Digital Media Analytics: Towards an Understanding of Content Design

and Social Media Promotion

by

Xueyan Yin

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2020 by the
Graduate Supervisory Committee:

Pei–yu Chen, Chair
Bin Gu
Zhan Shi
Victor Benjamin

ARIZONA STATE UNIVERSITY

August 2020

ABSTRACT

Digital media refers to any form of media which depends on electronic devices for its creation, distribution, view, and storage. Digital media analytics involves qualitative and quantitative analysis from the business to understand users' behaviors. This technique brings disruptive changes to many industries and its path of economic disruption is getting wider and wider. Under the context of the increasingly popular digital media market, this dissertation investigates what are the best content delivery strategy and the new cultural phenomenon: Internet Water Army. The first essay proposes a theory-guided computational approach that consolidates distinct data sources spanning unstructured text, image, and video data, systematically measures modes of persuasion, and unveils the multimedia content design strategies for crowdfunding projects. The second essay studies whether using the Internet Water Army helps sales and under what conditions it helps. This study finds that the Internet water army helps product sales at both post-level and fans-level. The effect is largely reflected by changing the number of emotional fans. Furthermore, the earlier to purchase the water armies, more haters, likers, and neutral fans it can attract. The last essay builds a game model to study the trade- off between honestly promoting the product according to their evaluation and catering to the consumer's prior belief on the product quality to stay on the market as long as possible. It provides insights on the optimum usage of promotion on social media and demonstrate how conventional wisdom about negative reviews will hurt business may be misleading in the presence of social media. These three studies jointly contribute to the

crowdfunding and social media studies literature by elucidating the content delivery strategy, and the impact and purchasing strategy of the Internet Water Army.

ACKNOWLEDGMENTS

One day I was talking about writing acknowledgments for the dissertation with professor Jessica Pye. She told me to finish a Ph.D. degree, we have a village of people to feel thankful to. I found it very true. I am grateful to have supportive family members, exceptional advisors, and cherished friends along the road to my Ph.D.

First of all, none of this would have happened without my amazing advisor, Dr. Pei-yu Chen. She always respects my thoughts and provides continuous support and guidance. Her kindness and her expertise navigate me through my Ph.D. study. She told me to follow my passion and conduct research in areas I really interested in. She also told me to make whatever decisions for whatever reasons as long as I am happy. There were many times during this five-year that I feel it is hard to hold on. None of this would have happened without her. I am more than grateful to have found Dr. Chen during my Ph.D. studies.

I also want to express my sincere thanks to my committee members, Bin Gu, Victor Benjamin, and Zhan (Michael) Shi. Thanks for your encouragement, patience, and guidance in helping me discover an interesting new area, shape the dissertation, and build my career. They are my mentors and also my friends. Thanks for giving me the chance to express my concerns with you and providing valuable insights. Furthermore, I would like to give my thanks to professor Xiao Liu, Jessica Pye, and Donghyuk Shin for our "after teach" coffee time on Monday and Wednesday. Thanks for not only listening to my gripe but also sharing your experience.

Additionally, I would like to express my appreciation to professor Raghu T.S., Shinyi Wu, John Zhang, Benjamin Shao, Sangpil Han, and other faculty members in the

IS department at ASU. Thanks for your insightful discussions on both my research and my life. Thanks for helping me with my teaching. I also want to thank Angelina Saric, Erica Grimwood, Jessica Wu, Janeen Christensen, and Emi Christensen for the administrative help. Moreover, I am very thankful for all my Ph.D. cohorts and fellows in the IS department. I am so grateful that you are part of my Ph.D. journey.

In addition, I want to thank my family members for their unconditional love and support. Specifically, I want to thank my mother, Cathy Chen. Thanks for your love and financial support. Thanks for giving me the freedom to choose the career path I want. Thanks for letting me know that you will always be here for me. Thanks for teaching me for being a strong woman.

Last but not least, I have to thank my husband, Taihua Yang. You made all of this possible for me. Words can't express how lucky I feel to have you with me through this journey. I could write a long thank you to my friends, but I would keep it short here. Thanks to my friend Dora Dai, Bob Liu, and Chris Chen. Thanks for listening to me at any time, making me laugh and supporting me.

To everyone else that I forgot to mention, I thank you all.

TABLE OF CONTENTS

CHAPTER Page

APPENDIX Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

CHAPTER 1

OVERALL INTRODUCTION

Digital media brings disruptive changes to many industries and its path of economic disruption grows wider and wider. Social media is one of the most influential changes digital media brings. The growing use of social media has significantly changed how consumers aware of news or products, communicate with others, exchange money, and so on (Tam & Ho, 2006; Nielsen & Schroder, 2014). In 2019, 70% of the U.S population has a social network profile on at least one platform and more than 50 million small businesses currently use social media platforms to connect with their customers. Social media marketing can help businesses build a stronger online presence, engage the consumers, and offer new opportunities for branding. Therefore, how to effectively use social media promotion strategy and create high-quality content becomes crucial in digital media marketing. In this dissertation, I will investigate how to create a high-quality content design and how to effectively use social media promotion strategy.

The first essay focuses on effective content design. Content creators have generated massive volumes of images and video content because they realize that 65% of the population are visual learners, and 40% of people can process visual information significantly faster than textual content, and retain ingested content substantially longer (Trafton, 2014, Women, 2017). In fact, 66% of people rank video as the most effective form of content for the marketing strategy[1]. Although there is a consensus that visual

[1] https://www.forbes.com/sites/forbescommunicationscouncil/2019/12/19/why-marketers-should-integrate-video-marketing-into-their-content-strategy/#585e829c67a3

content can be highly effective due to its capability to engage the audience, poorly organized visual content can, in fact, backfire, such as leaving negative impressions that outlast those of text-based content (Sundar 2000). A video not thoughtfully put together may also induce a high cognitive load, hampering information retainment (Liu et al., 2018; Anmarkrud et al., 2019; Wang et al., 2020). Consequently, essay 1 of the dissertation provides a theory-driven investigation by employing state-of-the-art AI tools and conducting a large-scale empirical study to advance our understanding of important factors that increase the effectiveness of the content design and delivery.

The second and third essays focus on the optimal promotion strategy. A new profession emerges in the past two years called Internet Water Armies. Internet Water Armies are defined as a group of internet ghostwriters who get paid to post anything online with some particular content. When using the Internet Water Armies, a large number of well-organized people "flood" the Internet with purposeful posts, reviews, comments, and other actions. To gain an online presence and manipulate the attitudes of online users, more and more companies start to hire Internet Water Army to perform social media promotions. According to a survey conducted by Xinhua News, 74 percent of the people interviewed believe the Internet Water Armies are ubiquitous on Chinese social media websites and are widely used by public relations managers. Facebook claimed that five percent of its online users are fake accounts and removed three billion fake accounts over six months (Romm, 2019). Using the Internet Water Armies becomes a popular promotion approach. However, the actual impact of using the Internet Water Armies are yet to be determined. On one hand, massive posts from the Internet Water

Armies can boost the awareness effect and then lead to more organic posts and, therefore, make the promoted product become a hit on social media networks and even on the online media press. On the other hand, the organic posts have declined over the years because too many “fake” posts occupied their social network and consumers have lost their trust in online media channels. In addition, once the audiences realize the buzz was created by the Water Armies, they will lose trust for the company and become a long-term hater. Consequently, essay 2 explores whether using the Internet Water Armies affects sales and under what conditions it helps sales?

Companies use many ways to influence social media promotions. The abundant use of social media promotion raises consumers’ concern about the quality of information delivered through the promotion. Both “likers” and “haters” will appear on social media talking about the same product. Prior research presents mixed results about having negative reviews about the products. Berger et al. (2010) show that negative reviews showing on social media can boost sales. While Basuroy et al. (2003) point out that negative reviews hurt performance more than positive reviews help performance. The actual effect of having doubters remains unanswered. This phenomenon is intensified in social media promotion. Essay 3 aims to study the following research question: How would the movie distributor use social media to promote products by developing a model to study the movie distributor’s optimal promotion strategy for information goods while using social media.

The three essays shed light on digital media analytics by exploring the crucial topics in social media marketing from two aspects: effective content design and optimal promotion strategy. This dissertation integrates various methods including state-of-arts AI technology, econometrics method, and analytical modeling.

CHAPTER 2

TOWARD AN UNDERSTANDING OF THE EFFECTS OF MIXED MEDIA FORMATS: AN AI-ENABLED APPLICATION AND EMPIRICAL STUDY

Abstract

Online content has grown into one of the most significant information sources that consumers utilize to shape their perceptions and decision-making. Despite a great diversity among content creators, they are all naturally concerned with questions surrounding what multimedia content is the best received by consumers, and how to create content that effectively persuades consumers. This study attempts to answer these questions in the context of crowdfunding campaigns with data from an online crowdfunding platform, Kickstarter. Drawing upon Aristotle's Rhetorical Theory of Persuasion and utilizing advanced AI tools, we measure the constructs of persuasion with features extracted from multimedia content and examine their impact on consumer decision-making. This work contributes to both literature and practice by creating a theory-guided computational approach. Our proposed research method consolidates distinct media formats spanning unstructured text, images, and videos, systematically measures modes of persuasion, and unveils the multimedia content design strategies for crowdfunding projects. This study provides content creators with practical guidance on how to design crowdfunding campaigns to communicate project visions and funding needs and gain backers effectively.

