price and mean review rating, are employed for validation tests and as control variables. The descriptive statistics are in Table 38. Figure 39 presents the distribution of an outcome variable (proportion of suspicious reviews).

Table 38. Descriptive Statistics for Brand Expenditures Analysis

| Variable | N | Mean | Std Dev | Min | Max | Skewness |
|---|---|---|---|---|---|---|
| Advertising expenditures ($ ‘000) | 12,386 | 5,187 | 18,080 | 0 | 1,176,954 | 27.26 |
| Positive suspicious review | 12,386 | .02 | .10 | 0 | 1 | 7.53 |
| Negative suspicious review | 12,386 | .01 | .06 | 0 | 1 | 14.60 |
| Average price | 12,386 | 37.11 | 71.36 | .01 | 999.99 | 6.11 |
| Average star rating | 12,386 | 4.29 | .94 | 1 | 5 | -1.71 |

Figure 39. Density Plot of Proportion of Suspicious Reviews



### 3.8.1. Determining the Cutoff

In order to implement RD analysis, we first need to establish the cutoff where small shifts in advertising expenditures lead to a significant change in the number of suspicious reviews. The cutoff point is identified from the data where the discontinuity occurs (Kerr, Lerner, and Schoar 2014). To find a reliable cutoff, we change the level of advertising expenditures in 10% blocks and examine the average percentage of positive manipulative reviews in each block. The largest shift occurs between advertising expenditure levels of 30%–40% and 40%–50%, where the average percentage of suspicious reviews drops from 2.620% to .447%. We set this as the cutoff.

### 3.8.2. Validation and Implementation of RD Design

We checked two necessary conditions for the validity of the RD design: 1) testing the density of the running variable, and 2) testing the treatment effect on the predetermined covariates. The first test evaluates whether the density of the running variable is continuous at the cutoff point (Hahn, Todd, and Van der Klaauw 2001). We tested this assumption using two methods: a density plot, and a McCrary (2008) sorting test. The second test is a falsification test that evaluates whether the other covariates exhibit a discontinuity at the threshold (Skovron and Titiunik 2015). We expect that the null hypothesis (no treatment effect) will not be rejected. After conducting the validation tests, we select the optimal bandwidth $\lambda$, which is in close vicinity to the cutoff point, using a cross-validation approach. Finally, we find the effect of brand advertising level on the incidence of suspicious reviews using local linear or local quadratic polynomial regressions. Following Skovron and Titiunik (2015), we estimate all the validation tests the same way as the outcome of interest, by choosing the optimal bandwidth with the same method and employing the same local regression function.

We estimate the effect, Equation (2), using the brand advertising expenditures and a nonparametric method to approximate the continuous relationship between the independent variable and the outcome variable with a local linear regression. The RD specification is as follows:

(2) $$y_i = \beta \times brand_i + L_l(a_i,\gamma_l) + L_r(a_i,\gamma_r) + \varepsilon_i$$

where $L_l(a_i,\gamma_l)$ is the left side of the threshold, $L_r(a_i,\gamma_r)$ is the right side of the threshold, and $\beta$ is a consistent estimate of the causal effect of brand advertising level on the incidence of suspicious reviews ($y_i$). We use a local quadratic polynomial regression in addition to the local linear method to check the robustness of the results.

The first validity test evaluates whether the running variable is continuous at the cutoff point (Hahn, Todd, and Van der Klaauw 2001). Visual inspection of the histogram supports the fact that no stark discontinuity point exists in the distribution of brand advertising expenditures, and the distribution is smooth and continuous at the cutoff. We also conduct the more formal McCrary sorting test (2008) that assesses continuity at the cutoff in the density function of the running variable. The null hypothesis that the discontinuity is zero cannot be rejected at $\alpha = .05$ (p-value = .43), meaning there is no evidence that the density of the running variable is discontinuous at the cutoff point.

Next, we analyzed two predetermined covariates: the average price of a product, and the average review rating of a product. We estimated the RD model with the two predetermined covariates as a dependent variable and brand advertising expenditure as an independent variable, using a local linear regression and a cross-validated optimal bandwidth method. We find no evidence of a discontinuity at the threshold for the predetermined covariates, demonstrating that the estimated effects are not significantly different from zero at the cutoff (Table 39).

We then analyze whether the level of brand advertising has a causal relationship with the incidence of review fraud. We hypothesize that low level of brand advertising effort leads to an increase in suspicious reviews with positive purposes, arguing that a manufacturer with lower advertising effort has a greater financial incentive for review fraud with a promotional purpose due to its less well-known reputation and more limited budget for advertising. To conduct the validation test, we first employ a local linear regression within an optimally chosen bandwidth with a cross-validation method for the estimation.

Table 39. RD Effect of Brand Advertising Expenditures on Predetermined Covariates – Local Linear Analysis

| Covariate | Bandwidth | Point Estimate | $p$-val | Robust Inference 95% CI |
|---|---|---|---|---|
| Average price | 155.4 | 2.829 | .318 | [-2.725, 8.384] |
| Average review rate | 155.2 | -0.004 | .945 | [-.129, .120] |

. $p$ <0.1; *$p$ <0.05; **$p$ <0.01; ***$p$ <0.001

Table 40 estimates the specification given by Equation (2) with the full sample and shows that there is a significant effect of brand advertisement level on the incidence of suspicious reviews. As expected, there is a strong impact on suspicious review incidence when the level of brand advertising for a product is low. In both the local linear and local quadratic specifications, the causal effects are highly significant, and the effect is even stronger with the quadratic specification (-7.6%) than with the linear specification (-6.1%).

In Table 40, we also present the results for the model including the two control variables, average price and average review rating, and again find that the negative and significant effect is robust to both local linear and local quadratic estimations with these control variables. These results imply that the outcome level of brand advertising is random, and thus the control variables do not affect the result of the estimation. To verify the robustness of these results, Tables 41 and 42 replicate the analysis reported in Table 40 with the middle 98% and middle 90% of the advertising expenditures, respectively.

To examine the effect on negative manipulative reviews, we re-estimated the same analysis with suspicious reviews with low star ratings. In doing so, we computed the percentage of negative suspicious reviews using the same method as was used for computing the percentage of positive suspicious reviews and used it as a dependent variable. For the cutoff, we repeated the same analysis as above and find that the most significant shift is between an advertising expenditure level of 50%–60% and 60%–70% for the average percentage of negative, suspicious reviews. Using this cutoff and the dependent variable for negative manipulative reviews, we re-estimated the RD analysis. The effect mostly fades away for suspicious reviews with negative purposes (Table 40). The effect of brand advertising effort on negative suspicious reviews is not significant for all the models except for a local linear model with control variables, and effect sizes are much smaller than those for positive suspicious reviews.

Table 40. RD Effect of Brand Advertising Effort on the Incidence of Suspicious Reviews: Local Linear and Polynomial

| Review Type | Model | Bandwidth | Point Estimate | *p*-val | Std. Err | Robust Inference 95% CI |
|---|---|---|---|---|---|---|
| Positive | Full model (linear) | 147.6 | -.061 | .000*** | .009 | [-.079, -.041] |
| | Full model with controls (linear) | 147.6 | -.062 | .000*** | .009 | [-.082, -.043] |
| | Full model (polynomial of order two) | 147.6 | -.076 | .000*** | .011 | [-.099, -.054] |
| | Full model with controls (polynomial of order two) | 147.6 | -.079 | .000*** | .012 | [-.102, -.056] |
| Negative | Full model (linear) | 726.4 | -.008 | .095 | .005 | [-.017, -.001] |
| | Full model with controls (linear) | 726.4 | -.011 | .016* | .005 | [-.020, -.002] |
| | Full model (polynomial of order two) | 726.4 | -.008 | .164 | .006 | [-.020, -.003] |
| | Full model with controls (polynomial of order two) | 726.4 | -.011 | .062 | .006 | [-.023, .0006] |

* $p<.05$; ** $p<.01$; *** $p<.001$

Table 41. RD Effect of Brand Advertising Effort on Suspicious Reviews: Local Linear and Polynomial: Drop Observations with the Lowest 1% and Highest 1% Brand Advertising Expenditure

| Review Type | Model | Bandwidth | Point Estimate | *p*-val | Std. Err | Robust Inference 95% CI |
|---|---|---|---|---|---|---|
| Positive | Full model (linear) | 147.6 | -.060 | .000*** | .009 | [-.080, -.041] |
| | Full model with controls (linear) | 147.6 | -.062 | .000*** | .009 | [-.082, -.043] |
| | Full model (polynomial of order two) | 147.6 | -.076 | .000*** | .011 | [-.099, -.054] |
| | Full model with controls (polynomial of order two) | 147.6 | -.079 | .000*** | .012 | [-.102, -.056] |
| Negative | Full model (linear) | 726.4 | -.008 | .095 | .005 | [-.017, -.001] |
| | Full model with controls (linear) | 726.4 | -.011 | .016* | .005 | [-.020, -.002] |
| | Full model (polynomial of order two) | 726.4 | -.008 | .164 | .006 | [-.020, -.003] |
| | Full model with controls (polynomial of order two) | 726.4 | -.011 | .062 | .006 | [-.023, .0006] |

* $p$<.05; ** $p$<.01; *** $p$<.001

Table 42. RD Effect of Brand Advertising Effort on Suspicious Reviews: Local Linear and Polynomial: Drop Observations with the Lowest 5% and Highest 5% Brand Advertising Expenditure

| Review Type | Model | Bandwidth | Point Estimate | *p*-val | Std. Err | Robust Inference 95% CI |
|---|---|---|---|---|---|---|
| Positive | Full model (linear) | 154.5 | -.060 | .000*** | .009 | [-.079, -.041] |
| | Full model with controls (linear) | 154.5 | -.061 | .000*** | .010 | [-.080, -.042] |
| | Full model (polynomial of order two) | 154.5 | -.073 | .000*** | .011 | [-.096, -.051] |
| | Full model with controls (polynomial of order two) | 154.5 | -.076 | .000*** | .012 | [-.099, -.053] |
| Negative | Full model (linear) | 725.7 | -.008 | .095 | .005 | [-.017, -.001] |
| | Full model with controls (linear) | 725.7 | -.011 | .017* | .005 | [-.020, -.002] |
| | Full model (polynomial of order two) | 725.7 | -.008 | .163 | .006 | [-.020, -.003] |
| | Full model with controls (polynomial of order two) | 725.7 | -.011 | .062 | .006 | [-.023, .0006] |

* $p$<.05; ** $p$<.01; *** $p$<.001

### 3.8.3. Robustness

Products with low demand or low popularity may have more manipulative reviews than the products with high demand or high popularity, because sellers of low-demand, unpopular products may have more of a financial incentive for review manipulation that promotes their products. To rule out this possibility, we re-estimate the full model with only high-demand products and then with just low-demand products. We used the number of reviews for a product as a proxy for its demand (product popularity) and defined products with more than 10 reviews as high-demand or popular products, and those with less than or equal to 10 reviews as low-demand or unpopular products. In column (1) and column (2) of Table 43, we report the results of the two models. Column (1) shows that the results of the model using the unpopular products are consistent with the full model, with a significant effect of brand advertising level on the incidence of suspicious reviews. We also find that the coefficient of the model with popular products is significant, meaning that the effect of brand advertising level also holds with the subset of high-demand or popular products. We further analyze the model with subsets whose brand advertising expenditure is within 30%, 25%, 20%, and 10% of the majority threshold. Columns (3) through (6) present that the results, which are consistent with the analysis from the full model, although the effect sizes become weaker as the data sizes decrease, likely due to the smaller number of observations.

## 3.9. Beta Regression Model and Category Effect

In addition to the RDD method, we explore the relationship between advertising expenses and the rate of spam using a beta regression model. Beta regression is frequently used when an outcome variable is a proportion. The goal of this section is to determine the effect of brand advertisement level on the incidence of review fraud. To achieve this goal, the same variables used in the RD model were incorporated into the model described here: brand-level advertising expenditures for brand advertisement and the percentage of manipulative reviews per product for a dependent variable ($Y_i \in [0; 1]$). To differentiate between positive and negative reviews, we divide the data into positive reviews with 4 or 5 stars and negative reviews with 1 or 2 stars and estimate the models separately. Table 44 presents the descriptive statistics for the variables that were used in the beta regression.

Table 43. RD Effect of Brand Advertising Effort on Positive Suspicious Reviews – Robustness Checks

| | No. Reviews | | Advertising Expenditure | | | |
|---|---|---|---|---|---|---|
| | ≤ 10 | > 10 | ± 30% | ±25% | ±20% | ±10% |
| Estimate (Suspicious %) | -.057*** | -.066* | -.063*** | -.079 | -.039*** | -.036*** |
| *p*-value | .000 | .031 | .000 | .000 | .000 | .000 |
| (Std error) | (.011) | (.030) | (.010) | (.012) | (.007) | (.006) |
| Number of Observations | 11,591 | 795 | 7,490 | 6,196 | 4,948 | 2,468 |

. $p$ <0.1; *$p$ <0.05; **$p$ <0.01; ***$p$ <0.001

Other than the univariate model in which the relationship between the proportion of manipulative reviews and brand advertisement level is estimated, we also estimate the product category effect in separate models. We cluster the product categories in two different ways. First, the products are clustered based on product involvement (low, medium, or high). Price level is used as a proxy for the product involvement, and three involvement levels with two cutoffs are created based on the number of data points. The total number of data points is 272,338. In the data set sorted based on price level, the first cutoff is at 90,778 observations, which corresponds to 13.5 dollars.

Table 44. Descriptive Statistics for Beta Regression

| Variable | N | Mean | Std. dev. | Min | Max | Skewness |
|---|---|---|---|---|---|---|
| Positive Reviews | | | | | | |
| Advertising Expenditure | 38,196 | 15,186.37 | 30,020.71 | 0.00 | 634,128.00 | 4.29 |
| Positive Suspicious Reviews (%) | 70,000 | 0.09 | 0.11 | 0.00 | 0.83 | 2.48 |
| Negative Reviews | | | | | | |
| Advertising Expenditure | 5,459 | 16,680.68 | 36,221.15 | 0.00 | 634,128.00 | 4.95 |
| Negative Suspicious Reviews (%) | 9,519 | 0.09 | 0.11 | 0.00 | 0.80 | 2.39 |

The second cutoff is at 181,556 observations, which corresponds to 32.5 dollars. Products priced above 32.5 dollars are clustered into the high level of product involvement group. Also, products are clustered as tech or non-tech products. Among the 17 product categories, electronics, cellphones, and software are categorized as tech products, while the others are clustered into the non-tech group.

The models are separately estimated. One model incorporates the categorical variables for product involvement (low, medium, and high product involvement). In this model, we estimate the interaction effect between product involvement and brand advertisement level to observe the dependency between the two independent variables. In another model, the dummy variable for tech products is incorporated with the interaction effect between these two independent variables. All estimations are performed based on beta regression models with both logit link and log-log link.

The results for positive and negative reviews are reported in Tables 45 and 46. The univariate models for both positive and negative reviews reveal the negative relationship between advertisement level and the incidence of review fraud. The results of the RD models are the same for the beta regressors, implying that a lower level of brand advertisement is significantly related to an increase in manipulative reviews with both positive and negative intentions.

For models incorporating the level of product involvement, a high level of product involvement is set as a reference level. The results consistently show that the effect of a low (medium) level of product involvement on the incidence of manipulative reviews is significantly higher than the effect of a high level of product involvement on the outcome variable for all four models (positive:logit, positive:loglog, negative:logit, and negative:loglog). This implies that review manipulation occurs when there is a low (medium) level of product involvement significantly more often than when there is a high level of product involvement. But the effects of product involvement on the incidence of manipulative reviews are not the same for every level of advertising expenses. The coefficients of the interaction terms explain that, for every one-unit increase in advertising expenses, the incidence of review fraud significantly decreases for a low (medium) level of product involvement, although the significant effects for interaction terms are weaker or fade out for negative reviews (Table 46).

We also examine the effects of product involvement on the incidence of manipulative reviews and determine whether they are different for tech vs. non-tech products. Columns (5) and (6) in Tables 45 and 46 display the results regarding this topic. For the main effects, review manipulation occurs at a significantly lower rate for tech-related products than with for non-tech products.

Table 45. Beta Regression Results: Positive Reviews

| | Positive reviews | | | | | |
|---|---|---|---|---|---|---|
| | Beta regression with logit link (1) | Beta regression with loglog link (2) | Beta regression with logit link (3) | Beta regression with loglog link (4) | Beta regression with logit link (5) | Beta regression with loglog link (6) |
| Advertising expenses | -1.406***<br>(-9.39) | -0.518***<br>(-9.92) | -0.070***<br>(-0.31) | -0.027***<br>(-0.33) | -0.880***<br>(-4.08) | -0.324***<br>(-4.04) |
| Involvement:low | | | 0.107***<br>(8.88) | 0.040***<br>(3.94) | | |
| Involvemen:medium | | | 0.069***<br>(5.96) | 0.025***<br>(5.90) | | |
| Advertising expenses* Involvement:low | | | -3.781***<br>(-10.31) | -1.406***<br>(-10.75) | | |
| Advertising expenses* Involvement:medium | | | -0.952***<br>(-2.74) | -0.327***<br>(2.65) | | |
| Technology | | | | | -0.492***<br>(-49.11) | -0.178***<br>(-50.10) |
| Advertising*Technology | | | | | 0.309<br>(1.05) | 0.124<br>(1.18) |
| Pseudo R-squared | 0.004 | 0.004 | 0.01 | 0.009 | 0.12 | 0.10 |
| N | 38,196 | 38,196 | 38,196 | 38,196 | 38,196 | 38,196 |

. *p*<.05; **p*<.01; ***p*<.001; ****p*<.0001

Table 46. Beta Regression Results: Negative Reviews

| | Negative reviews | | | | | |
|---|---|---|---|---|---|---|
| | Beta regression with logit link (1) | Beta regression with loglog link (2) | Beta regression with logit link (3) | Beta regression with loglog link (4) | Beta regression with logit link (5) | Beta regression with loglog link (6) |
| Advertising expenses | -2.061***<br>(-6.11) | -7.209***<br>(-6.52) | -1.362**<br>(-2.85) | -0.496**<br>(-3.04) | -0.994.<br>(-1.95) | -0.323.<br>(-1.76) |
| Involvement:low | | | 0.281***<br>(8.83) | 0.105***<br>(8.82) | | |
| Involvemen:medium | | | 0.072*<br>(2.38) | 0.024*<br>(2.16) | | |
| Advertising expenses* Involvement:low | | | -1.976.<br>(-1.91) | -0.789*<br>(-2.14) | | |
| Advertising expenses* Involvement:medium | | | -0.779<br>(-1.06) | -0.132<br>(-9.56) | | |
| Technology | | | | | -0.527***<br>(-20.07) | -0.189***<br>(-20.50) |
| Advertising*Technology | | | | | -0.504<br>(-0.75) | -0.170<br>(-0.74) |
| Pseudo R-squared | 0.004 | 0.004 | 0.01 | 0.009 | 0.12 | 0.10 |
| N | 9,519 | 9,519 | 9,519 | 9,519 | 9,519 | 9,159 |

*p*<.05; **p*<.01; ***p*<.001; ****p*<.0001

## 3.10. Conclusions

The direct influence of online reviews on purchasing decisions make opinion-sharing platforms attractive targets for review fraud. We use Amazon.com review data from 16 product categories to investigate factors that characterize reviews that consumers would find suspicious, concentrating on the unstructured text of a review. We then investigate a weak brand as an incentive for engaging in review fraud.

We do not directly know what product reviews are spams, so a labeling process which classifies reviews as suspiciously manipulative or genuine must precede the econometric analyses in our study. Due to the large number of reviews, the labeling procedure first uses human investigators for a subset of the reviews and then builds a semi-supervised classifier to label the full dataset. We find that approximately 3% of the product reviews on Amazon.com would be viewed as suspicious by consumers.

After building the labeled data, we empirically analyze factors that characterize product reviews as suspicious versus genuine and explore several classifiers and models to find the model that has the best predictive power. A large body of research on spam reviews focuses on fraud detection algorithms, and most of these studies focus on a review's numeric information (e.g., star rating) and on simple textual features such as basic sentiment polarity or text similarity, largely ignoring the more elaborate semantic information from the text data. Finally, we investigate an important factor that is associated with review manipulation, finding that brand advertising effort is related to the incidence of positive (promotional) review fraud.

The analyses reveal several interesting findings. First, suspicious product reviews are more likely to have extreme ratings rather than mediocre ratings, tend to have a lower word count, and are more likely to be created by users who post more reviews.

Second, emotionality accounts for potential product review manipulation. We find that emotions with a clear positive or negative valence better describe suspicious reviews than neutral or low arousal emotions. Specifically, our analysis shows that fear and joy, which are extreme emotions, are positively related to the likelihood that a consumer views as review as suspicious, while mixed emotions, such as anticipation and surprise, are less likely to be associated with opinions suspected of being manipulative. More extreme reviews tend to have a greater proportion of emotional content (Ullah et al. 2016), and we find an effect of emotional content after accounting for a review's star ratings.

Third, we find that semantic values from topic models further improve the predictive power of the models. We examine various machine-learning classifiers (Logistic classification, SVM, Random Forest, and Deep Learning) and find that, overall, models that incorporate semantic features have superior predictive performance.

We compare a Baseline Model that includes only numeric features with models that also include emotionality and topic distribution values and confirm the superior predictive power of the latter models for several machine-learning classifiers.

Finally, we discover an effect of brand advertising on the incidence of suspicious reviews. A quasi-experimental method (RD) reveals that a low level of brand advertising is associated with greater incidence of suspicious reviews with a promotional purpose. The effect does not hold for suspicious reviews with defamatory purposes.

The results have several managerial and academic implications. First, the findings from our analysis should be of interest to firms where a recommendation system is their main business operation. In recent years, fake reviews have become critical to the quality control aspect of review platforms. Information regarding the characteristics of manipulative reviews that are offered from our empirical analysis and the models with improved forecasting power can help managers at these firms to better control review quality and maintain their platforms' reputation.

Second, interested groups can directly utilize the labeling procedure that we establish to classify reviews as manipulative or genuine. In an era of enormous data culled from many different sources (e.g., social media, chats in online forums), academic researchers and business practitioners frequently deal with very large unlabeled data. Labeling with manual annotation is time consuming and expensive. In these cases, we recommend a semi-supervised learning method for labeling the data. Semi-supervised methods can be widely applied to many areas in business where there are large amounts of unlabeled data and human annotation is not feasible. Examples in business research include natural language processing, where a researcher needs to group and label certain features such as product aspects, or influencer detection in social media or in online communities. Crowdsourcing methods such as Amazon Mechanical Turk are often used to coordinate the use of human intelligence to amass a large number of responses. This method, however, is not always not feasible. Some annotations require expert opinions or cautious inspections, like our case, and thus only small amounts of labeling can be done using these procedures because of limited resources. We recommend that researchers and practitioners apply our method in these practices.

Human investigation and semi-supervised classification can be directly applied to different events or business industries. The application of our method is not limited to the spam detection domain and can be used in any context where a labeling process is necessary due to the absence of related variables. Further, the methodology and the RD procedure that we employ should be of interest to the academic researchers who have concerns about endogeneity or selection bias issues in their analyses but have limited resources to conduct a randomized experiment.

Finally, our study has certain limitations. First, the data for semi-supervised classifiers in Section 3.4.4. is created in the form of balanced data. The balanced data is based on number of reviewers (i.e., 50 percent from genuine reviewers and 50 percent from suspicious reviewers), but future model can be conducted with balanced data based on number of reviews instead of number of reviewers. Second, the analyses rely on the models that mainly consider conditional mean. Future research will employ a quantile regression which is more robust to non-normal errors and outliers and also allows us to consider the impact of a covariate on the entire distribution of *y*, not merely its conditional mean. Third, in order to create a training set for human evaluators, we relied on a spam detection algorithm with a research setting similar to ours, but that approach can be replaced with another one contingent upon the research setting or business goal of the interested group. We focus on the effect of brand advertising effort on review manipulation, but there might be other factors that drive the malpractice such how recently the product launch was launched. Other than the main effects, there may be moderating or mediating variables that strengthen or weaken the outcome of interest. For instance, the effect of brand advertising effort on review fraud may be mediated by the product category (e.g., electronics or beauty). We also suspect that there are different reasons for creating negative review spam, but we do not examine this issue. For example, a recent increase in the reputation of a competitor may motivate review fraud with a defamatory purpose against that competitor. We hope that future research can investigate the financial incentives of manipulative reviews with negative purposes. Finally, the results are based on the analysis of tangible products, so we cannot directly generalize the results to other sectors. We hope that future studies using other datasets can be helpful in examining the effects that lie beyond the scope of our study.

# Chapter 4

## Conclusions

Over two essays, my dissertation addresses the recent challenges that marketing managers face when consumers frequently migrate from one media platform to another, causing strategic and informed marketing decisions to become more challenging issues. I particularly investigate the content characteristics that featurize user-generated content (UGC), both from structured and from unstructured data, and their associations with consumer responses and economic incentives. In doing so, I employ artificial intelligence (AI) techniques (e.g., machine learning, deep learning) and econometric models that can efficiently analyze large-scale field data and that automate many of the processes which would otherwise have to be done manually.

Although a wide range of studies examine the relationships between digital consumer traces and market performance, most of the existing empirical research focuses on UGC's direct link to performance instead of focusing on Smotivations. In addition, the research that examines the semanticity of UGC pays limited attention to content beyond the text. Apart from other research, this dissertation carefully extracts content features both from texts and from images and actively seeks out the relationships between the semantic characteristics and various behavioral and economic measures. The two essays that constitute my dissertation collectively argue that commercializing UGC may create good opportunities for firms and brands. For example, my first essay identified the bimodal effect of visual complexity on consumer engagement. However, the semantic aspects of UGC are differentially associated with idiosyncratic contextual characteristics, such as product category or types of social media posts. Visual sentiment, for example, is a positive predictor of consumer engagement for Instagram posts that are created by general users, whereas it is a negative predictor for commercial-related posts. In the second essay, I argue that the effect of the advertising expense level on review manipulation is different for technology-related products and non-technology products; it is higher for non-technology products. Practically, these findings imply that marketing managers need to create different UGC mixes or UGC activation plans in different contexts.