Keywords: Content Design, Artificial Intelligence, Crowdfunding, Persuasion, Video Mining, Text Analytics, Emotion Detection

1. Introduction

The Internet has transformed the way people live, work, and communicate profoundly in the past decade. Online platforms such as search engines, social media, e-commerce platforms, app stores, ad networks, and crowdfunding thrive on the ubiquity of information technologies and play an ever more central role in the online world and hence social and economic life. These online platforms attempt to facilitate and extract value from direct interactions or transactions between users based on content created and shared. Online content has entered into a more mature and richer phase and grown into a significant knowledge source that consumers rely on to shape their perceptions and inform their decision-making. 65% of the population are visual learners, and 40% of people can process visual information significantly faster than textual content, and retain ingested content substantially longer (Trafton, 2014, Women, 2017). Recognizing this, content creators have generated massive volumes of images and video content. This development is of great interest to businesses as the commonality of multimedia content can have a significant impact on consumers.

Although there is a consensus that visual content can be highly effective due to its capability to engage the audience (Greenwald & Leavitt, 1984; Liu et al., 2018), poorly organized visual content can, in fact, backfire. For example, ill-designed videos can leave negative impressions that outlast those of text-based content (Sundar 2000). A video not thoughtfully put together may also induce a high cognitive load, hampering information retainment (Liu et al., 2018; Anmarkrud et al., 2019; Wang et al., 2020). Issues with poor content design can further be exacerbated by the large volumes of online content made available to users and competing for attention. Hence, the question of what effective content design and delivery consist of is among the most critical ones faced by all content creators.

To date, the content design has been based on individual experience and judgment rather than scientific evidence. The reason is two-fold. First, content creators are facing many design choices, especially concerning video content, such as the visual style (animated vs. realistic), background music genre (cinematic, acoustic, ambient, etc.), and topic coverage (personnel, product, goals, etc.). Second, the adoption of specific tactics important for content delivery is subject to creators' experience, skills, and knowledge. For instance, skills such as creating animation and improving video and image aesthetics require significant time and monetary investment to develop. Until the impact of different factors on content effectiveness is evaluated systematically and comprehensively, the content design remains a subjective decision.

Although there is a great interest in advancing our understanding of factors that increase content effectiveness, research has been limited to lab experiments or small-scale studies that focus on a limited set of design features (Liu et al., 2018). Artificial intelligence (AI) is defined as a machine's capacity to perform cognitive functions that are associated with human minds - most notably learning and problem-solving (Rai et al. 2019). A distinctive characteristic of AI is its ability to sense, process, and detect patterns from a large volume and variety of data, leading to enhanced computational capabilities for deduction and response (Agrawal et al., 2018). Examples of recent AI applications are numerous and include chatbots, facial and voice recognition, medical image/record diagnosis, personalized recommendations, cyber fraud/attack detection, and more, etc. (Hoque et al., 2013; Xu et al., 2014; Mohan et al., 2017; Guo et al., 2017; Cruciana et al., 2017; Hou et al., 2019). Advances in computational power and artificial intelligence techniques carry the unprecedented potential for companies, platforms, and individuals to decipher the ever-changing world of content design and consumption.

The goal of this research is to provide a theory-driven investigation by employing state-of-the-art AI tools and conducting a large-scale empirical study to advance our understanding of important factors that increase the effectiveness of the content design and delivery. This work instantiates a system for processing and analyzing the online content of various formats through the perspective of different features/aspects. An econometrics modeling approach is utilized to understand the effectiveness of the content design.

To ground this research in a practical, real-world scenario, we focus on scrutinizing crowdfunding campaign success on the Kickstarter web platform. Crowdfunding platforms serve as an intermediary to connect backers with entrepreneurs and organizations that have creative ideas. Campaign creators introduce their ideas on crowdfunding platforms in hopes of luring an audience of individual backers who each contribute some funds to the project (Hu et al., 2015). Creators have the option to create their campaigns through a variety of media formats, including text, images, and videos. Further, Kickstarter presents a closed system where content design and consumption primarily occur within the platform rather than on unobservable mediums, and also where project outcomes are observed. For these reasons, Kickstarter provides an excellent opportunity to scrutinize different multimedia strategies taken by campaign creators and to find the outcomes of their crowdfunding efforts.

A theoretical perspective can guide the analysis of multimedia strategy beyond the arbitrary selection of features while also enhancing the generalizability of findings. The proposed research is conceptualized upon Aristotle's rhetorical theory on persuasion (Aristotle 1991, Burton 2007), which inspires several analyses that are operationalized through numerous AI-technologies. The Aristotelian theory of persuasion suggests that effective persuasion consists of a combination of three strategies: appeal to credibility, appeal to emotion and appeal to logic. Utilizing advanced AI technologies such as video mining, emotion detection, and topic modeling, we can measure the modes of persuasion by extracting the features relevant to persuasion from texts, images, and videos, and study their effects.

Our proposed method extracts a comprehensive list of features from three popular media formats, text, images, and videos on a large scale. It examines their impact on effective content design and delivery, which enables an innovative approach for content design based on theory and evidence. We find that rewards and products are best discussed using a written format, while motivations are best expressed through videos, in terms of improving project outcomes. Our results suggest that topic-related reinforcement and emotion-related reinforcement can significantly improve the project funding outcomes. Aesthetic features of the campaign video have a significant and positive impact on project outcomes and consumer engagement (measured by the number of comments/discussions). Our data analytical proposed approach can be further generalized to infer multimedia strategies in other contexts.

This research makes several unique contributions. First, to our knowledge, this study is the first to integrate a social psychology theory to guide the computational design of multimedia content. The overarching design developed in this study can be generalized to other contexts where social psychological perspectives are adopted to scrutinize multimedia strategy. Second, we can systematically measure the constructs of persuasion and provide a large-scale scientific study to evaluate their effects. Third, this work provides a computational template that can be adopted by future research seeking to tightly couple similar analyses with social or psychological theory. This work contributes to practice by consolidating distinct data sources spanning unstructured text, image, and video data to analyze multimedia strategy, and by providing a template that can be easily applied to other contexts outside crowdfunding campaigns.

2. Research Context

Crowdfunding platforms, such as Kickstart, Patreon, Indiegogo, have created and shaped new markets into more efficient arrangements that bring funding to users based on their creative project ideas presented in text-based campaigns, images, and videos. Headquartered in the U.S., Kickstarter is one of the world's largest crowdfunding platforms. As of September 2019, Kickstarter has raised more than $4 Billion for 459,000 crowdfunded projects. To use Kickstarter, an entrepreneur (called a "creator" on Kickstarter) creates a webpage for their crowdfunding campaign explaining the purpose of their project and the specific deliverables that they aim to produce with contributed funding. The webpage can include content in the format of texts, images, and videos. The creator selects an ending date for the project as well as a funding goal to be met, i.e., the amount of money required to execute the project. On Kickstarter, the funding is all-or-nothing. If the project has not met the funding goal by the end date specified by the project creator, any contributed funds will be returned to users with no funds for the creator.

The success rate among projects that seek funding on Kickstarter is only 37% per Kickstarter's publicly released statistics (KickStarter Statistic 2019[2]). A lift to project funding success rate could yield a significant boon for the platform, creators, and project backers. First of all, a higher success rate benefits Kickstarter because it makes revenue by taking 5% of the total funding amount raised to successfully funded projects. Second, a higher project funding success rate enables more creators to get the necessary funds to execute their project ideas. As such, these creators are more likely to come back to the platform for new crowdfunding campaigns in the future. A higher campaign success rate may also help attract more creators to the service.

Furthermore, a higher success rate may also improve backers' experience, as it suggests that the projects they liked and have pledged funding to have a realistic chance of being fulfilled. Therefore, having a positive experience can help reduce the backer churn rate. As a two-sided platform, Kickstarter is subject to network effects. Therefore, the growth of backers and creators could reinforce each other, thus making the platform more sustainable (Albuquerque et al. 2012).

2 https://www.kickstarter.com/help/stats

One viable way to improve a project's odds to succeed on Kickstarter is to create an effective campaign. Kickstarter has recognized the need for coaching project creators and responded by providing guidelines to project creators on what types of content to include and how the content should be delivered (screenshots in Appendix 1). However, the guidelines exist as a high-level, passive list that provides suggestions not always relevant or otherwise not useful in some contexts. In contrast, an AI service could offer more current, targeted recommendations for creators to help them improve their crowdfunding campaigns.

Creators can leverage multiple media formats and different presentation styles to convey their messages. A campaign that is effective in attracting potential backers and inducing them to back the project is much more likely to be successful in acquiring funding. The goal of this research is, therefore, to look into the content design and provide a large-scale study to understand how content in different formats and styles affect campaign success (in achieving funding goals). While much has been done on project attributes, creator characteristics, and text descriptions, evidence suggests that repeated exposure to the same content increases the familiarity, generates the memory fluency, and helps the audience make quick decisions (Zajonc & Rajecki, 1969; Hawkins & Hoch 1992; Johar & Roggeveen 2007; Ferraro et al. 2008; Spenner & Freeman 2012). Other studies found that new information can provide justifications for consumers to make decisions (Bostrom et al. 1981; Branco et al. 2016).