Future work will look at more content features and their roles beyond the UGC-level. At aggregated levels, consumers or business profiles, combined with content characteristics from UGC, may offer lucrative opportunities for firms to gain in-depth customer knowledge that can be leveraged across different medial channels. Future research will investigate further contextual effects. The semantic effects on consumer engagement, for example, may not be the same across all brands. Brands are frequently categorized based on brand personality or strength, and the content effects on consumers' behavioral metrics may vary according to the different types of brands. My future research will consider such contextual effects across more diverse data-sets.

Bibliography

Albuquerque P, Pavlidis P, Chatow U, Chen KY, Jamal Z (2012) Evaluating promotional activities in an online two-sided market of user-generated content. *Marketing Sci.* 31(3):406-432.

Al-Naser M, Chanijani SSM, Bukhari SS, Borth D, Dengel A (2015) What makes a beautiful landscape beautiful: Adjective noun pairs attention by tracking and gaze analysis. *Proc. 1st Intetnat. Workshop on Affect and Sentiment in Multimedia* (ACM, New York), 51-56.

Alvarez GA, Cavanagh P (2004) The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Sci.* 15(2):106-111.

Ana BR, Francesca S, Kristine V, Tammo HAB (2016) The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *J. Marketing Res.* 53(3):297-318.

Anderson ET, Duncan IS (2014) Reviews without a purchase: Low ratings, loyal customers, and deception. *J. Marketing Res.* 51(3):249-269.

Aral S (2013) The Problem with Online Ratings. *MIT Sloan Management Review.* 55(2):47-52.

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

Arjun M, Kumar A, Liu B, Wang J, Hsu M, Castellanos M, Ghose R (2013) Spotting opinion spammers using behavioral footprints. *Proc. 19th ACM SIGKDD Internat. Conf. Knowledge Discovery and Data Mining* (KDD, Chicago, IL), 632-640.

*Bazaarvoice* (2012) Talking to strangers: Millennials trust people over brands (January), Retrieved February 5, 2019, http://media2.bazaarvoice.com/documents/Bazaarvoice_WP_Talking-to-Strangers.pdf

Berger J, Milkman K (2012) What makes online content viral? *J. Marketing Res.* 49(2):192-205.

Berglund B, Berglund U, Engen T (1982) Multi-dimensional analysis of emotions expressed in photo graphs and words. *University of Stockholm*.

Berlyne D (1974) *Studies in the New Experimental Aesthetics* (Hemisphere Pub. Corp., Washington, DC).

Bhattacharya N, Desai H, Venkataraman K (2013) Does earnings quality affect information asymmetry? Evidence from trading costs. *Contemporary Accounting res*. 30(2):482-516.

Bhattacharya S, Nojavanasghari B, Chen T, Liu D, Chang SF, Shah M (2013) Towards a comprehensive computational model for aesthetic assessment of videos. *Proc. 21st ACM Internat. Conf. Multimedia* (ACM, New York), 361-364.

*Bloglovin* (2016) How micro-influencers really want to work with brands? (November), Retrieved May 1, 2018, https://www.activate.social/docs/whitepaper/Bloglovin_Global_Influencer_Survey_2016.pdf

Borth D, Chen T, Ji R, Chang SF (2013) SentiBank: Large-scale ontology and classifiers for detecting sentiment and emotions in visual content. *Proc. 21st ACM Intern. Conf. Multimedia* (ACM, New York), 459-460.

BrightLocal (2017) Local consumer review survey (June), Retrieved June 8, 2018, https://www.brightlocal.com/learn/local-consumer-review-survey/

Brodie RJ, Ilic A, Juric B, Hollebeek L (2015) Consumer engagement in a virtual brand community: An exploratory analysis. *J. Business Res.* 66(1):105-114.

Cappallo S, Mensink T, Snoek CGM (2015) Latent factors of visual popularity prediction. *Proc. 5th ACM Intern. Conf. Multimedia* (ACM, New York), 195-202.

Chen M, Zhang L, Allebach JP (2015) Learning deep features for image emotion classification. *2015 IEEE Intern. Conf. Image Processing* (ICIP, Quebec City), 4491-4495.

Chen T, Borth D, Darrell T, Chang SF (2014) Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. *arXiv preprint arXiv:1410.8586*.

Chen YC, Shang R, Kao CY (2008) The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment. *Electronic Commerce Res. Appl.* 8:48-58.

Cheung C, Lee M, Jin X (2011) Customer engagement in an online social platform: A conceptual model and scale development. *Proc. 32nd Internat. Conf. Inform. Systems* (Shanghai, China), 1-8.

Cheng CM, Shen XL, Lee ZW, Chan TK (2015) Promoting sales of online games through customer engagement. *Electronic Commerce Res. Appl.* 14(4):241-250.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(3):345-354.

Cho H, Schwarz N (2006) If I don't understand it, it must be new: Processing fluency and perceived product innovativeness. *ACR North American Advances*.

Chwialkowska A (2018) The effectiveness of brand- and customer-centric content strategies at generating share, 'likes', and comments. *J. Promotion Management* 15:280-300.

Colicev A, Malshe A, Pauwels A, O'Connor P (2018) Improving consumer mindset metrics and shareholder value through social media: The different roles of owned and earned media. *J. Marketing* 82(1):37-56.

Coursaris CK, van Osch W, Balogh BA (2016) Do Facebook likes lead to shares or sales? Exploring the empirical links between social media, content, brand equity, purchase intention, and engagement. *49th Internat. Conf. System Sciences* (HICSS 16), 3546-3555.

Cox DS, Cox AD (1988) What does familiarity breed? Complexity as a moderator of repetition effects in advertisement evaluation. *J. Consumer Res.* 15(1):111-116.

Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Najada HA (2015) Survey of review spam detection using machine learning techniques. *J. Big Data* 2(23):1-24.

Davis SF, Palladino JJ, Christopherson K (2012) *Psychology 7th Edition* (Pearson Education, Upper Saddle River, NJ).

Deng J, Berg A, Satheesh S, Su H, Khosla A, Fei-Fei L (2012) Imagenet large scale visual recognition competition 2012. *Google Scholar* (ILSVRC2012).

Deng L, Poole MS (2010) Affect in web interfaces: A study of the impacts of web page visual complexity and order. *MIS Quarterly* 34(4):711-730.

Dewan P, Bharadhwaj V, Mithal A, Suri A, Kumaraguru P (2016) Visual themes and sentiment on social networks to aid first responders during crisis events. *arXiv preprint arXiv:1610.07772.*

Dewang RK, Singh AK (2018) State-of-art approaches for review spammer detection: A survey. *J. Intelligent Information Systems* 50:231-64.

Dixit S, Agrawal AJ (2013) Survey on review spam detection. *Internat. J. Computer & Communication Technology* 4(2):68-72.

Dolinsky C, Feinberg RA (1986) Linguistic barriers to consumer information processing: Information overload in the Hispanic population. *Psych & Marketing* 3(4):261-271.

Donderi DC (2006) Visual complexity: A review. *Psychological Bulletin* 132(1):73-97.

Engen T, Levy N, Schlosberg H (1958) The dimensional analysis of a new series of facial expressions. *J. Experimental Psych.* 55(5):454-458.

Eric G, Karahalios K (2010) Understanding deja reviewers. *Proc. 2010 ACM Conf. Computer Supported Cooperative Work* (ACM, New York), 225-228.

Fan S, Jiang M, Shen Z, Koenig BL, Kankanhalli MS, Zhao Q (2017) The role of visual attention in sentiment prediction. *Proc. 2017 ACM Multimedia Conf.* (ACM, New York), 217-225.

Filieri R (2016) What makes an online consumer review trustworthy? *Annals of Tourism Res.* 58:46-64.

Fontaine JR, Scherer KR, Roesch EB, Ellsworth PC (2007) The world of emotions is not two-dimensional. *Psychological Sci.* 18(12):1050-1057.

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

Frijda NH (1969) Recognition of emotion. *Advances in Experimental Social Psyh.* 4:167-223.

Geissler GL, Zinkhan GM, Watson RT (2006) The influence of home page complexity on consumer attention, attitudes, and purchase intent. *J. Advertising* 35(2):69-80.

Gelli F, Uricchio T, Bertini M, Del Bimbo A, Chang SF (2015) Image popularity prediction in social media using sentiment and context features. *Proc. 23rd ACM Intern. Conf. Multimedia* (ACM, New York), 907-910.

Gensler S, Völckner F, Liu-Thompkins Y, Wiertz C (2013) Managing brands in the social media environment. *J. Interactive Marketing* 27(4):242-256.

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

Goh KY, Heng CS, Lin Z (2013) Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Inform. Systems Res.* 24(1):88-107.

Goldberg JH (1999) Visual search of food nutrition labels. *Human Factors* 41(3):425-437.

Gorn GJ, Goldberg ME, Basu K (1993) Mood, awareness and product evaluation. *J. Consumer Psych.* 2(3):237-256.

Hahn J, Todd P, Klaauw WV (2001) Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica* 69(1):201-209.

Heydari A, Tavakoli M, Salim N, Heydari Z (2015) Detection of review spam: A survey. *Expert Systems with Applications* 42(7):615-635.

Ho-Dac, NN, Stephen JC, Moore WL (2013) The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *J. Marketing*, 77(6):37-53.

Hu N, Bose I, Koh NS, Liu L (2012) Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems* 52(3):674-684.

Isola P, Parikh D, Torralba A, Oliva A (2011) Understanding the intrinsic memorability of images. *In Advances in Neural Information Processing Systems*, 2429-2437.

Jacob R, Zhu P, Somers MA, Bloom HS (2012) A practical guide to regression discontinuity, working paper.

Jensen Schau H, Gilly MC (2003) We are what we post? Self-presentation in personal web space. *J. Consumer Res.* 30(3):385-404.

Jia Y, Shelhamer E, Donahue J, Karayev S, Karayev S, Long J, Ross G, Guadarrama S, Darrell T (2014) Convolutional architecture for fast feature embedding. *Proc. 22nd CAN Internat. Conf. Multimedia* (ACM, New York), 675-678.

John LK, Emrich O, Gupta S, Norton MI (2017) Does 'Liking' lead to loving? The impact of joining a brand's social network on marketing outcomes. *J. Marketing Res.* 54(1):144-155.

Jou B, Chen T, Pappas N, Redi M, Topkara M, Chang SF (2015) Visual affect around the world: A large-scale multilingual visual sentiment ontology. *Proc. 23rd ACM Internat. Conf Multimedia* (ACM, New York), 159-168.

Kalay K, Moses SC, Ashok K (2015) A survey on the design, fabrication and utilization of different crops planter. *Eur. Acad. Res.* 3:147-158.

Kalayeh MM, Seifu M, LaLanne W, Shah M (2015) How to take a good selfie? *Proc. 23rd ACM Internat. Conf. Multimedia* (ACM, New York), 923-926.

Kerr WR, Lerner J, Schoar A (2014) The consequences of entrepreneurial finance: Evidence from angel financings. *Review of Financial Studies* 27(1):20-55.

Kopec DA (2006) *Environmental Psychology for Design* (Fairchild Books, New York, NY).

Kossly S (1975) Information representation in visual images. *Cognitive Psych.* 7(3):341-370.

Kristopher F, Freling R, Alhoqail S, Cho HY, Freling T (2014) How online product reviews affect retail sales: A meta-analysis. *J. Retailing* 90(2):217- 232.

Kumar A, Bezawada R, Rishika R, Janakiraman R, Kannan PK (2016) From social to sale: The effects of firm-generated content in social media on customer behavior. *J. Marketing* 80(1):7-25.

Lee D, Hosanagar K, Nair HS (2018) Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Sci.* forthcoming.

Lev M, Aral S, Taylor SJ (2013) Social influence bias: A randomized experiment. *Science* 341(6146):647-51.

Li, Q, Huang Z, Christianson K (2016) Visual attention toward tourism photographs with text: An eye-tracking study. *Tourism Management* 54:243-258.

Li S, Wang Z, Zhou G, Lee SYM (2011) Semi-supervised learning for imbalanced sentiment classification. *Proc. 2nd Internat. Joint Conference on Artificial Intelligence* (Barcelona, Catalonia), 1826-1831.

Lim EP, Nguyen VA, Jindal N, Liu B, Lauw HW (2010) Detecting product review spammers using rating behaviors. *Proc. 19th ACM Internat. Conf. Information Knowledge Management* (ACM, New York), 939-948.

Linshi J (2014) Personalizing yelp star ratings: A semantic topic modeling approach. *Yale University*.

Liu L, Dzyabura D, Mizik N (2017) Visual listening in: Extracting brand image portrayed on social media. Available at SSRN: https://ssrn.com/abstract=2978805.

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

Lovato P, Perima A, Sebe N, Omar Z, Montagnini A, Bicego M, Cristani M (2013) Tell me what you like and I'll tell you what you are: Discriminating visual preferences on Flickr data. *Asian Conf. Computer Vision* (Berlin, Beidelberg), 45-56.

Lovett M, Peres R, Shachar R (2013) On brands and word of mouth. *J. Marketing Res.* 50(4):427-444.

Lovett M, Peres R, Shachar R (2014) A data set of brands and their characteristics. *Marketing Sci.* 33(4):609-617.

Luca M, Zervas G (2016) Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Sci.* 62(12):3412-3434.

Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390(6657):279-281.

Malhotra NK, Jain AK, Lagakos SW (1982) The information overload controversy: An alternative viewpoint. *J. Marketing* 46:27-37.

Malthouse EC, Haenlein M, Shiera B, Wege E, Zhang M (2013) Managing customer relationships in the social media era: Introducing the social CRM house. *J. Interactive Marketing* 27(4):270-280.

Mayzlin D, Dover Y, Chevalier J (2014) Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review* 104(8):2421-55.

McAuley J, Leskovec J (2013) Hidden factors and hidden topics: Understanding rating dimensions with review text. *Proc. 7th ACM Conf. Recommender System* (ACM, New York), 165-172.

McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: A density test. *J. Econometrics* 142(2):698-714.

Michailidou E, Harper S, Bechhofer S (2008) Visual complexity and aesthetic perception of web pages. *Proc. 26th annual ACM Internat. Conf. Design of Communication* (ACM, New York), 215-224.

Mochon D, Johnson K, Schwartz J, Ariely D (2017) What are likes worth? A Facebook page field experiment. *J. Marketing Res.* 54(2):306-317.

Mohammad AN, Chanijani SSM, Bukhari SS, Borth D, Dengel A (2015) What makes a beautiful landscape beautiful: Adjective noun pairs attention by eye-tracking and gaze analysis. *Proc. 1st Internat. Workshop Affect & Sentiment in Multimedia* (ACM, New York), 51-56.

Munzel A (2015) Malicious practice of fake reviews: Experimental insight into the potential of contextual indicators in assisting consumers to detect deceptive opinion spam. *Recherche et Applications en Marketing* 40(4):24-50.

Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Sci.* 31(3):521-543.

Newell A, Simon HA (1972) *Human problem solving* (Prentice-Hall, Englewood Cliffs, NJ).

Oh C, Roumani Y, Nwankpa JK, Hu HF (2017) Beyond likes and tweets: Consumer engagement behavior and movie box office in social media. *Inform. & Management* 54(1):25-37.

Palmer SE (1999) *Vision science: Photons to phenomenology* (Cambridge, MA).

Pang B, Vaithyanathan S (2002) Thumbs up: Sentiment classification using machine learning techniques. *Proc. ACL-02 Conf. Empirical Methods in Natural Language Processing* (EMNLP 10), 79-86.

Payne JW (1976) Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance* 16(2):366-387.

Peracchio LA, Meyers-Levy J (2005) Using stylistic properties of ad pictures to communicate with consumers. *J. Consumer Res.* 32(1):29-40.

Percy L, Rossiter JR (1983) Effects of picture size and color on brand attitude responses in print advertising. *Advances Consumer Res.* 10:17-20.

Peters K, Chen Y, Kaplan AM, Ognibeni B, Pauwels K (2013) Social media metrics – A framework and guidelines for managing social media. *J. Interactive Marketing* 27(4):281-298.

Pfitzner R, Garas A, Schweitzer F (2012) Emotional divergence influences information spreading in Twitter. *Proc. 6th Internat. AAAI Conf. Weblogs and Social Media* (ICWSM 12), 2-5.

Pieters R, Warlop L (1999) Visual attention during brand choice: The impact of time pressure and task motivation. *Internat. J. Res. Marketing* 16(1):1-16.

Pieters R, Warlop L, Wedel M (2002) Breaking through the clutter: Benefits of advertisement originality and familiarity for brand attention and memory. *Management Sci.* 48(6):765-781.

Pieters R, Wedel M (2007) Goal control of attention to advertising: The Yarbus implication. *J. Consumer Res.* 34(3):224-233.

Pieters R, Wedel M, Zhang J (2007) Optimal feature advertising design under competitive clutter. *Management Sci.* 53(11):1815-1828.

Pieters R, Wedel M, Batra R (2010) The stopping power of advertising: Measures and effects of visual complexity. *J. Marketing* 74(5):48-60.

Plutchik R (1980) *Emotion: A psychoevolutionary synthesis* (Harper & Row, New York, NY).

Plutchik R, Kellerman H (1980) *Emotion: Theory, Research, and Experience: Vol.1 Theories of Emotion* (Academic Press, Cambridge, MA).

Rayner K, Miller B, Rotello CM (2008) Eye movements when looking at print advertisements: The goal of the viewer matters. *Applied Cognitive Psych.* 22(5):697-707.

Rayner K, Rotello CM, Stewart RA, Keir J, Duffy SA (2001) Integrating text and pictorial information: Eye movements when looking at print advertisements. *J. Experimental Psych.* 7(3):219-226.

Reinecke K, Bernstein A (2013) Knowing what a user likes: A design science approach to interfaces that automatically adapt to culture. *MIS Quarterly* 37(2):427-453.

Reisenzein R (1994) Pleasure-arousal theory and the intensity of emotion. *J. Personality and Social Psych.* 67(3):525-539.

Russell N, Cribbin L, Murphy TB (2012) Upclass: An R package for updating model-based classification rules, working paper.

Ryan D, Ansari A, Toubia O (2017) Letting logos speak: Deep, probabilistic models for logo design, working paper.

Saeideh B, Shamma DA, Gilbert E (2014) Faces engage us: photos with faces attract more likes and comments on Instagram. *Proc. SIGCHI Conf. Human Factors in Computing Systems* (ACM, New York), 965-974.

Saxton GD, Niyirora JN, Guo C, Waters RD (2015) #AdvocatingForChange: The strategic use of hashtags in social media advocacy. *Advances in Social Work* 16(1):154-160.

Schuckert M, Lui X, Law R (2016) Insights into suspicious online ratings: Direct evidence from TripAdvisor. *Asia Pacific J. Tourism Res.* 21(3):259-272.

Schwarz N, Clore G (1983) Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *J. Personality and Social Psych.* 45(3):513-523.

Scholsberg H (1954) Three dimensions of emotion. *Psychological Rev.* 61(2):81-88.

*Scrunch* (2016) Influencer marketing case study: Launching a new product with HP Australia. (September), Retrieved May 1, 2018, https://blog.scrunch.com/case-study-hp-australia.

Shiyang G, Zhang J, Zhao P, Jiang X (2017) Tweeting as a marketing tool: A field experiment in the TV industry. *J. Marketing Res.* 54(6):833-850.

Shojaee S, Murad MAA, Azman AB, Sharef NM, Nadali S (2013) Detecting deceptive reviews using lexical and syntactic features. *Proc. 13th Internat. Conf. Intelligent Systems Design and Apps* (Kembangan, Selangor), 53-58.

Skovron C, Titiunik R (2015) A practical guide to regression discontinuity designs in political ccience, working paper.

*Socialbakers* (2014) How social engagement drives site visits. (December 16), Retrieved May 1, 2018, https://www.socialbakers.com/blog/2320-how-social-engagement-drives- site-visits.

Sonnier GP, McAlister L, Rutz OJ (2011) A dynamic model of effect of online communications on firm sales. *Marketing Sci.* 30(4):702-716.

Sprott D, Czellar S, Spangenberg E (2009) The importance of a general measure of brand engagement on market behavior: Development and validation of a scale. *Marketing Res.* 46(1):92-104.

Stieglitz S, Xuan LD (2013) Emotions and information diffusion in social media-sentiment of microblogs and sharing behavior. *J. Management Inform. Systems* 29(4):217-248.

Stonedahl F, Rand W, Wilensky U (2010) Evolving viral marketing strategies. *Proc. 12th Annual Conf. Genetic and Evolutionary Computation* (ACM, New York), 1195-1202.

Sun M, Zhu F (2013) Ad revenue and content commercialization: Evidence from blogs. *Management Sci.* 59(10):2314-2331.

Tatemura J (2000) Virtual reviewers for collaborative exploration of movie reviews. *Proc. 5th Intenat. Conf. Intelligent User Interfaces* (ACM, New York), 272-275.

Teixeira T, Picard R, el Kaliouby R (2014) Why, when, and how much to entertain consumers in advertisements? A web-based facial tracking field study. *Marketing Sci.* 33(6):809-827.

Terveen LG, Hill WC, Amento B, McDonald D, Creter J (1997) Building task-specific interfaces to high volume conversational data. *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems* (ACM, New York), 226-233.

Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A (2010) Sentiment strength detection in short informal text. *J. Association for Information Sci.* Technology 61(12):2544-2558.

Theodoros L, Sabnis G, Valkanas G (2016) The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Res.* 27(4):940-961.

Toubia O, Stephen AP (2013) Intrinsic vs. image-related utility in social media: Why do people contribute content to twitter? *Marketing Sci.* 32(3):368-392.

Townsend C, Kahn BE (2014) The 'Visual Preference Heuristic': The influence of visual versus verbal depiction on assortment processing, perceived variety, and choice overload. *J. Consumer Res.* 40(1):993-1015.

Triandis HC, Lambert WW (1958) A restatement and test of Schlosberg's theory of emotion with two kinds of subjects from Greece. *J. Abnormal and Social Psych.* 56(3):321-328.

Ullah R, Amblee N, Kim W, Lee H (2016) From valence to emotions: Exploring the distribution of emotions in online product reviews. *Decision Support Systems* 81:41-53.

Wagner C, Strohmaier M (2010) The wisdom in tweetonomies: Acquiring latent conceptual structures from social awareness streams. *Proc. 3rd Internat. Semantic Search Workshop* (ACM, New York).

Wallace GK (1992) The JPEG still picture compression standard. *IEEE Transactions on Consumer Electronics* 38(1):18-34.

*Wall Street Journal* (2016) When shopping online, can you trust the reviews? (November 29), Retrieved May 14, 2018, www. wsj.com

Wan J, Wang D, Hoi SCH, Wu P, Zhu J, Zhang Y, Li J (2014) Deep learning for content-based image retrieval: A comprehensive study. *Proc. 22nd ACM Internat. Conf. Multimedia* (ACM, New York), 157-166.

Wedel M, Pieters R (2000) Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Sci.* 19(4):297-312.

Xiao L, Ding M (2014) Just the faces: Exploring the effects of facial features in print advertising. *Marketing Sci.* 33(3):338-352.

Xu C, Cetintas S, Lee KC, Li LJ (2014) Visual sentiment prediction with deep convolutional neural networks. *arXiv preprint arXiv:1411.5731.*

Xu J, Forman C, Kim JB, Ittersum KV (2014) News media channels: Complements or substitutes? Evidence from mobile phone usage. *J. Marketing* 78(4):97-112.

Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440(7080):91-95.

Yin D, Bond S, Zhang H (2013) Angry or anxious? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly* 38(2):539-60.

Yoon S, Pavlovic V (2014) Sentiment flow for video interestingness prediction. *Proc.1st ACM Internat. Workshop Human Centered Event Understanding from Multimedia* (ACM, New York), 29-34.

Zhu F, Zhang XM (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *J. Marketing* 74(2):133-48.

Zhu W, Ash J, Li Z, Wang Y, Lowry M (2015) Applying semi-supervised learning method for cellphone-based travel model classification. *The First IEEE Internat. Smart Cities Conf.* (Guadalajara, Jalisco), 1-6.

Zhu X (2007) *Semi-Supervised Learning Tutorial* (Department of Computer Sciences University of Wisconsin, Madison, WI).