It is under-examined how to create an effective campaign on Kickstarter and what strategies the project creators should adopt in terms of content design across multiple media channels. An AI-enabled coach has the potential to learn from successful project campaigns more comprehensively and provide guidance for campaign creation. Such a system may benefit both project creators, platforms, and backers by increasing the project success rate. To this end, we propose to address two research questions related to content design on Kickstarter in this study. First, what are the effective strategies for creating successful campaigns across multiple media formats on Kickstarter? Second, how can we design a theory-driven AI-enabled coach for content design on Kickstarter?

3. Theory and Literature Review

In this section, we first present our theoretical foundation based on the theory of persuasion and surveyed recent literature that studied content design factors that affect persuasion. We then review recent work in crowdfunding -- our research context and propose our theoretical framework to explore how various design factors individually and jointly affect crowdfunding performance.

3.1 Theoretical Foundation: Theory of Persuasion

On crowdfunding platforms, the campaign is designed to influence backers' behavior through the delivery of the content. Persuasive design becomes the holy grail of content marketing (Akpinar & Berger, 2017) because compelling content generally depreciates much more slowly (Mueller & Stratmann, 1994). Therefore, what effective content design consists of is among the most critical ones faced by all content creators. To address these questions, we draw upon the broad stream of the persuasion literature that examines how the people's attitudes and decision-making are influenced (Manchanda et al. 2006; Ghosh & Stock 2010; Akpinar & Berger 2017). When these perspectives are coupled with recent advances in machine power and AI tools, there is an excellent opportunity to have a large scale empirical study to understand “influence” through content design.

Persuasion is defined as the intention to convince others. For instance, persuasive information can alter consumers’ preference for a firm and influence consumers’ willingness to pay for a product (Ghosh & Stock, 2010). Arguments and strategies for persuasion commonly take one of three approaches: ethos (appeal to credibility), logos (appeal to logic), and pathos (appeal to emotion) (Dow, 2015). These three types of rhetorical appeals can be traced back to Aristotelian reasoning (Aristotle 1991, Burton 2007), and have been adopted in various business contexts including online advertisement (Darke and Ritchie 2007; Galak et al. 2011; Tanner et al. 2012; Galak et al. 2013; Gonzalez et al. 2014; Jiang et al. 2016; Strebe 2016), movie trailers (Liu et al. 2018), crowdfunding platforms (Greenberg et al. 2013; Muller et al. 2013; Xu et al. 2014; Hou et al. 2019) and so on.   In Table 1, we summarize a list of features that have shown efficacy in prior literature and that operate on the three appeals of persuasion.

Table 1. Features Effective in Persuasion

Strategy	Feature	Media format can be applied to	Used in:	Method	Context	Impact on Persuasion
Appeal to emotion	Aesthetic Features	Image	Jiang et al. 2016; Strebe 2016	Stimuli Manipulation	Online Advertisement	Aesthetic (+)
			Shin et al. 2019	Pre-trained CNN	Social Media	
		Video				

	# of Scenes	Video	Galak et al. 2011; Galak et al. 2013	Stimuli Manipulatio n	Online Advertise ment	Number of scenes (-)
			Liu et al. 2018	Scene Detection Software	Movie Trailer	
	Music	Video	Liu et al. 2018	Sound Processing software	Movie Trailer	Having music (+/-)
	Emotion	Text	Hansen and Hansen 1988; Escalas et al. 2003; Argo et al. 2008; Griskevicius et al. 2009; Yiend 2010; Thales et al. 2012; Hou et al. 2019	Stimuli Manipulatio n	Online Advertise ment;	Joy (+); sad (+)
		Image	Hou et al. 2019	Machine Learning	Crowdfu nding	Joy (+);
		Video				
Appeal to logic	# of Topics	Text	Darke and Ritchie 2007 Fransen et al. 2015; Akpinar and Berger 2017;	Stimuli Manipulatio n	Online Advertise ment	Topics (+/-)
		Image	Akpinar and Berger 2017	Stimuli Manipulatio n	Online Advertise ment	Topics (+)
		Video				
	Whether video has narratives	Video	Darke and Ritchie 2007;	Stimuli Manipulatio n	Online Advertise ment	Have narratives (+)
Appeal to Credibil ity	Historical performan ce		Greenberg et al. 2013; Muller et al. 2013; Xu et al. 2014	Data Collection	Crowdfu nding	Experien ce (+)

	Human Character	Image	Tanner et al. 2012; Gonzalez et al. 2014	Stimuli Manipulation	Online Advertisement	Including human (+)
		Video				

Pathos (appeal to emotion) includes communication strategies intended to stimulate the audiences' emotions. Specifically, this may involve the use of drama, mood, music, and other emotion-eliciting strategies (Matthes et al. 2014; Akpinar & Berger 2017). Aesthetic score (Jiang et al. 2016; Strebe et al. 2016; Shin et al. 2019), number of scenes (Galak et al. 2011; Galak et al. 2013; Liu et al. 2018), use of music (Liu et al. 2018), and emotions (Thales et al. 2012) are believed to stimulate the audience' emotions and influence their decision making and hence are used to measure the effort to appeal to emotion.

Logos (appeal to logic) are methods of persuasion based on evidence and reasoning. Informative appeals are used to improve the product-related outcomes such as evaluations and purchases (Matthes et al. 2014; Akpinar and Berger 2017). The number of topics (Akpinar and Berger 2017), and whether videos contain narratives (Darke and Ritchie 2007), have been developed as measures of appeal to logic.

Ethos (appeal to credibility) are means of convincing someone by the character or trustworthiness of the persuader. Historical performance is used to measure the appeal to credibility (Greenberg et al. 2013; Muller et al. 2013; Xu et al., 2014a). The presence of the human character is believed to influence the degree to which the consumer finds information credible, and therefore, persuasive (Tanner et al. 2012; Gonzalez et al. 2014).

Based on the survey of the literature, we identify the following gaps: First, most of the existing rely on small-scale stimuli manipulation in lab settings to evaluate the efficacy of certain factors. Large-scale data analyses on persuasion strategies with multiple media formats with machine learning methods have been absent from related literature. Second, although videos have been widely adopted by content creators as a media format on various online platforms for persuasion, most prior work has not examined the capability of videos. Third, no previous work has conducted a comprehensive evaluation regarding the efficacy of all three common media formats to persuasion.

3.2 Factors of Successful Projects on Crowdfunding Platforms

Crowdfunding has gained traction as a mechanism to raise resources for entrepreneurial, artistic, or even charity projects. Understanding the factors leading to successful campaigns is essential for crowdfunding platforms and project creators. Extant studies have examined how the characteristics of the creators seeking funding, the project, and the campaigns relate to fundraising success (Sauermann et al. 2019). Higher funding goals (Zhang & Liu 2012) and longer project durations (Burtch et al. 2013) lead to lower chances of successful funding on crowdfunding platforms. Inclusion of an introductory or overview video as a component of a crowdfunding campaign, and generally providing more frequent updates to potential backers both significantly increase the likelihood of successful project funding (Xu et al. 2014a; Muller et al. 2013).

Machine learning approaches have been developed to predictive models for predicting chances of projects being funded (Greenberg et al. 2013; Desai et al. 2015; Hou et al. 2019). Results of these studies empirically demonstrate how a campaign’s early performance within the first few days of launching (Etter et al. 2013), the campaign's readability (Desai et al. 2015), and the project's use of images (Hou et al. 2019) are all significant predictors of campaign performance.

Studies examining the actual content embedded within the visual components of crowdfunding campaigns are still limited. Table 2 summarizes several representative studies. Early studies have primarily focused on text content. Mitra et al. (2014) found out that the language used in the projects can increase the prediction power of consumer engagement using text analysis. Desai et al. (2015) explored the influence of readability, such as emotion conveyed in the textual content on project success rate using machine learning techniques. Besides textual content, Hou et al. (2019) started to investigate the image content on crowdfunding platforms. Specifically, they studied the impact of image aesthetic features (brightness level, color contrast .etc), including humans and including live animals in the image on crowdfunding project success rates. Only Steigenberger and Wilhelm (2018) attempted to include all media formats as a whole by employing two coders to manually rate the degree they perceive emotional cues and information cues from each campaign. Their goal is to understand how different signals interact with each other to influence project funding.

Table 2. Summary of Content Analysis in Crowdfunding Platforms

Paper	Method	Features				Key Findings
		Text	Image	Video	Project as a whole	
Hou et al. 2019	LIWC Object Detection	Emotion in the text description	Image aesthetic features; Objects			Aesthetic attributes of images can predict emotion

			on the images.			in images, and emotions, such as sadness and contentment, can predict the performance of crowdfunding projects.
Steigenberger & Wilhelm 2018	Human - Coded (7-point Likert Scale)				Coders rated how strongly they perceived the signals in a project pitch	Rhetorical signals are an essential complement for substantive signals and may increase or decrease the effect of substantive signals.
Siering et al. 2016	SVM, Neural Networks	The sentiment, and linguistic features (sentence complexity, diversity, etc.)				Content-based cues and linguistic cues are valuable for fraud detection.
Desai et al. 2015	Linguistic Inquiry and Word Count (LIWC)	Linguistic features (certainty, leisure, achievement, etc.)				Successful project pitches are more emotive, thoughtful, and colloquial.
Mitra & Gilbert 2014	Text Analysis (Uni,bi,tri-grams; Phrase frequency)	Phrases used in each category.				The language used in the project has surprising predictive power—accounting for 58.56% of the variance around successful funding
This Study	State-of-the-art AI tools	Topics and emotion.	Use of image; Image contents;	Video aesthetics, music, emotion,	Content and emotion comparison between	The compelling content design and delivery methods across

			Image aesthetic	etc..	different formats.	different media formats.

There is a lack of scalable and machine learning-based research that aims to understand how content in different formats and styles affect campaign success. Using machine learning techniques to utilize different content types is beneficial in understanding the actual impact of content design and providing personalized content design guides to project owners. Recent breakthroughs in AI and improvement in computer power enable capturing, storing, and analyzing video data for various applications, such as face recognition, facial expression recognition, and gesture detection, among others (Lu et al. 2016). Video analyses may contribute to this stream of research by unveiling insights on how persuasion can be carried out with videos.

Focusing only one data type in a crowdfunding platform may not adequately capture the content in the projects and, in turn, generates a bias in studying and designing an effective project campaign. In many contexts, increasing data volume can bring about a greater variety of that data may increase the complexity to analyze and synthesize insights from different data formats. Scrutiny of the messaging delivered by images and video is a necessary input for developing a more comprehensive perspective of the factors that affect campaign success.

3.3 Information Reinforcement and Complement

As appeals to emotion, logic, and credibility can be accomplished via all three common media formats (text, image, and video), multimedia content creators face a unique challenge, which is to understand the interplay between different persuasion strategies and delivery media formats. In addition to persuasion strategies and means to achieve them, it is essential to understand whether there are modes of persuasion more effective through specific media formats and whether it would be advantageous to convey different information or repeat content across multiple formats.

Individuals' mere repeated exposure to stimuli often enhances their attitudes toward the stimuli (Zajonc 1968). Prior studies show that the repeated messages can improve consumers' evaluations and opinions toward the stimuli (Zajonc & Rajecki 1969; Hawkins & Hoch 1992; Johar & Roggeveen 2007; Ferraro et al. 2008). For example, Johar and Roggeveen (2007) suggest that repeated advertisements can impact consumers' prior beliefs. Ferraro et al. (2008) showed that repeated exposure of the brand activities increases the chance that consumers will choose the focal brand because the repeated representation of the brand in the memory generates the fluency. The repetition of messages in the advertising increases its familiarity and subsequently improves the beliefs in the claim (Hawkins & Hoch 1992); therefore, are efficient in persuasive communications. Meanwhile, previous studies also found that repeated exposures can increase counter-argumentation and topic-irrelevant thinking (Cacioppo & Petty 1979). When the stimulus is simple, repeated exposure also can lead to a decrease or an initial increase and then decrease in the positive attitudes toward the stimulus (Saegert & Jellison 1970; Smith & Dorfman 1975).

In contrast to repeated exposure, providing new information in communication is another commonly used strategy to serve the persuasive purpose. Bostrom et al. (1981) demonstrate that informative messages resonate with consumers more than the reinforcement of messages (repeated messages) because informative messages provide more justification. Extensive studies argued about how much information is enough for the consumers (O'reilly 1980; Spenner & Freeman 2012; Branco et al. 2016). Spenner & Freeman (2012) contend that emphasizing fewer product functions can increase consumers' brand stickiness. For example, their study observes the context of cameras and finds that discussing too many functionalities will distract consumers' attention and fails to yield a positive influence on consumer decision-making. Conversely, Oreilly (1980) suggested that compared with information underload, information overload can significantly increase the consumers' satisfaction.

Previous studies found mixed results for both reinforcement (repeated messages) and complement (new messages) on consumers' decision-making process. There is no conclusive answer, whether repeated messages or new messages are more effective in persuasion. A comprehensive content design strategy should entail the complement or reinforcement effects.

3.4 Theoretical Framework

Prior work related to the theory of persuasion shows that text, image, and video features can facilitate effective persuasion via appeals to logic, emotions, and credibility. However, they have not explored the effectiveness of each multimedia format and how each interacts with others. This study examines the respective as well as joint effects of different media formats, including text, image, and videos. This work also adds to the existing literature in crowdfunding research by developing a research framework guided by the theory of persuasion and providing comprehensive guidelines for project creators.

Figure 1 depicts the theoretical framework we propose in the crowdfunding research context. We examine how Kickstarter campaigns comply to appeals to logic, appeal to emotion, and appeal to credibility using a comprehensive list of features from the text, images, and videos. We explore the reinforcement and complement effects amongst campaign videos, textual descriptions, and images.

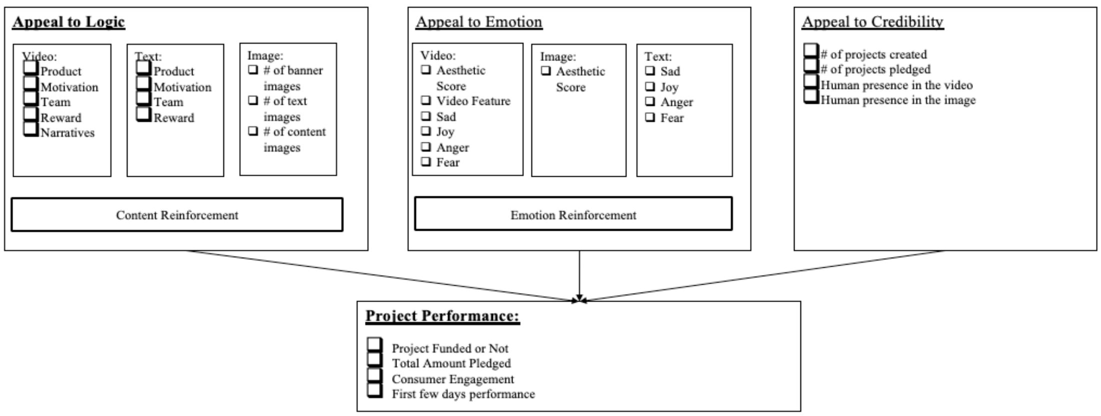


Figure 1. Theoretical Framework

We are interested in studying how the three types of persuasive appeals coupled with content design structure (reinforcement vs. complementary) affect project performance across four different aspects. The first aspect is concerned with whether the project is successfully funded or not, and to what extent by total amount pledged. Second is the level of consumer attention and engagement directed towards the campaign. Tucker (2015) argues that successful campaigns encourage consumers to comment and provide feedback to the campaign itself. We proxy consumer attention and engagement using the number of comments/discussions generated.

Additionally, a high volume of comments suggests that the project attracts notable attention. Lastly, we also examine the first few days' performances of a project. Due to network effects, a project campaign that can gain momentum early on is more likely to be successful. Etter et al. (2013) contended that the time series performance provides high accuracy when predicting the success rate of crowdfunding projects. In other industries, such as the motion picture, the opening weekend collection is used as a significant predictor of total box office revenue. Therefore, it will be interesting to understand how content design can help speed up the momentum.

4. Data and Measurement

4.1 Data

We extract information on all live projects from Kickstarter for the 'Games' category from March 27, 2019, to July 25, 2019.[3] Different categories of projects on Kickstarter enjoy varying levels of success in raising money. Among all project categories, the Games category has attracted the highest amount of funding contributions totaling $1.01 billion, amounting to nearly a quarter of the total successful funding dollars in the platform's history. The success rate for Games projects is 39.36%, which is slightly higher than the average success rate (37%) across all categories on the platform. We select the Games category as our focus for this study due to the volume and variety of data included.

For each project creator, we collect his or her project creation history and project backing history. For each project, we collect the project campaign, including all the

[3] Projects are grouped into 13 broad categories: Art, Comics, Dance, Design, Fashion, Film and Video, Food, Games, Music, Photography, Publishing, Technology, and Theater

videos, images, and text descriptions, project comment counts, project updates, reward levels, and corresponding rewards. In total, we have collected 1,436 projects from March to the end of July. Descriptive statistics for these projects are reported in Table 3.