# Appendix 1.

## BRAND LIST

Table A1. Brand List

| Brand | Brand Image[1] | Number of Posts | Average Number of Likes/Post | Average Number of Comments/Post |
|---|---|---|---|---|
| A and W Root Beer | fun | 966 | 33.03 | 1.67 |
| Bud Light | fun | 1,574 | 53.96 | 1.93 |
| Cheetos | fun | 1,076 | 99.29 | 2.92 |
| Dunkin Donuts | fun | 1,346 | 138.06 | 2.63 |
| Fanta | fun | 593 | 555.04 | 6.63 |
| Fisher Price | fun | 796 | 72.16 | 2.41 |
| Harry Potter | fun | 1,113 | 388.90 | 5.04 |
| Hershey | fun | 1,579 | 70.89 | 2.33 |
| iPod | fun | 1,040 | 192.40 | 2.49 |
| Jello | fun | 1,123 | 72.26 | 2.19 |
| Kool Aid | fun | 1,312 | 120.05 | 3.04 |
| Lays Chips | fun | 1,267 | 57.00 | 2.54 |
| Leap Frog | fun | 987 | 62.02 | 2.12 |
| Lego | fun | 1,543 | 137.46 | 3.20 |
| Little Tikes | fun | 940 | 141.74 | 2.58 |
| M&M | fun | 1,418 | 75.85 | 2.81 |
| Mattel | fun | 1,118 | 148.84 | 3.02 |
| Mug Root Beer | fun | 828 | 22.44 | 1.28 |
| Nestle | fun | 658 | 50.81 | 2.48 |
| Nintendo | fun | 1,192 | 303.83 | 5.12 |
| Old Navy | fun | 1,013 | 114.04 | 4.55 |
| Oreos | fun | 1,692 | 152.83 | 4.81 |
| Pepsi | fun | 777 | 712.77 | 6.07 |
| Playskool | fun | 541 | 23.58 | 1.34 |
| PSP | fun | 1,121 | 76.70 | 2.80 |
| Snickers | fun | 914 | 130.18 | 3.27 |
| Victoria Secret | fun | 945 | 272.35 | 4.98 |
| Wii | fun | 1,093 | 184.86 | 4.30 |
| Xbox 360 | fun | 1,129 | 332.09 | 6.77 |
| Armani | glamorous | 515 | 226.73 | 3.41 |
| BMW | glamorous | 685 | 613.72 | 5.61 |
| Burberry | glamorous | 592 | 126.21 | 3.10 |
| Chanel | glamorous | 615 | 433.15 | 5.26 |
| Clairol | glamorous | 1,289 | 61.95 | 3.09 |
| Clinique | glamorous | 808 | 180.80 | 3.91 |
| Cover Girl | glamorous | 1,605 | 254.53 | 5.50 |
| Dolce & Gabbana | glamorous | 1,157 | 141.08 | 3.46 |
| Estee Lauder | glamorous | 1,046 | 255.61 | 4.26 |
| Ferrari | glamorous | 956 | 572.77 | 6.61 |
| Garnier Fructis | glamorous | 535 | 59.98 | 3.38 |
| Gucci | glamorous | 688 | 533.06 | 5.03 |

| | | | | |
|---|---|---|---|---|
| Herbal Essence | glamorous | 1,092 | 292.41 | 5.56 |
| Jaguar | glamorous | 937 | 428.79 | 4.55 |
| Lancome | glamorous | 936 | 249.71 | 4.95 |
| Lexus | glamorous | 729 | 414.49 | 4.89 |
| L'Oréal | glamorous | 1,083 | 195.77 | 4.26 |
| Louis Vuitton | glamorous | 926 | 391.62 | 4.90 |
| Mary Kay | glamorous | 289 | 42.81 | 1.44 |
| Maybelline | glamorous | 1,357 | 317.43 | 5.07 |
| Mercedes Benz | glamorous | 715 | 621.86 | 6.04 |
| Pantene | glamorous | 415 | 108.81 | 2.25 |
| Porsche | glamorous | 946 | 696.73 | 6.22 |
| Prada | glamorous | 777 | 331.86 | 3.62 |
| Ralph Lauren | glamorous | 632 | 536.14 | 4.86 |
| Revlon | glamorous | 600 | 151.96 | 4.11 |
| Tresemme | glamorous | 494 | 103.04 | 3.15 |
| Adidas | rugged | 1,054 | 682.28 | 6.47 |
| Ajax | rugged | 1,133 | 74.02 | 2.11 |
| Black & Decker | rugged | 788 | 94.17 | 2.27 |
| Chevrolet | rugged | 874 | 657.71 | 6.21 |
| Converse | rugged | 593 | 210.28 | 3.07 |
| Dodge | rugged | 1,293 | 555.24 | 5.54 |
| Eddie Bauer | rugged | 1,204 | 91.84 | 2.95 |
| Ford | rugged | 1,201 | 514.73 | 5.08 |
| Gillette | rugged | 883 | 75.18 | 2.65 |
| GMC | rugged | 1,230 | 304.50 | 2.77 |
| Harley Davidson | rugged | 1,132 | 185.40 | 2.56 |
| Jeep | rugged | 1,430 | 263.09 | 3.57 |
| Kenmore | rugged | 1,510 | 47.14 | 1.92 |
| Land Rover/Range Rover | rugged | 1,096 | 278.24 | 3.41 |
| Levis | rugged | 651 | 483.84 | 6.79 |
| Maytag | rugged | 1,237 | 50.73 | 1.98 |
| Motorola | rugged | 660 | 239.41 | 3.29 |
| New Balance | rugged | 502 | 233.26 | 4.09 |
| Nike | rugged | 780 | 715.83 | 6.62 |
| Old Spice | rugged | 946 | 98.13 | 2.94 |
| Osh Kosh | rugged | 617 | 52.75 | 1.34 |
| Panasonic | rugged | 658 | 131.98 | 2.60 |
| Reebok | rugged | 625 | 349.59 | 4.26 |
| Subaru | rugged | 1,004 | 401.09 | 4.13 |
| Suzuki | rugged | 716 | 658.47 | 5.10 |
| Under Armour | rugged | 823 | 177.79 | 3.80 |
| Volvo | rugged | 732 | 135.17 | 2.17 |
| Whirlpool | rugged | 580 | 66.27 | 3.04 |
| Wilson | rugged | 1,144 | 192.06 | 2.82 |
| Yamaha | rugged | 664 | 532.76 | 4.21 |

[1] From Lovett, Peres, Shachar's (2014) online supplement (dataset).

# Appendix 2.

## INSTRUCTIONS FOR THE STUDENTS (INSTAGRAM DATA)

### Instagram Study

The aim of this task is to detect a set of hashtags (or words) that include multiple brands. For example, below Example 1 includes multiple brands, and Example 2 includes just one brand (i.e., reebok). Your job is to detect sets of hashtags that include multiple brands (i.e., Example 1) by coloring the cell in the excel sheet.

Example 1.

| |
|---|
| yamaha honda suzuki kawasaki bmw ducati motogp dadecountyriderz motorcycle bikelife bikefam miami miamibikelife sportbikelife cbr r1 r6 bikersofinstagram riderich bikekings bikeswithoutlimits shift_life instamotogallery motorcyclesofinstagram bikerguys stuntbikes bikes_vs_cops universalbikers bikelifeshoutout sexy |

Example 2.

| |
|---|
| peaceout bodybuilding olympia healthyfood health healthylifestyle shredded ripped fitness fitnessmodel lifestyleblogger photography gurushots delhi protein whey snacks healthyme me love photooftheday instamood fashion hats dope reebok |

The text comes from hashtags in Instagram. For example, the words from Example 1 are originally hashtags (e.g., #yamaha, #bikekings, #motogp, #riderich, #bikekings). We deleted the pound sign (#) in front of all the hashtags. Please be sure the following rules when you are working on the task.

1) The word of 'Instagram' is not considered as a brand.
   Example)) reebok instagram sports photography niceday happylife nestle goinghome
   In the example above, only 'reebok' and 'nestle' are considered as a brand.
2) Count as one brand if a set of hashtags includes several product-level brands under a corporate-level brand.
   Example)) sports photography niceday happylife goinghome iphone7 iphone8 ipad macpro
   In the example above, we consider the set of words include only one brand since all the brands are under one corporate brand (i.e., Apple).
3) Count as one brand if a set of hashtags includes both a corporate brand and product brands under the corporate brand (e.g., kitkat, nestle)
   Example)) Halloween teacher costume kids kitkat nestle movieday
   In the example above, we consider the set of hashtags include one brand since nestle is a corporate brand of kitkat.
4) Do not color if the row includes only NA or a set of hashtags does not include any brand.
   Example)) NA
   The example above contains only NA, so do not color it.
   Example)) babyboy trike autumnstroll sunnyafternoon
   The example above does not contain any brand, so do not color it.
5) Do not consider a celebrity's name a brand (e.g., justinbieber)

6) Instagram users frequently use several different words in one hashtag (e.g., #lovedunkin) and use several different hashtags that indicate one brand (e.g., #nikepro, #nike, #nikeworld, #newnike, #airjordannike). In both cases, please consider them a brand.
   Example)) lovedunkin dunkindonutday hello morningcoffee freshair busy feelsogood
   The example above includes only one brand (i.e., dunkin donuts).
   Example)) nikepro nike nikeworld newnike airjordannike hair sport healthylife newversion
   The example above includes only one brand (i.e., nike)

You can look up online if you are uncertain about brands (e.g., if the words are brands or not), but the important thing is you should 'not' spend too much time. 30-minute is recommended for 500 sets of hashtags (i.e., 500 rows in the excel sheet) at once you are familiar with the task.

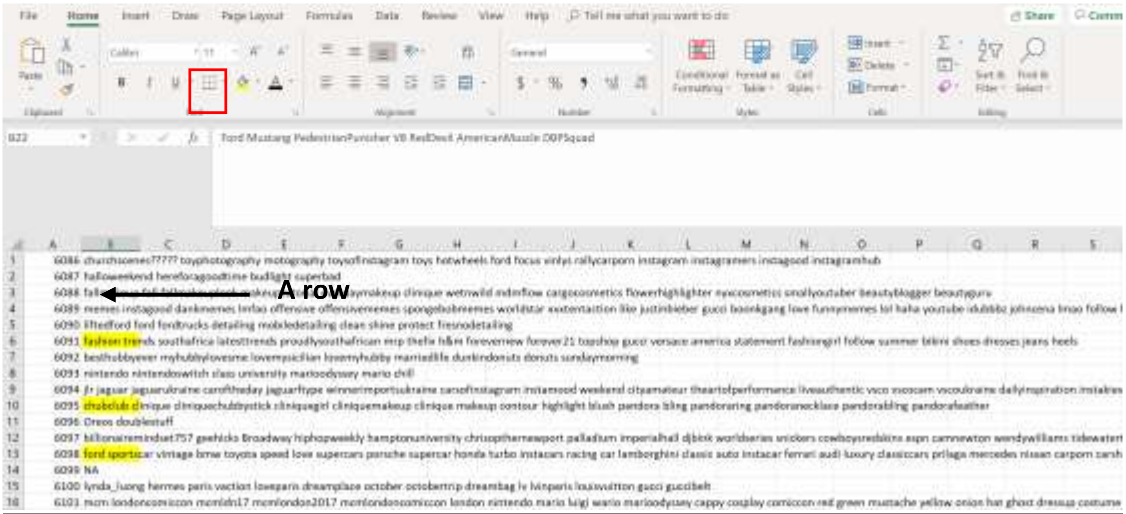
Please be familiar with the brand list below that frequently appear in the hashtags you are investigating before starting the task.

Figure A1. Brand List

lego / harleydavidson / prada / playschool / jeep / dolceandgabbana / fisherprice / blackanddecker / louisvuitton / xbox360 / landrover / victoriassecret / mattel / levis / esteelauder / leapfrog / wilson / gucci / nintendo / eddiebauer / armani / littletikes / nike / covergirl / wii / yamaha / revlon / psp / ford / porsche / mandms / maytag / jaguar / harrypotter / suzuki / mercedesbenz / oreos / reebok / chanel / snickers / volvo / bmw / hershey / oldspice / ferrari / cheetos / gmc / tresemme / koolaid / dodge / lancome / jello / whirlpool / loreal / ipod / newbalance / marykay / nestle / gillette / lexus / mugrootbeer / Chevrolet / herbalessences / victoriasecret / ajax / clairol / aandwrootbeer / kenmore / clinique / budlight / underarmour / maybelline / dunkindonuts / adidas / pantene / pepsi /subaru / ralphlauren / oldnavy / oshkosh / garnierfructis / fanta / converse / topshop / 7cup / motorola / burberry / layschips / panasonic / tiffanyandco / yamaha / honda / lamborghini / mac / unban decay / puma / play station / sony / samsung / nissan / kitkat / givenchy / balenciaga / marcjacobs / rolex / disney / range rover / nikon / canon

Your job is to color the row having multiple brands by coloring the row using <u>'Fill Color'</u> (the icon in the red box below) in the excel sheet. Figure 1 is the example of the task. The rows that are colored in yellow in the Figure 1 include multiple brands. Finally, please be sure that save the file as .xlxs instead of other extensions (e.g., .csv / .txt / .pdf) as the original files are saved. Also, please do not change any other things (e.g., changing the file names / deleting columns or rows) except for coloring the rows.

Figure A2. Snapshot of Data



# Appendix 3.

## OBJECTS DETECTED

Table A2. Sample of Object Types Detected

| Living Things | Food | Plant | Scenery and Events | Non-Living Things |
|---|---|---|---|---|
| adult | alcohol | bushes | assortment | abacus |
| animal | apple | cactus | beach | accessory |
| antelope | apricot | coral | birthday | air |
| baby | banana | daisy | business | aircraft |
| bear | batter | flower | canyon | airplane |
| bovine | beer | fungus | catcher | airport |
| boy | berry | grass | christmas | antique |
| cattle | beverage | green | city | appliance |
| child | bread | greens | clean | apron |
| cow | breakfast | hash | close | area |
| dancer | broccoli | hay | clouds | armor |
| deer | bun | lush | desert | artifact |
| eagle | burrito | nature | dining | backpack |
| elephant | cabbage | palm | drinking | bag |
| female | cauliflower | pasture | driving | ball |

# Appendix 4.

## ROBUSTNESS CHECKS

Table A3. Robustness Checks (*LIKES*)

| | *LIKES* Brand Equity As 2 Variables | *LIKES* Brand Equity As 1 Variable | *LIKES* Nonlinear Monotone | *LIKES* With Middle 98% of the Data |
|---|---|---|---|---|
| Intercept | 2.01***(.07) | 1.88***(.07) | 2.26*** (.10) | 1.98*** (.09) |
| *Image Content* | | | | |
| VizSenti | .04***(.01) | .04***(.01) | .03***(.01) | .05***(.01) |
| VizComplexity | .10***(.01) | .09***(.01) | .10***(.01) | .08***(.01) |
| VizComplexity^2 | .04***(.00) | .04***(.01) | .03***(.00) | .001(.00) |
| VizComplexity^3 | -.02***(.00) | -.02***(.00) | -.02***(.00) | -.007***(.00) |
| *Text Content* | | | | |
| TextSenti | .09***(.00) | .09***(.00) | .08***(.00) | .06***(.00) |
| TextComplexity | .25***(.01) | .25***(.01) | .21***(.01) | .22***(.01) |
| TextComplexity^2 | .24***(.00) | .24***(.00) | .24***(.00) | .17***(.00) |
| TextComplexity^3 | -.10***(.00) | -.10***(.00) | -.10***(.00) | -.08***(.00) |
| *Brand Characteristics* | | | | |
| Visibility | .06***(.01) | .06***(.01) | .10***(.01) | .08***(.01) |
| Involvement | .44***(.02) | .42***(.02) | .30***(.02) | .33***(.02) |
| PerceivedRisk | .01(.03) | .14***.(.02) | -.13***(.03) | .003(.02) |
| *Object Types* | | | | |
| Living | .06***(.01) | .06***(.01) | .02.(.01) | .09***(.01) |
| Food | -.20***(.01) | -.20***(.01) | -.18***(.01) | -.15***(.01) |
| Plant | .10***(.01) | .09***(.01) | .12***(.01) | .10***(.01) |
| NumFaces | .05***(.00) | .05***(.00) | - | .05***(.00) |
| ln(NumFaces) | - | - | .04***(.00) | - |
| TextLength | .002***(.00) | .002***(.00) | - | .002***(.00) |
| ln(TextLength) | - | - | .15***(.00) | - |
| *Brand Equity* | | | | |
| Relevance | | | -.19***(.01) | -.14***(.01) |
| Differentiation | | | .67***(.03) | .48***(.03) |
| Esteem | | | -.38***(.03) | -.26***(.03) |
| Knowledge | | | .15***(.01) | .15***(.01) |
| Brand Stature | -.09***(.00) | | | |
| Brand Strength | .15***(.01) | | | |
| Brand Asset | | -.009***(.00) | | |
| *User Characteristics* | | | | |
| NumFollowers[1] | 1.62***(.00) | 1.62***(.00) | .1.59***(.02) | 2.87***(.01) |
| NumFollowing | .00008***(.00) | .00008***(.00) | .00008***(.00) | .00008***(.00) |
| PostCount | -.00004***(.00) | -.00004***(.00) | -.00004***(.00) | -.00005***(.00) |
| *Product Characteristics* | | | | |
| ExpGood | -.12***(.01) | -.13***(.01) | -.16***(.01) | -.10***(.01) |
| Premium | -.16***(.00) | -.18***(.01) | -.15***(.01) | -.11***(.01) |
| AIC | 786,642 | 786,971 | 785,469 | 744,410 |
| Deviance | 85,350 | 85,391 | 85,201 | 80,738 |
| alpha | 1.35 | 1.35 | 1.33 | 1.03 |
| N | 72,194 | 72,194 | 72,194 | 68,287 |

.*p*<.05; **p*<.01; ***p*<.001; ****p*<.0001

Table A4. Robustness Checks (*COMMENTS*)

| | *COMMENTS* Brand Equity As 2 Variables | *COMMENTS* Brand Equity As 1 Variable | *COMMENTS* Nonlinear Monotone | *COMMENTS* With Middle 98% of the Data |
|---|---|---|---|---|
| Intercept | .23*(.10) | .14(.10) | -.88*** (.14) | -0.03 (.14) |
| *Image Content* | | | | |
| VizSenti | .11***(.01) | .11***(.01) | .09***(.01) | .10***(.01) |
| VizComplexity | .03*(.01) | .02.(.01) | .02* (.01) | .02(.01) |
| VizComplexity^2 | .006***(.00) | .008(.00) | .007(.00) | -.01*(.00) |
| VizComplexity^3 | -.007***(.00) | -.008*(.00) | -.007*(.00) | -.004(.00) |
| *Text Content* | | | | |
| TextSenti | .12***(.01) | .12***(.01) | .12***(.01) | .10***(.01) |
| TextComplexity | .015(.01) | .02.(.01) | -.04***(.01) | .004(.01) |
| TextComplexity^2 | .21***(.01) | .21***(.01) | .23***(.01) | .18***(.00) |
| TextComplexity^3 | -.07***(.00) | -.07***(.00) | -.07***(.00) | -.06***(.00) |
| *Brand Characteristics* | | | | |
| Visibility | .04*(.02) | .04*(.02) | .06***(.02) | .04*(.02) |
| Involvement | .13***(.02) | .12***(.02) | .14***(.03) | .10***(.03) |
| PerceivedRisk | -.18***(.04) | -.07*(.04) | -.16***(.04) | -.14***(.04) |
| *Object Types* | | | | |
| Living | .11***(.02) | .11***(.02) | .05**(.02) | .13***(.02) |
| Food | -.07***(.01) | -.07***(.01) | -.05***(.01) | -.06***(01) |
| Plant | .04*(.02) | .04*(.02) | .05**(.02) | .02(.02) |
| NumFaces | .04***(.01) | .04***(.01) | - | .04***(.01) |
| ln(NumFaces) | - | - | .05***(.00) | - |
| TextLength | .005***(.00) | .005***(.00) | - | .005***(.00) |
| ln(TextLength) | - | - | .32***(.01) | - |
| *Brand Equity* | | | | |
| Relevance | | | .06**(.02) | .09***(.02) |
| Differentiation | | | .50***(.05) | .40***(.05) |
| Esteem | | | -.51***(.04) | -.39***(.04) |
| Knowledge | | | .06***(.02) | .03.(.02) |
| Brand Stature | -.08***(.01) | | | |
| Brand Strength | .16***(.02) | | | |
| Brand Asset | | -.008***(.00) | | |
| *User Characteristics* | | | | |
| NumFollowers[1] | .65***(.01) | .65***(.01) | .64***(.01) | 1.28***(.03) |
| NumFollowing | .00005***(.00) | .00005***(.00) | .00005***(.00) | .00005***(.00) |
| PostCount | -00006***(.00) | -.00007***(.00) | -.00006***(.00) | -.00009***(.00) |
| *Product Characteristics* | | | | |
| ExpGood | -.11***(.02) | -.13***(.02) | -.14***(.02) | -.11***(.02) |
| Premium | -.04*(.02) | -.05**(.02) | -.01(.02) | -.006(.02) |
| AIC | 306,250 | 306,407 | 305,632 | 298,242 |
| Log-likelihood | -15,310 | -15,320 | -15,280 | -146,000 |
| Alpha | 1.93 | 1.94 | 1.91 | 1.73 |
| N | 72,194 | 72,194 | 72,194 | 72,194 |

.$p$<.05; *$p$<.01; **$p$<.001; ***$p$<.0001

Table A5. Consumer Engagement: *Log*(*LIKES*)

| | *Log(LIKES)* Model 1 | *Log(LIKES)* Model 2 |
|---|---|---|
| Intercept | -.347**(-2.83) | 1.78***(3.80) |
| *Image Content* | | |
| VizSentiment | .059***(7.92) | -1.35***(-4.31) |
| VizSentiment^2 | | .304***(4.28) |
| VizSentiment^3 | | -.021***(-4.04) |
| VizComplexity | .124***(13.08) | .123***(12.95) |
| VizlComplexity^2 | -.009*(-2.24) | -.009*(-2.25) |
| VizComplexity^3 | -.010***(-3.93) | -.010***(-3.87) |
| *Object Types:* | | |
| Living | .108***(8.52) | .109***(8.56) |
| Food | -.191***(-15.69) | -.192***(-15.78) |
| Plant | .126***(8.35) | .124***(8.22) |
| NumFaces | .068***(12.95) | .067***(12.75) |
| *Text Content* | | |
| TextSentiment | .092***(17.03) | .092***(16.99) |
| TextComplexity | .304***(34.44) | .304***(34.44) |
| TextComplexity^2 | .204***(44.79) | .204***(44.74) |
| TextComplexity^3 | -.109***(-40.96) | -.109***(-40.96) |
| TextLength | .002***(13.09) | .002***(13.08) |
| *Brand Characteristics* | | |
| Visibility | .071***(4.33) | .071***(4.37) |
| Involvement | .497***(22.27) | .495***(22.16) |
| PerceivedRisk | .311***(9.99) | .312***(10.00) |
| *Brand Equity:* | | |
| Relevance | -.049**(-3.05) | -.049**(-3.09) |
| Differentiation | .388***(9.63) | .389***(9.66) |
| Esteem | -.461***(-13.49) | -.458***(-13.41) |
| Knowledge | .182***(13.54) | .182***(13.52) |
| *User Characteristics* | | |
| NumFollowers[1] | .214***(77.31) | .214***(77.29) |
| NumFollowing | .00008***(22.04) | .00008***(22.03) |
| PostCount | -.000004*(-1.99) | -.00004***(-2.04) |
| *Product Characteristics* | | |
| ExpGood | .062***(2.37) | .061***(4.52) |
| Premium | -.161***(-10.46) | -.161***(-10.50) |
| R-squared | .192 | .192 |
| Adjusted R-squared | .192 | .192 |
| N | 72,194 | 72,194 |

*p*<.05; **p*<.01; ***p*<.001; ****p*<.0001

*Note.* [1] NumFollowers re-scaled to the range 0-100 / Models were estimated with ordinary least squares.

Table A6. Model Fit Comparisons with Different Functional Forms

| *LIKES* | Model1 | | Model2 | | Model3 | | Model4 | | Model5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Variable | Significance | Variable | Significance | Variable | Significance | Variable | Significance | Variable | Significance |
| Visual Sentiment | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + | Yes | Yes: - |
| Visual Sentiment^2 | No | | Yes | No: - | No | | No | | Yes | Yes: + |
| Visual Sentiment^3 | No | | No | | No | | No | | Yes | Yes: - |
| Visual Complexity | Yes | No: + | Yes | No: + | Yes | No: + | Yes | Yes: + | Yes | Yes: + |
| Visual Complexity^2 | No | | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + |
| Visual Complexity^3 | No | | No | | No | | Yes | Yes: - | Yes | Yes: - |
| Text Sentiment | Yes | Yes: + | Yes | Yes: - | Yes | Yes: + | Yes | Yes: + | Yes | No: + |
| Text Sentiment^2 | No | | Yes | Yes: + | No | | No | | Yes | No: - |
| Text Sentiment^3 | No | | No | | No | | No | | Yes | Yes: + |
| Text Complexity | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + |
| Text Complexity^2 | No | | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + |
| Text Complexity^3 | No | | No | | No | | Yes | Yes: - | Yes | Yes: - |
| Interaction Variables | Yes | | Yes | | Yes | | Yes | | Yes | |
| Control Variables | Yes | | Yes | | Yes | | Yes | | Yes | |
| Brand Characteristics | Yes | | Yes | | Yes | | Yes | | Yes | |
| BAV | Yes | | Yes | | Yes | | Yes | | Yes | |
| AIC | 788,504 | | 786,107 | | 786,297 | | 784,210 | | 784,024 | |
| Deviance | 85,579 | | 85,273 | | 85,297 | | 85,038 | | 85,015 | |
| N | 72,194 | | 72,194 | | 72,194 | | 72,194 | | 72,194 | |

*Notes.* Control Variables include Object Types, Text Length, Number of Faces, User Characteristics (Number of Followers, Number of Followings, and Post Count), and Product characteristics (Experience Good, Premium Good) / Brand Characteristics include Visibility, Involvement, and Perceived Risk, and BAV include Relevance, Differentiation, Esteem and Knowledge.

Table A7. Model Fit Comparisons with Different Functional Forms Continued

| *LIKES* | Model6 | | Model7 | | Model8 | | Model9 | | Model10 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Variable | Significance | Variable | Significance | Variable | Significance | Variable | Significance | Variable | Significance |
| Visual Sentiment | Yes | Yes: - | Yes | No: + | Yes | Yes: - | Yes | No: + | Yes | Yes: - |
| Visual Sentiment^2 | Yes | Yes: + | Yes | No: - | Yes | Yes: + | Yes | No: - | Yes | Yes: + |
| Visual Sentiment^3 | Yes | Yes: - | No | | Yes | Yes: - | No | | Yes | Yes: - |
| Visual Complexity | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + |
| Visual Complexity^2 | Yes | Yes: + | Yes | Yes: + | No | | No | | No | |
| Visual Complexity^3 | Yes | Yes: - | Yes | Yes: - | No | | No | | No | |
| Text Sentiment | Yes | Yes: + | Yes | Yes: + | Yes | Yes: - | Yes | Yes: - | Yes | Yes: - |
| Text Sentiment^2 | No | | No | | Yes | Yes: + | Yes | Yes: + | Yes | Yes: + |
| Text Sentiment^3 | No | | No | | Yes | Yes: - | No | | No | |
| Text Complexity | Yes | Yes: + | Yes | Yes: + | Yes | Yes: - | Yes | Yes: - | Yes | Yes: - |
| Text Complexity^2 | Yes | Yes: + | Yes | Yes: + | No | | No | | No | |
| Text Complexity^3 | Yes | Yes: - | Yes | Yes: - | No | | No | | No | |
| Interaction Variables | Yes | | Yes | | Yes | | Yes | | Yes | |
| Control Variables | Yes | | Yes | | Yes | | Yes | | Yes | |
| Brand Characteristics | Yes | | Yes | | Yes | | Yes | | Yes | |
| BAV | Yes | | Yes | | Yes | | Yes | | Yes | |
| AIC | 784,184 | | 784,211 | | 787,749 | | 787,821 | | 787,788 | |
| Deviance | 85,035 | | 85,038 | | 85,482 | | 85,492 | | 85,487 | |
| N | 72,194 | | 72,194 | | 72,194 | | 72,194 | | 72,194 | |

*Notes.* Control Variables include Object Types, Text Length, Number of Faces, User Characteristics (Number of Followers, Number of Followings, and Post Count), and Product characteristics (Experience Good, Premium Good) / Brand Characteristics include Visibility, Involvement, and Perceived Risk, and BAV include Relevance, Differentiation, Esteem and Knowledge.