Table 3. Summary Statistics for Kickstarter Projects

Numeric Variables	**N**	**Mean**	**Min**	**Max**	**Std.**	**Median**
# of Projects Creator Backed	1,436	28.77	0	762	67.404	4
# of Projects Creator Created	1,436	4.143	1	52	6.713	2
Amount Asked	1,436	17,805	1	2,000,000	78,775.2	5,000
Amount Pledged	1,436	36,480.6	1	2,575,192.7	170,414.8	3,193
Goal Completed (%)	1,436	506.5	0	50,600	2,547.7	110.5
# of Backers	1,436	517.79	1	53,643.0	2,275.59	72
# of Reward Levels	1,436	6.267	0	53	4.492	6
Project Duration(days)	1,436	29.92	1	60	11.083	30
# of Updates Before Project Funding Deadline	1,436	4.904	0	45	5.840	3
Comment Intensity (Comment counts)	1,436	126.2	0	18,734	681.464	8
# of Words in Project Text Description	1,436	363.3	0	14,340	651.561	70
Categorical Variables	**N**	**Yes**		**No**		
Funded	1,436	802		634		
Has Campaign Video	1,436	948		488		

The average project has a goal of over $17,000 and receives more than $36,000 in pledged contributions. Projects tend to last for around 30 days. Videos are prevailing in these campaigns, with 948 projects, including campaign videos in their campaigns. Among these 1,436 projects, 802 projects are funded. The average project offers more than six reward levels as incentives for their backers. Creators generally post about five updates before the project funding deadline. Creators within our captured dataset have, on average, started four crowdfunding campaigns and backed 28 projects themselves.

4.2 Empirical Model

We adopted ordinary least squares and logistic regression for the econometric analysis. Equation 1 below illustrates our model.

$$y_i = \beta_0 + \beta_1 AppealtoLogic_i + \beta_2 AppealtoEmotion_i + \beta_3 AppealtoCredibility_i + \beta_4 ContentComparison_i + \beta_5 EmotionComparison_i + \beta_6 Controlariable_i + \varepsilon_i (1)$$

For project i, y_i denotes a set of project performance variables, including the total amount pledged, project funding outcome, consumer engagement intensities, and the first day and the first three days performance in terms of amount pledged and discussions. Variables in the appeal to logic category are the topic loadings from campaign descriptions and video narratives, and the total number of text images. Variables in the appeal to emotion categories include image and video features, and emotion scores of campaign descriptions and narratives. Variables in the appeal to credibility category include the project creators' historical performance. Project amount requested, reward level, project duration, total updates before the project gets funded (for successful projects), or the total number of updates before the last day of funding cycle are also included as control variables. β_1, β_2, β_3, β_4 and β_5 are our variables of interest. It will help us understand the effect of content design in mixed media formats.

4.3 Measuring Constructs

We propose a multi-method approach to studying effective content design through three steps. First, the measurement/diagnosis phase: we employ advanced AI tools to extract features of content in different formats and measure our theoretical constructs. Second, we use econometric models to understand the effect and choose the optimal model. Third, the optimal model provides the foundation of an AI-enabled coach, which can provide useful guidance to content creators. Figure 2 depicts the proposed design for measuring constructs with AI-enabled methods. The following subsections will detail each component.

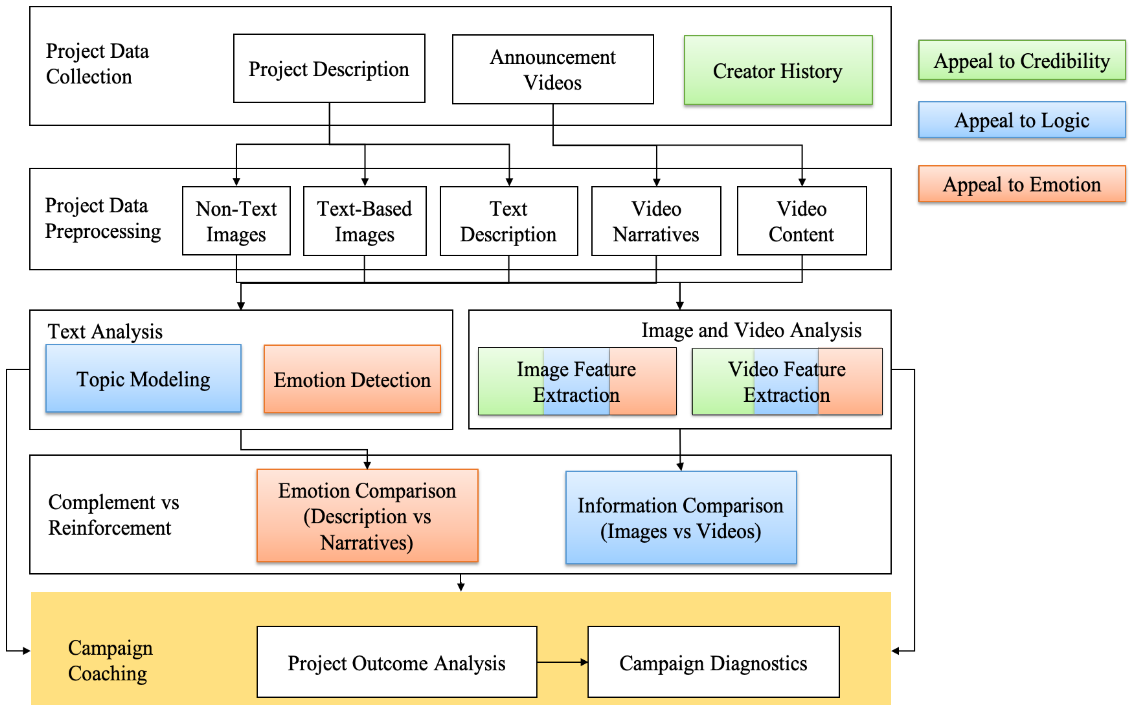


Figure 2: An AI-based Campaign Coaching System

4.3.1 Project Data Preprocessing

To prepare the data for subsequent analysis, we first preprocess the project campaigns and divide the campaign data into three types: text descriptions, project images, and videos. We develop analyses of textual data and image/video data independently.

Project descriptions and images are commonly used in campaigns for persuasion. Some images contain textual information about the project, while others such as website banners and icons might be irrelevant to the campaigns. We first categorize the project images into banner images, text-based images, and non-text images. We use image pixel height and width to filter irrelevant images.

Banner images appear at the top of every web page except the landing page. They are irrelevant to the Kickstarter projects. Text-based images are images composed of text instead of objects. They are used to deliver information in textual format. Non-text images use animations, objects, and shapes to convey information. To identify banner images, we use a simple heuristic based on the dimensions of the images. We calculate the ratio r= image width/ image height. If the ratio surpasses a threshold value (in our study, 2 is the threshold) and the height of the image is less than another threshold (230), we consider it as a banner image.

To classify text-based images, we use an OCR library (Tesseract) to extract text from images. If the number of characters in the image is above a certain threshold (128), we consider it as a text image. We use Google Cloud OCR to extract text descriptions from text-based images. These texts will be included in the subsequent text analysis.

Images that do not fall into the banner or text category are placed into a third category, content image. It is important to note here that there can be an overlap of images that are classified as both text images and banner images. The text classification takes precedence over the banner image classification. This is because our text classification method is generally more reliable than the banner image classification heuristic.

A significant proportion of the project campaigns have included videos. To analyze the videos, we first apply Amazon Rekognition API to extract the narratives from the videos. Their narrative text will be included in the text analysis.

4.3.2 Text Analytics in Kickstarter Campaigns

Text analytics explores the project descriptions, content from text-based images, and narratives from campaign videos. Project text description and content from text-based images are grouped as written form communication, referred to as project textual description in the paper. Narratives from campaign videos are treated as verbal communication. We are interested in examining both written and verbal communications as well as exploring the complementary or reinforcement effect they have.

We leverage the IBM Tone analyzer for emotion detection in text analysis. The tone analyzer is built based on linguistic analysis to detect joy, fear, sadness, and anger in text. The tool returns results for tones whose score meets a minimum confidence threshold of 0.5. We apply the IBM Tone analyzer to all the text data. The table below summarizes the emotion detection from text analytics.

Table 4. Summary of the Emotions in Project Campaigns

	Emotion	Min	Q1	Median	Q3	Max	Mean	Sd
Campaign Textual Description (1436)	Anger	0.0	0.0	0.0	0.0	0.8228	0.0084	0.0722
	Fear	0.0	0.0	0.0	0.0	0.9917	0.0130	0.0940
	Joy	0.0	0.0	0.0	0.6287	0.9830	0.2758	0.3481
	Sadness	0.0	0.0	0.0	0.0	0.7385	0.0161	0.0969
Campaign Video Narrative (948)	Anger	0.0	0.0	0.0	0.0	0.6896	0.0140	0.0889
	Fear	0.0	0.0	0.0	0.0	0.8091	0.0157	0.0960
	Joy	0.0	0.0	0.0	0.6149	1.0	0.2537	0.3192

	Sadness	0.0	0.0	0.0	0.0	0.7262	0.0548	0.1675

As shown in Table 4, Joy is the most common emotion expressed in both campaign textual descriptions and video narratives. On the other hand, other emotions such as anger, fear, and sadness are rarely detected in the campaign.

The platform suggests in their general guideline that creators should discuss the product, team, motivation, and rewards in the campaigns. We evaluate how the project creators elaborate on these four topics. Topic models such as Latent Dirichlet Allocation (LDA) (Blei et al. 2003) have emerged as a powerful tool to analyze document collections in an unsupervised fashion. Topic modeling techniques allow us to measure the allocation of a document to each topic, i.e., the percentage of text in the document belonging to specific topics.

While campaigns can cover various topics, we focus on the discussion related to product, team, motivation, and rewards, as suggested by Kickstarter. We adopt a semi-supervised LDA model, also known as guided LDA, to compute the topic allocation. Guided LDA model incorporates seed words for a small set of topics (e.g., topics of interests) to guide the extraction of specific topics (Jagarlamudi et al. 2012). The resulting topic model can compute the document allocation on these topics more accurately. Seed words for these four topics are included in Appendix 3.

We empirically tested the selection of hyperparameter, the number of topics, for Guided LDA. When the number of topics is 150, the seed words we chose have the highest likelihood to be generated from these four topics of interest. In other words, the four guided topics are generated best matching the seed words and with minimal overlap with the rest 146 topics. We compute the topic allocation for product, team, motivation, and rewards in both project textual descriptions and campaign video narratives and report them in Table 5 below.

Table 5: A summary of topic allocation in campaign descriptions and video narratives.

	Topics	Min	Q1	Median	Q3	Max	Mean	Sd
Campaign Textual Description (1436)	Product	0	0	0.050	0.127	0.577	0.078	0.092
	Team	0	0.065	0.137	0.218	0.727	0.151	0.123
	Motivation	0	0	0.070	0.134	0.703	0.085	0.091
	Rewards	0	0	0.078	0.155	0.801	0.105	0.116
Campaign Video Narrative (948)	Product	0	0	0	0.002	0.788	0.046	0.102
	Team	0	0	0	0.134	0.801	0.075	0.132
	Motivation	0	0	0	0.113	0.668	0.063	0.109
	Rewards	0	0	0	0	0.423	0.010	0.040

A significant portion of the projects covers suggested topics in their campaign descriptions. The most common topic is the team. On average, the allocation to this topic is 0.151. In other words, project creators spend 15.1% of their content on introducing their team in campaign textual description. Rewards is another critical topic in campaign textual description, taking about 10.5% of the content. Compared to campaign descriptions, video narratives have a lower coverage on these topics.

4.3.3 Image and Video Analytics

To understand how project campaigns comply with the theory of persuasion using images and videos, we extract image features and video features using state-of-the-art pre-trained deep learning models. The image features include the number of images in the campaign, the number of text images, the number of content images, the image aesthetic score, and the presence of humans.

The image aesthetic scores are calculated using a pre-trained NIMA (Neural Image Assessment) neural network (Talebi et al. 2018). NIMA yields state-of-the-art performance across many domains for aesthetic scoring tasks. The underlying convolution neural network architecture for NIMA is NASNet, which is designed to rate images based on how aesthetically pleasing they appear. Given an image, NIMA assigns an image score from one to ten for the aesthetic quality of the image. We use the mean aesthetic score of all the images in a campaign to represent its image aesthetic score.

We employ Amazon Rekognition API to extract the image and video features. Amazon Rekognition can identify the objects, people, text, scenes, and activities in images and videos. With this API, we detect whether an image has a human presence. The image human presence feature is the percentage of images with humans for a given campaign. Table 6 below provides a summary of the image features.

Table 6. Summary Statistics for Campaign Image Features

Numeric Variables	N	Mean	Min	Max	Std.	Median
# of Images in Campaign	1436	17	0	164	16.850	12
# of Text Images	1436	2.781	0	38	4.464	1
# of Content Images	1436	9.35	0	119	10.046	7
Image Aesthetic Score	1436	3.142	0	3.918	0.88	3.381
Image Human Presence	1436	0.411	0	1	0.078	0.174

The video features include video aesthetic score, the presence of humans in the video, containing narratives and music, the number of scenes, and the video duration. With the Amazon Rekognition API, we generated video features such as whether the video includes humans, the percentage of screen time with humans, whether the video contains narratives, contains music, the total number of scenes, and video duration.

To assess the aesthetic scores of project videos, we sample frames from videos before we can employ the NIMA model. There are two standard techniques to sample frames from videos: sample per time interval and sample per scene. In our study, we found that a sample rate of a frame every ten seconds works for most Kickstarter videos[4]. Once the frames are sampled, each sample gets its own NIMA score. The minimum, maximum, average, and standard deviation of the NIMA scores in the samples for each project are added to our dataset.

Exploratory factor analysis (EFA) is performed to identify the latent relational structure among the video features and narrow down to a smaller number of features (Child, 1990). Two factors are identified by using 0.3 as the cutoff point on loading. Factor 1 (video features) includes whether the video contains real humans, the duration of the human showing in the video, whether the video consists of music, the number of scenes, and the video duration. Factor 2 (Aesthetic score) includes the minimum, maximum, and mean of the video aesthetic score. Table 7 below provides summary statistics of the video features.

Table 7. Summary Statistics for Projects Video Features

Variable	**N**	**Mean**	**Min**	**Max**	**Std**	**Median**
aestheticScoremin	948	3.352	2.288	6.491	0.901	2.618

[4] If we sample after a larger interval, we start to lose information since some of the Kickstarter videos only last a few minutes. We also experimented with sampling per scene. However, that results in a compounding of classification errors i.e. if the number of scenes is misclassified, the error carries over to the aesthetic analysis.

aestheticScoremax	948	5.821	4.486	6.940	0.407	5.828
aestheticScoremean	948	4.939	3.480	6.508	0.372	4.963
aestheticScoremedian	948	5.072	2.607	6.513	0.395	5.112
aestheticScorestd	948	0.658	0	1.472	0.303	0.678
VideoDuration	948	148.2	0	8561	296.887	108
Percentage_human	948	0.117	0	0.981	0.185	0.026
Scenes	948	22.74	1	488	26.915	16
Categorical Variables	**N**	**Yes**		**No**		
RealHuman	948	406		542		
Music	948	512		436		
Narratives	948	859		89		

4.3.4 Complement and Reinforcement

As a project campaign can be delivered in multiple formats, a comprehensive recommendation for project creators should include the complement and reinforcement effect across different content formats. To this end, we examine the effectiveness of covering suggested topics in campaign descriptions, in video narratives, and both of them. We also investigate the impact of emotions in descriptions, narratives, and in both of them. Besides, we measure how many images in the project campaign description come from the project video.

To identify the overlap between campaign images and videos, we match the video frames to images used in the project campaign description. We apply the SIFT (Scale-Invariant Feature Transform) algorithm to extract the features from the edges in the image, and then use the RANSAC (Random Sample Consensus) algorithm to find a projective transformation between the features in the frame and the image. If an image appears in the video, it signals a potential reinforcement because the text around the image usually explains the image. For additional analysis, we also calculate the similarity between campaign descriptions and video narratives. The control variables include the project creator information (appeal to credibility), the duration of the project, the funding goal, and the number of updates before being funded. Table 8 shows the summary statistics for New and Repeated Information.

Table 8. Summary Statistics for New and Repeated Information

Variable	N	Mean	Min	Max	Std	Median
Percentage_RepeatedInfo	948	0.341	0	1	0.280	0.300
Percentage_NewInfo	948	0.659	0	1	0.281	0.700

5. Empirical Results

5.1 Impact of Image and Video Features

We first examined the results for the regressions with image and video features as the variables of interest (Table 9). As shown by the significant coefficients, video aesthetic scores have a positive impact on the total amount pledged and the total number of comments. This finding is consistent with prior studies that demonstrate product aesthetics can change consumers' attitudes and improve consumers' product evaluation (Bloch 1995; Page and Herr 2002; Strebe 2016). A project campaign video with a high aesthetic score can positively influence the funding outcome and backer engagement in the first few days and throughout the entire funding period. Backers tend to discuss the videos more in the discussion board if the videos are visually appealing, and therefore, impact the project success rate. Other video features such as whether the video contains real humans, the duration of the human showing in the video, whether the video includes music, the number of scenes, and the video duration do not have a significant impact on project outcome and backer engagement.

The image aesthetic scores have a significant and positive impact on the total amount pledged in the first three days and over the entire funding period. However, it does not have a significant effect on the total number of comments. The number of text-based images and the number of content images have a significant and positive on total amount pledged, the total number of comments, and the first day pledged percentage. Including more images and using text images can make the campaign visually more appealing and improve the campaign outcome.