# Appendix 5

## SPAM REVIEW DETECTION ALGORITHMS

To inform creating a training sample for use the human evaluators, we first employ a fully automated labeling procedure based on the five behavioral features of spam from Lim et al. (2010). The first spam detection method, Targeting Product (TP), is based on the assumption that spamming behavior against a targeted product is characterized by the number of reviews on the products as well as the rating and text similarities of the reviews. Users (reviewers) who create multiple reviews on the same product are considered potential spammers. In our dataset, 3,981 reviewer-product pairs (1.47% of the total reviewer-product pairs) involve multiple reviews and ratings; 2,582 pairs belong to a high cluster group (i.e., 5-star ratings), and 373 pairs belong to a low cluster group (i.e., 1- or 2-star ratings). Figure WB1 presents the distribution of the number of review-product pairs on the number of ratings on the same product for the high rating cluster and the low rating cluster. Within the subset, we assign two types of similarity scores (i.e., "rating spam score" and "text spam score") to the reviewer based on the rating and text similarities of the reviews on the same products. The rating spam function assigns high spam scores to reviewers with a large proportion of reviews with multiple similar ratings on products. The text spam function uses similar logic to the rating spam function; it assigns high spam scores to users who write a large proportion of reviews with multiple reviews with similar text. The text similarity score is computed as a function of cosine similarity of the bi-gram term frequency and inverse document frequency (TFIDF) vectors of two documents. The final TP score is calculated by a linear combination of the "rating spam score" and the "text spam score" with equal weight (.5) on each term.

The second method, Targeting Group (TG), characterizes spamming behavior that promotes or denigrates a set of products sharing common attributes within a short span of time. If a spammer writes reviews on a set of products with the same brand in a short period of time, a high spam score will be assigned to this reviewer. We employ brand as the common attribute and set a day as the short span of time. Assuming that spammers have a promotional or defamatory intention, this method flags high (5-star) or low (1- or 2-star) ratings on products sharing the same brand by the same user within a day. Only a sufficiently large number of ratings from a user on a brand in the specified time window can capture the 'burstiness' of rating behavior, and we set the minimum size thresholds of the high and low rating clusters to be three and two, respectively. 917 unique reviewers wrote reviews on products with the same brand in a single day for the high rating cluster, and 447 unique reviewers wrote reviews on products with the same brand in a day for the low rating cluster. Figure WB2 displays the distributions of the number of reviewers at each review frequency for the high and the low rating clusters.

The third and fourth methods, General Deviation (GD) and Early Deviation (ED), are based on the deviation of ratings. A typical customer is likely to give ratings that are similar to those of other customers. Since a spammer usually intends to promote or denigrate products, she/he does not have as much incentive to give moderate ratings, and her/his rating may then be rather different from other reviewers' ratings. The GD method captures the deviation of each rating from the average rating of the rated product. Spammers also create reviews early to maximize their impact. The ED model captures this aspect of spamming behavior by giving a weight to the GD score based on the order of reviews on the rated product.

The last method is combined score (CS), which is constructed by multiplying each of the four terms (TP, TG, GD and ED) by a constant (.65, .25, .05, and .05 for TP, TG, GD, and ED, respectively) and adding them. We use analyst judgement to adapt the weight scheme to our context, giving more emphasis to the TP spam score than to the others. GD and ED are given the least weighting, since it is known that deviations are generally weaker evidences of spamming behavior (Chengzhang and Kang 2015).

The technical details of the five features are as follows.

*Targeting Product (TP)*

[i] Rating spam score function $(c_{p,e}(u_i))$for user $u_i$:

(1)
$$c_{p,e}(u_i) = \frac{s_i}{\mathrm{Max}_{u'_i \in U} s_{i'}},$$

(2)
$$s_i = \sum_{e_{ij} \in E_{ij}, |E_{ij}| > 1} |E_{ij}| \cdot \mathrm{sim}(E_{ij}).$$

where $s_i$ is the unnormalized rating spammer score of user $u_i$, and $E_{ij}$ is the set of ratings from user $u_i$ to object $o_j$.

The similarity score in Equation 2 (i.e., sim($E_{ij}$)) compares ratings in a given set and is defined as follows:

(3)
$$\mathrm{sim}(E_{ij}) = 1 - \underset{e_{k,}, e_{k'} \in E_{ij}, k < k'}{\mathrm{Avg}} |e_k - e_{k'}|.$$

where $e_k$'s is the normalized rating score ($e_k \in [0, 1]$) in a given set.

The final rating spam score for user $u_i$ is a normalized spammer score which is earned by taking the ratio of the unnormalized spammer score ($s_i$) for user $u_i$ to the unnormalized spammer score for user $u'_i$, who has the maximum spammer score in the given set (Equation 1). The rating spam function assigns high spam scores to the reviewers with a large proportion of ratings involved with multiple similar ratings on products.

[ii] Text spam score $(c_{p,v}(u_i))$for user $u_i$:

When a spammer produces multiple reviews on a product, not only do the review ratings but the review texts are likely to be similar, because the spammer tries to conserve their effort. The text spam score captures the text similarity among the reviews on a product written by a reviewer and the number of reviews on the same product. Accordingly, the text spam score is defined as:

(4)
$$c_{p,v}(u_i) = \frac{s_i}{\mathrm{Max}_{u'_i \in U} s_{i'}},$$

(5) $$s'_i = \sum_{v_{ij} \in V_{ij}, |V_{ij}|>1} |V_{ij}| \cdot \mathrm{sim}(V_{ij}).$$

where $s'_i$ is the unnormalized text spammer score of user $u_i$, and $V_{ij}$ is the set of review texts from user $u_i$ to object $o_j$.

The text similarity score (i.e., $\mathrm{sim}(V_{ij})$) implies an average similarity score of all combinations of two review texts in a given set ($\mathrm{i.e.}, |V_{ij}| > 1$). The similarity score is defined as:

(6) $$\mathrm{sim}(V_{ij}) = \underset{v_{k,}, v_{k'} \in V_{ij}, k<k'}{\mathrm{Avg}} \mathrm{sim}\ (v_k,\ v_{k'}).$$

Here, the $\mathrm{sim}\ (v_k,\ v_{k'})$ indicates the similarity between two reviews ($v_k$ and $v_{k'}$) in a given set and is defined by a cosine similarity of the bigram TFIDF vectors of $v_k$ and $v_{k'}$ as follows:

(7) $$\mathrm{sim}\ (v_k,\ v_{k'}) = \mathrm{cosine}\ (v_k,\ v_{k'}).$$

The final text spam score for user $u_i$ is also normalized by taking the ratio of the unnormalized spammer score ($s'_i$) for user $u_i$ to the unnormalized spammer score for user $u'_i$, who has the maximum text spamming score in the given set (Equation 4).

The final TP score is a linear combination of the rating spam score and the text spamming score with an equal weight on each term.

(8) $$c_p(u_i) = \frac{1}{2}\left(c_{p,e}(u_i) + c_{p,v}(u_i)\right).$$

*Targeting Group (TG)*

The high (low) rating cluster by user $u_i$ to a product group $b_k$ in time window w is defined as follows:

(9) $$E_{ik}^{\mathcal{H}}(w) = \{e_{ij} \in E_{i*} \mid o_j \in b_k \wedge t(e_{ij}) \in w \wedge e_{ij} \in H\ \mathrm{RatingSet}\},$$

(10) $$E_{ik}^{\mathcal{L}}(w) = \{e_{ij} \in E_{i*} \mid o_j \in b_k \wedge t(e_{ij}) \in w \wedge e_{ij} \in L\ \mathrm{RatingSet}\}.$$

Where, $E_{ik}^{\mathcal{H}}(w)$ ($E_{ik}^{\mathcal{L}}(w)$) is the set of reviews created by user $u_i$ on a product group $b_k$ sharing the same product attribute (i.e., brand) within the time window w for a high rating cluster (a low rating cluster).

The high (low) rating clusters that satisfy the minimum threshold requirement can be defined as:

(11) $$C_i^{\mathcal{H}}(w) = \cup_{k,w}\{E_{ik}^{\mathcal{H}}(w) \mid E_{ik}^{\mathcal{H}}(w) \geq \text{minsize}^{\mathcal{H}}\},$$

(12) $$C_i^{\mathcal{L}}(w) = \cup_{k,w}\{E_{ik}^{\mathcal{L}}(w) \mid E_{ik}^{\mathcal{L}}(w) \geq \text{minsize}^{\mathcal{L}}\}.$$

The TG spam score of the high (low) rating cluster for user $u_i$ is the ratio of the unnormalized TG score of $u_i$ (i.e., $C_i^{\mathcal{H}}(w)$ and $C_i^{\mathcal{L}}(w)$) to the unnormalized TG score of $u'_i$, who has the maximum TG core in each high or low rating cluster. Formally, the TG spam score of the high (low) rating cluster can be defined as follows:

(13) $$c_{g,\mathcal{H}}(u_i) = \frac{C_i^{\mathcal{H}}}{\text{Max}_{u'_i} \in C_{i'}^{\mathcal{H}}},$$

(14) $$c_{g,\mathcal{L}}(u_i) = \frac{C_i^{\mathcal{L}}}{\text{Max}_{u'_i} \in C_{i'}^{\mathcal{L}}}.$$

We combine the two TG spam scores by taking the average as we do for the TP spam score.

(15) $$c_g(u_i) = \frac{1}{2}\left(c_{g,\mathcal{H}}(u_i) + c_{g,\mathcal{L}}(u_i)\right).$$

*General Deviation (GD)*

The General Deviation spam score for a user $u_i$ is defined as:

(16) $$c_d(u_i) = \underset{e_{ij} \in E_{i*}}{\text{Avg}} |d_{ij}|.$$

where $d_{ij}$ ($d_{ij} = e_{ij} - \underset{e \in E_{*j}}{\text{Avg}}\, e$) is the difference between a rating $e_{ij}$ and the average rating on the same product.

*Early Deviation (ED)*

The weight of each rating $e_{ij}$ (i.e., $w_{ij}$) for the ED score is defined as:

(17) $$w_{ij} = \frac{1}{r_{ij}{}^{\alpha}}.$$

where $r_{ij}$ indicates the review order of $e_{ij}$ among all the reviews on the rated product.

Incorporating $w_{ij}$ into the model, the final ED spam score of user $u_i$ is defined as follows:

(18) $$c_e(u_i) = \frac{\sum_{e_{ij} \in E_{i*}} (|d_{ij}| \times w_{ij})}{\sum_{e_{ij} \in E_{i*}} w_{ij}}.$$

*Combined Score (SC)*

The Combined Score (SC) is based on all four methods (TP, TG, GD, and ED) and defined as follows:

(19) $$c(u_i) = .65c_p(u_i) + .25c_g(u_i) + .05c_d(u_i) + .05c_e(u_i).$$

The weighting scheme is empirically determined to give more emphasis to product-specific spamming than group-specific spamming. Deviations are generally weaker evidence of spam and thus are given the smallest weighting.