Table 9. Video feature and image feature results

	Total Amount Pledged -OLS-	Funded or Not -Logistic-	Total Comments -OLS-	First Day Pledged -OLS-	Three Day Pledged -OLS-	First Day Comments -OLS-	Three Day Comments -OLS-
Video Aesthetic Score	1.289** (0.560)	0.807 (1.143)	1.118** (0.488)	0.655 (0.623)	1.311** (0.610)	0.555 (0.563)	1.090* (0.600)
Video Features	-0.011 (0.070)	0.047 (0.137)	0.092 (0.061)	0.080 (0.074)	0.028 (0.073)	0.115* (0.067)	0.082 (0.071)
Image Aesthetic Score	0.364* (0.196)	-0.234 (0.687)	0.258 (0.170)	0.123 (0.172)	0.339* (0.186)	0.003 (0.155)	0.109 (0.183)
Image Human Feature	-0.021 (0.093)	0.249 (0.269)	-0.115 (0.081)	0.118 (0.115)	0.041 (0.121)	-0.140 (0.104)	-0.172 (0.119)
Text Images	0.260** (0.102)	0.248 (0.224)	0.336*** (0.089)	0.365*** (0.106)	0.297*** (0.104)	0.375*** (0.096)	0.401*** (0.103)
Content Images	0.460*** (0.125)	0.113 (0.266)	0.339*** (0.109)	0.321** (0.142)	0.222 (0.142)	0.173 (0.129)	0.090 (0.140)
Amount Asked	0.459*** (0.059)	-0.986*** (0.156)	0.265*** (0.052)	-0.406*** (0.064)	-0.432*** (0.063)	0.208*** (0.058)	0.256*** (0.062)
Total Reward Level	0.051*** (0.015)	0.072** (0.034)	-0.013 (0.013)	0.022 (0.015)	0.036** (0.015)	-0.017 (0.014)	-0.012 (0.015)
Project Duration	-0.028*** (0.008)	-0.044** (0.020)	-0.018** (0.007)	-0.028*** (0.010)	-0.031*** (0.009)	-0.021** (0.009)	-0.027*** (0.009)

Total Updates Before Funded	0.093*** (0.014)	0.430*** (0.060)	0.113*** (0.012)	0.066*** (0.016)	0.079*** (0.016)	0.066*** (0.014)	0.087*** (0.015)
Creator Project Backed	0.176*** (0.057)	0.136 (0.123)	0.176*** (0.049)	0.119** (0.059)	0.095 (0.058)	0.127** (0.054)	0.132** (0.057)
Creator Project Created	0.370*** (0.143)	0.902** (0.351)	0.568*** (0.124)	0.527*** (0.150)	0.511*** (0.148)	0.494*** (0.136)	0.565*** (0.145)
Constant	-1.172 (1.100)	4.488 (3.055)	-4.092*** (0.958)	3.581*** (1.144)	3.185*** (1.133)	-2.417** (1.033)	-3.040*** (1.114)
N	429	429	429	288	279	288	279
R-Square	0.615		0.632	0.592	0.617	0.565	0.615
Adj R-Square	0.573		0.592	0.523	0.549	0.490	0.547
AIC		368.218					

5.2 Impact of Topic Allocations in Videos and Texts

Table 10 illustrates the impact of topic allocation in video narratives and textual description on project performance variables. Dummy variables such as DescriptionYesOrNo indicate whether narratives and text descriptions include the given topic. The topic loading variables indicate the coverage of the given topic. Both dummy variables and loadings are included in the regression analysis. The coefficients of the dummy variables suggest whether project creators should include the topic, while the estimates of the loading variables indicate how to allocate the topics in the campaign.

Kickstarter platform recommends the content creators to emphasize on product, team, motivation, and rewards in the campaign messages. According to our empirical analysis, we observe that including a discussion of the project team has no significant impact on any project performance variable. As shown by the negative and significant coefficients, more discussion of the team in descriptions may lower the amount pledged and the number of comments received. The results suggest that it will hurt the project funding outcomes as well as the backers' engagement when the project video and project textual description consist of too much team information. Introducing rewards in the textual description has a significant and positive impact on the amount pledged and the total number of comments. However, the benefit of including reward discussion in the videos is not substantial. The best place to promote rewards is the campaign description. The discussion of products in campaign descriptions has a positive impact on the total number of comments but does not significantly impact the project funding outcomes. Expressing motivations in the video can positively influence project funding outcomes.

Table 10. Informativeness Results

	Product	Team	Motivation	Reward
Amount Pledged				
DescriptionYesOrNo	Insignificant	Insignificant	Insignificant	+
DescriptionLoading	Insignificant	-	Insignificant	+
NarrativeYesOrNo	Insignificant	Insignificant	Insignificant	Insignificant
NarrativeLoading	Insignificant	Insignificant	+	Insignificant
Project Funded Or Not				
DescriptionYesOrNo	Insignificant	Insignificant	Insignificant	+
DescriptionLoading	Insignificant	Insignificant	Insignificant	Insignificant
NarrativeYesOrNo	Insignificant	Insignificant	Insignificant	Insignificant
NarrativeLoading	Insignificant	Insignificant	Insignificant	Insignificant
Total Comments				
DescriptionYesOrNo	+	Insignificant	Insignificant	+

DescriptionLoading	+	-	Insignificant	+
NarrativeYesOrNo	Insignificant	Insignificant	Insignificant	Insignificant
NarrativeLoading	Insignificant	-	Insignificant	+
First Day Pledged Percentage				
DescriptionYesOrNo	Insignificant	Insignificant	Insignificant	+
DescriptionLoading	-	-	Insignificant	+
NarrativeYesOrNo	Insignificant	Insignificant	Insignificant	Insignificant
NarrativeLoading	Insignificant	Insignificant	Insignificant	Insignificant
First Three - Day Pledged Percentage				
DescriptionYesOrNo	Insignificant	Insignificant	Insignificant	+
DescriptionLoading	Insignificant	-	Insignificant	+
NarrativeYesOrNo	Insignificant	Insignificant	Insignificant	Insignificant
NarrativeLoading	Insignificant	Insignificant	Insignificant	Insignificant
First Day Comments				
DescriptionYesOrNo	+	Insignificant	Insignificant	+
DescriptionLoading	Insignificant	Insignificant	Insignificant	+
NarrativeYesOrNo	Insignificant	-	Insignificant	Insignificant
NarrativeLoading	Insignificant	Insignificant	Insignificant	+
First Three-Day Comments				
DescriptionYesOrNo	Insignificant	Insignificant	Insignificant	+
DescriptionLoading	Insignificant	Insignificant	Insignificant	+
NarrativeYesOrNo	Insignificant	Insignificant	Insignificant	Insignificant
NarrativeLoading	Insignificant	Insignificant	Insignificant	Insignificant

5.3 Impact of Emotions in Videos and Texts

To further examine the impact of emotions, we conducted regressions using emotions as variables of interest. Similar to the content variables, we use dummy variables and loading variables to show different influences. The results are reported in Table 11. Videos can make emotions more expressive. Having joyful or sad emotions in the project video can positively influence the backers' engagement and project funding success rate. The results are consistent for both dummy variables and loading variables. Stronger emotions tend to influence consumers' engagement positively. This finding is consistent with prior studies, which demonstrated that teachers' sad emotions could stimulate students' class participation (Zhou et al. 2019).

Table 11. Emotion results

	Total Amount Pledged -OLS-	Funded or Not - Logistic-	Total Comments -OLS-	First Day Pledged Percentage -OLS-	Three Day Pledged percentage -OLS-	First Day Comments -OLS-	Three Day Comments -OLS-
Video - Joy	InSig	InSig	+	InSig	InSig	InSig	InSig
Video - Sad	InSig	+	InSig	InSig	InSig	InSig	InSig
Video - Fear	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Video - Anger	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Text - Joy	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Text - Sad	InSig	InSig	InSig	InSig	InSig	InSig	InSig

Text - Fear	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Text - Anger	InSig	-	InSig	InSig	InSig	InSig	InSig
Emotion Loadings							
Video - Joy	InSig	InSig	+	InSig	InSig	InSig	InSig
Video - Sad	InSig	+	InSig	InSig	InSig	InSig	InSig
Video - Fear	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Video - Anger	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Text - Joy	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Text - Sad	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Text - Fear	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Text - Anger	InSig	-	InSig	InSig	InSig	InSig	InSig

5.4 Complement and Reinforcement Effects

Table 12 shows the result of the relationship between campaign videos and project textual descriptions. The variable generated from the comparison between images and videos illustrates the content similarities. Emotion comparison is captured by the interaction term of the emotion indicators between two different project campaign formats. We find that the reinforcement effect is potent in persuasive communication. Both topic-related reinforcement and emotion-related reinforcement can significantly improve project crowdfunding outcomes.

Table 12. Complement vs. Reinforcement Effect

	Total Amount Pledged -OLS-	Funded or Not -Logistic-	Total Comments -OLS-	First Day Pledged Percent -OLS-	Three Day Pledged percent -OLS-	First Day Comments -OLS-	Three Day Comments -OLS-
RepeatedInformation_YesOrNo	InSig	InSig	InSig	InSig	InSig	InSig	InSig
RepeatedInformation_Loading	+	InSig	InSig	+	InSig	InSig	InSig
NewInformation_YesOrNo	InSig	InSig	InSig	InSig	InSig	InSig	InSig
NewInformation_Loading	-	InSig	InSig	InSig	InSig	InSig	InSig
Emotion_Joy	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Emotion_Sad	InSig	InSig	+	+	InSig	InSig	InSig
Emotion_Anger	InSig	InSig	InSig	InSig	InSig	InSig	InSig
Emotion_Fear	InSig	InSig	InSig	InSig	InSig	InSig	InSig

The results show that appeal to emotion and logic can play a significant role in influencing project outcomes, and the impact is different across different media types. Figure 3 presents a summary of the results. We run 22 models that gradually include different variables (Appendix 5a) and test for the model performance for the project funding outcomes: total amount pledged.