# Appendix 6

## INSTRUCTIONS GIVEN TO HUMAN EVALUATORS

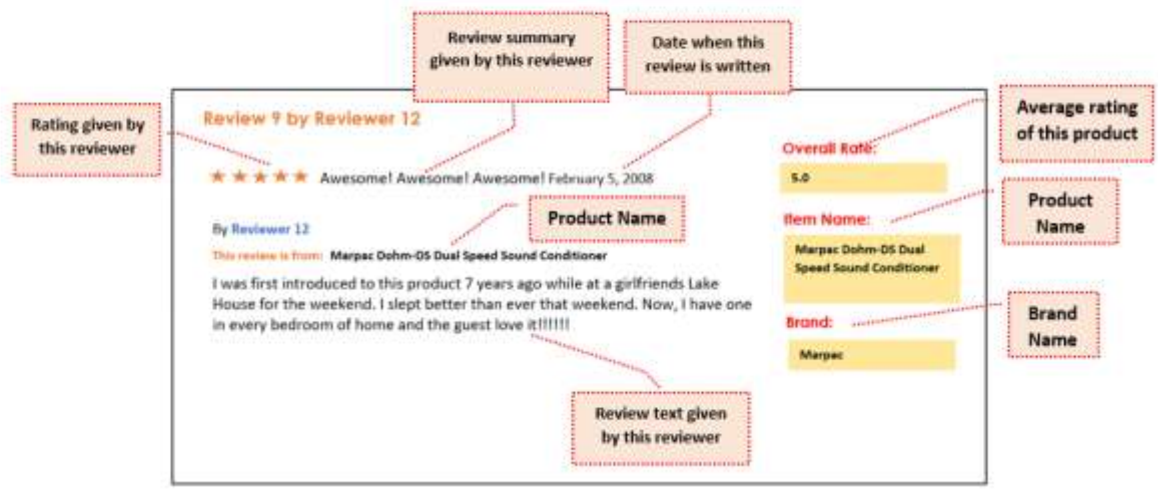
For full instruction, please follow the link https://docs.google.com/forms/d/e/1FAIpQLSelBy63KUNT3M2g443heGOrif8lsw62esgR5NFKocKzNZK_uQ/viewform

## Evaluating Suspicious Reviewers

* Required

The Marketing Research lab at Emory University is currently conducting research related to **fake customer reviews** on e-commerce websites. As part of this research, we are asking human investigators to evaluate how suspicious each reviewer is **based on their reviews** on a certain e-commerce website. Please read each of the reviews carefully and indicate **to what extent you believe each of the reviewers is a spammer (or a non-spammer)**. Again, please **read each of the reviews very carefully** before making a decision. Each review includes the following components: product name, brand name, date, etc. Following example visually explains the review components.

[Review Components]

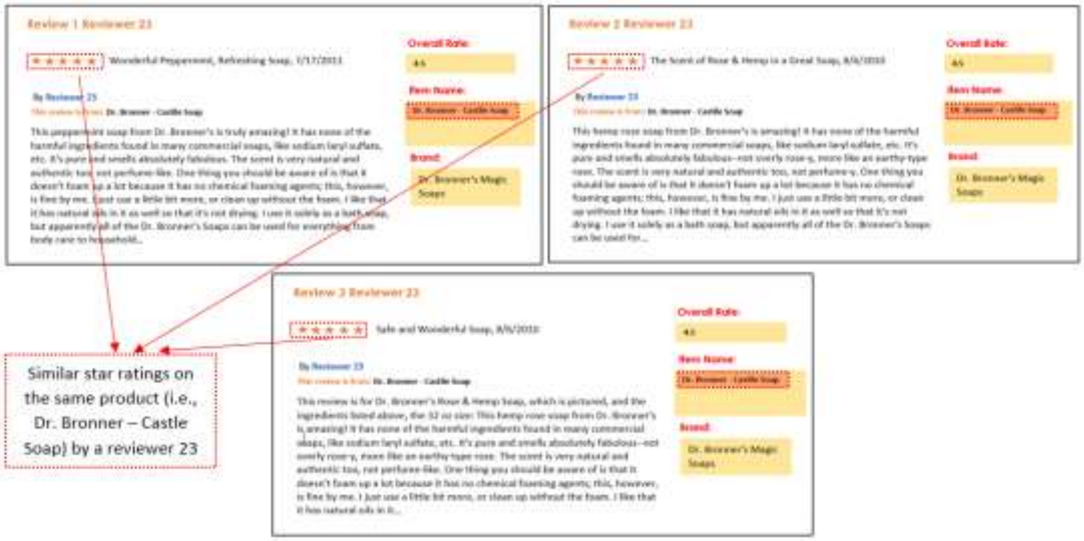



[Characteristics of Suspicious Reviews]

Based on previous academic research, reviews created by highly suspicious reviewers display the following characteristics:

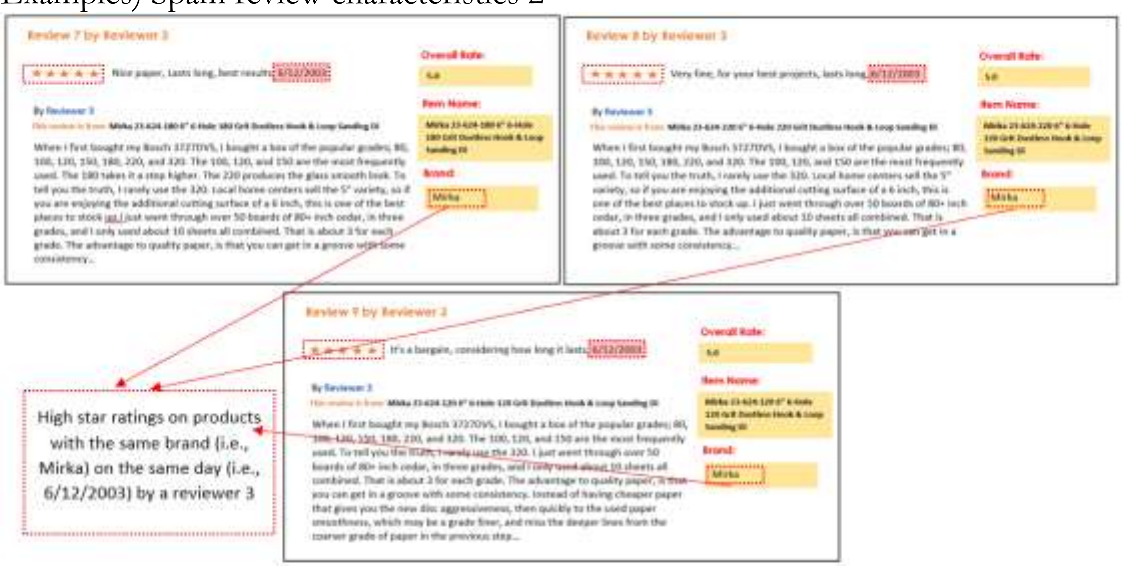
Characteristic 1: Reviews that are (i) exact (or near-exact) duplicates of other reviews; or (ii) have identical (or similar) ratings to other reviews on the same product by the evaluated reviewer. For example, if "reviewer A" left **multiple reviews with similar text or similar star ratings on one product** (e.g., Motorola Surfboard SB5101 Cable Modem), then the reviewer is suspicious as a spammer. Visual examples are as follows:

Examples) Spam review characteristics 1



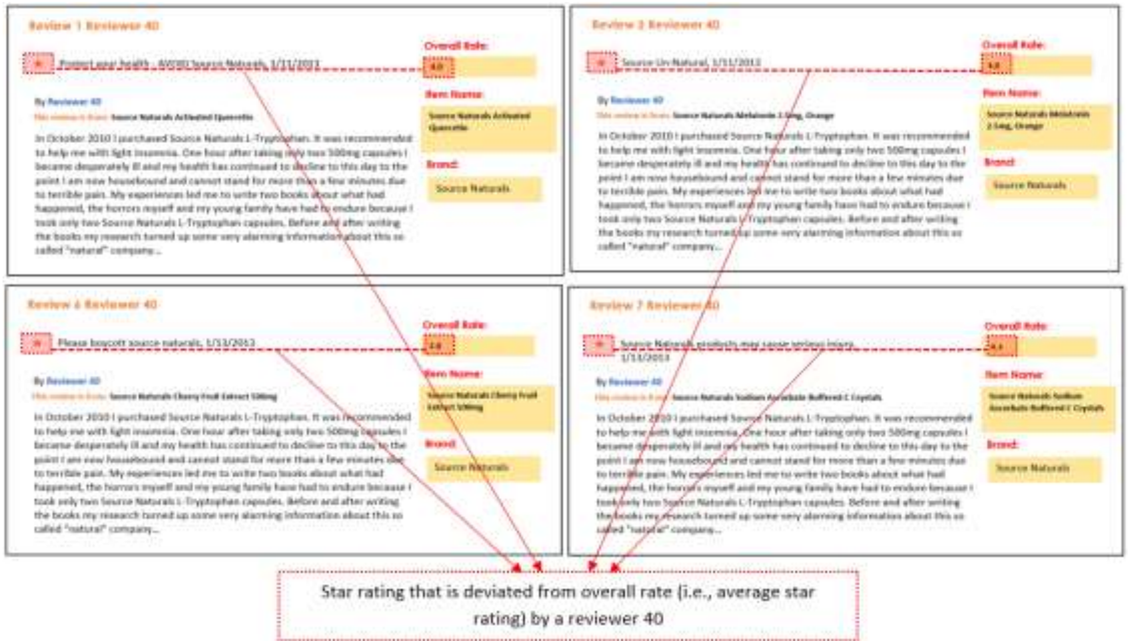

Characteristic 2: Reviews of products of the same brand (e.g., Gucci, Energizer) with multiple identical high or low ratings from the evaluated reviewer within the same day. For example, if "reviewer A" left **<u>several 1-star (or 5-star) ratings on multiple Gucci products on the same day (e.g., 02/05/2008),</u>** then the reviewer is suspicious as a spammer. Visual examples are as follows:

Examples) Spam review characteristics 2




Characteristic 3: Reviews by the evaluated reviewer with ratings that deviate from the average ratings of the reviewed products. For example, if "reviewer A" left **<u>reviews with 1- or 2-star ratings when the average rating of the reviewed product is 5-star</u>**, then the reviewer is suspicious as a spammer. Visual examples are as follows:

Examples) Spam review characteristics 3

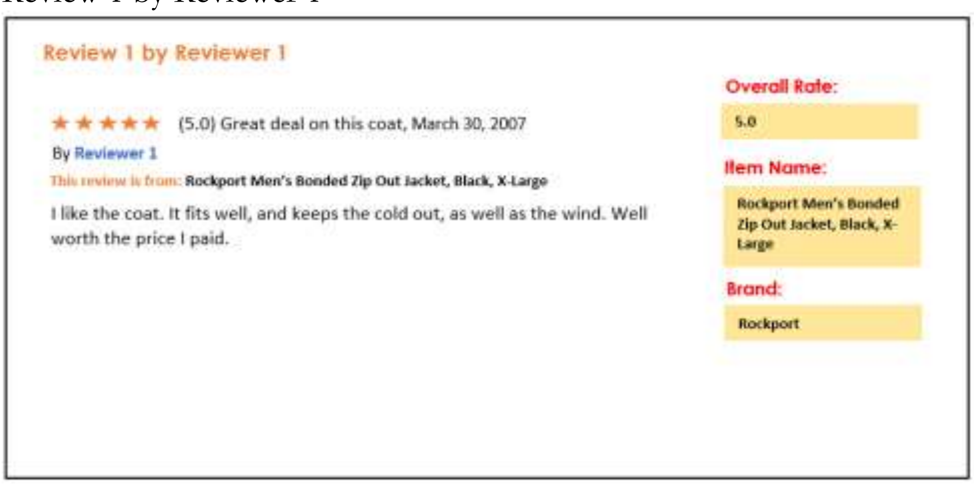



You are evaluating Reviewer 1 to Reviewer 54. Thank you for your cooperation; we promise that your information will be kept strictly confidential.

Reviewer 1. Please select to what extent you believe the following reviewer is a spammer. *

Non-spammer
Slightly suspicious
Somewhat suspicious
Highly suspicious
Spammer

Review 1 by Reviewer 1



Review 2 by Reviewer 1

**Review 2 by Reviewer 1**

★★★★★ (5.0) Brandon Thomas Coat, February 6, 2007

By Reviewer 1

This review is from: Brandon Thomas Women's Ellie Cord Jacket, Black

Could not pass on the price. Ordered a size large, based on reviews, and it fit great on my wife. She likes it, and wears it often at night.

**Overall Rate:**

4.0

**Item Name:**

Brandon Thomas Women's Ellie Cord Jacket, Black

**Brand:**

Brandon Thomas

Reviewer 2. Please select to what extent you believe the following reviewer is a spammer. *

Non-spammer
Slightly suspicious
Somewhat suspicious
Highly suspicious
Spammer

Review 1 by Reviewer 2

**Review 1 by Reviewer 2**

★★★★★ (5.0) High Class, August 20, 2010

By Reviewer 2

This review is from: Royce Toiletry Bag with Zippered Bottom Compartment - L

Sometimes you can't describe you really feel about something you love, and the best way that I can describe how happy I am with this toiletry bag is, WOOOOW!!! I love it.

**Overall Rate:**

5.0

**Item Name:**

Royce Toiletry Bag with Zippered Bottom Compartment - L

**Brand:**

Royce Leather

Review 2 by Reviewer 2

**Review 2 by Reviewer 2**

★★★★★ (5.0) Love this passport case, August 20, 2010

By Reviewer 2

This review is from: **Milano Feather-Lite Manmade Leather Passport Ticket Hol**

I bought this item about 6 months ago, I travel a lot and I'm very happy with the quality of the item, wear and tear does not exist for this item. It is very useful when you see those long lines when you are checking in you don't have to waste time looking for boarding passes or tickets because everything is handy. I love it

**Overall Rate:**

**5.0**

**Item Name:**

**Milano Feather-Lite Manmade Leather Passport Ticket Hol**

**Brand:**

**Royce Leather**