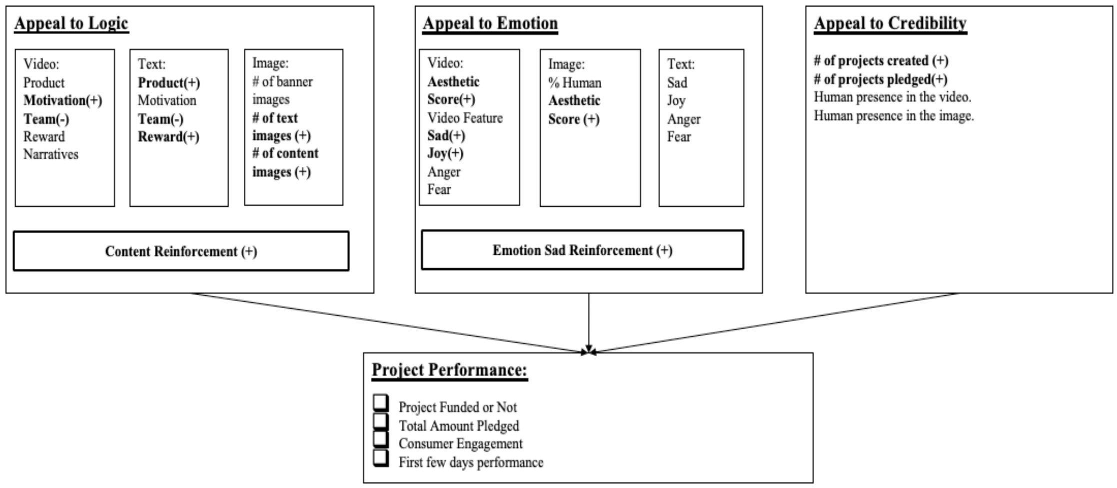


Figure 3. Result Summary Graph

We report 8 model results in Table 13 [5]. The models selected here are 1) model 1 only include project variables; 2) model 2 adds the textual description; 3) model 3 only consists with project variables and the image features; 4) model 4 adds the video features; 5) model 5 adds the video content information; 6) model 6 adds content comparison; 7) model 7 adds emotion comparison and model 8 includes everything. From Table 9, we can contend that video variables significantly increase the explainable power of the models. The results indicated that including video features could significantly increase the explanatory power, which suggests that video information is important to assess the project outcomes.

[5] The detailed 22 model results are listed in Appendix 5b.

Table 13. Model Comparison Result
(LogAmountPledged)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Project Variable (include appeal to credibility)	√	√	√	√	√	√	√	√
Text Topic Loading		√			√	√	√	√
Text Emotion Loading		√			√	√	√	√
Image Data			√	√	√	√	√	√
Video Features				√	√	√	√	√
Video Topic Loading					√	√	√	√
Video Emotion Loading					√	√	√	√
Content Comparison						√		√
Emotion Comparison							√	√
R-Square	0.483	0.564	0.536	0.541	0.601	0.603	0.612	0.615
Adjusted R-Square	0.479	0.555	0.527	0.528	0.572	0.573	0.571	0.573

6. Practical Application: An AI - Coach

Our approach and analyses provide the foundation for an AI-enabled coach that can guide content creators to improve their content design. This AI-enabled Coach is designed based on three types of data and built upon the empirical results from econometric models. We illustrate the use of such an AI-enabled coach using an example we randomly picked. Consider the project "SkyHearth: An Indie Skyrim Game" as an example. This project is an open-world RPG game that the players can explore the game environment and create their world in the game. The project was started on April 30, 2019, and lasted for 30 days. The funding goal is $10,000. The creators have backed five projects and created one project before. The project campaign includes one campaign video, 18 images, and 141 characters long text description. We first perform a diagnosis of the project content and generate an output shown in Table 15.

We can see that the project creators include a 169 seconds long campaign video. The video has background music as well as human narratives. 38.38% of the video time has a human presence. The aesthetic score of the campaign video is 5.111 out of 10. The creators cover the topic product, team, and motivation, and show happiness in the video. Topics on product, team, reward, and motivation are all covered in the campaign textual description. Project creators express joy in the text campaign as well. The project campaign includes 11 banner images and seven content images. None of the content images includes human beings. The project creator implements an information-compliment content design strategy in which they provide 94% new information in project textual description in addition to the campaign video. Emotion Joy is reinforced in two formats. The content gap graph is presented in Appendix 6.

Table 14. Project Diagnosis and Prescription: An example

Appeal to Emotion: Video	Image	Text
Music: Yes Scenes: 43 Video Duration: 169 seconds Human: Yes % Human: 38.38% Aesthetic Score Min: 2.608 Aesthetic Score Max: 5.829 Aesthetic Score Mean: 5.111 Video - Joy: 0.783 Video - Sad: 0 Video - Anger: 0 Video - Fear: 0	% Human: 0 Aesthetic Score Min: 1.463 Aesthetic Score Max: 6.032 Aesthetic Score Mean: 3.605	Text - Joy: 0.707 Text - Sad: 0 Text - Anger: 0 Text - Fear: 0
Appeal to Logic: Video	Image	Text
Video - Product: 0.21 Video - Team: 0.264 Video - Reward: 0 Video - Motivation: 0.238 Narratives: Yes	# of banner images: 11 # of text images: 0 # of content images: 7	Text - Product: 0.017 Text - Team: 0.222 Text - Reward: 0.03 Text - Motivation: 0.149
Appeal to Credibility: Number of Projects Creator Pledged: 5 Number of Projects Creator Created: 1	**Project Information:** Project Campaign Video: Yes Project Description Length: 16 Project Campaign Length: 125 Project Total Images: 18 Project Text Images: 0 Project Images: 7	**Overall Campaign Information:** Information_Repeated: 0.056 Information_New: 0.944 Emotion_Joy: Yes Emotion_Sad: No Emotion_Anger: No Emotion_Fear: No
Overall Prediction: Based on your current campaign profile, the predicted possibility your project will be successfully funded is **7.53%**; the predicted amount of money pledged is **$1525.38**, which is $8,474.62 less than the required amount[6].		

6 The predictions are based on the best performance model out of 22 models (lowest AIC and highest R-

Overall Suggestion:

What is done well:
1. Good job in including the "happy emotion" in the video
2. A good discussion of motivations in the video to start this project.

Things that can improve your success rate:
1. Increase content on your product in the text description.
2. List your rewards in the text description.
3. If applicable, introducing some "sadness" in your video content can help engage potential backers and increase their chance of backing.
4. Please reinforce your information in both videos and text descriptions.
5. Decrease time spent introducing your team in both video and text content.
6. Make your campaign pictures more visually appealing.

7. Contribution and Conclusion

In this paper, we propose a systematic, theory-driven approach for assessing and guiding the creation of multimedia content strategy. We then contextualize this development within a study examining the funding success of Kickstarter campaigns. Kickstarter campaigns were the chosen context to highlight gaps and challenges that exist in the current literature. Many content design studies identified in literature often make use of only one or a few similar data types (e.g., text or images exclusively). However, by drawing upon theories from Aristotelian reasoning and utilizing Kickstarter campaign data that incorporates many distinct data formats including videos, images, text descriptions, and associated project creators' information, our proposed method lays a foundation for designing an effective multimedia strategy with means to evaluate a vast range of data types. Scrutinizing content design in Kickstarter campaigns necessitates a mixed-methods approach, including textual topic modeling, emotion detection, exploratory factor analysis, regression analysis, and other machine learning and

squared and adjusted R-Squared).

econometric approaches. This work also contributes to the literature on persuasion and content design by providing a large-scale empirical study examining the effectiveness of different forms of persuasion.

Concerning the specific context of Kickstarter, this study explores prescriptions for multimedia strategy in crowdfunding campaigns. Results articulate how campaign creators can effectively communicate project visions and funding needs. This work scrutinizes what types of content design styles are most effective in achieving project success, increasing consumer attention and engagement, and gaining momentum. Overall, we find that the reinforcement effect is potent in persuasive communication. Both topic-related reinforcement and emotion-related reinforcement can significantly improve project crowdfunding outcomes. Projects that articulate the logic behind why the project is desired and also communicate rewards for backing the project tend to be more successful. Our results also indicate campaign aesthetics may not affect the project outcomes directly, but they do impact consumer attention and engagement.

Our analysis carries several implications for both research and practice. This research is among the first to attempt theory-driven consolidation of distinct data sources spanning unstructured text, image, and video data. Previous studies often lack the comprehensive utilization of multiple data sources, or they may take an ad hoc approach towards combining the different data streams without properly grounding design logic in existing theory. However, the theory-driven design articulated within this study allows for the development of a multimedia strategy that is logically-grounded in the existing theory. The capabilities developed and introduced in this work have widespread usefulness given the ubiquity of social media and multimedia content, as they manage to synthesize insights from different data formats and make personalized recommendations.

Furthermore, this study contributes to crowdfunding literature. Our study is among the first to evaluate the impact of video content on crowdfunding project success. Specifically, we investigate what video characteristics are essential for increasing the persuasive power of crowdfunding campaigns. We also explore mechanisms for how to synergize content design among different information channels, particularly video and text. A multimedia strategy for crowdfunding campaigns is conceptualized in this paper to provide guiding prescriptions for enhancing content design through these means. This study introduces numerous recent developments from computer science to crowdfunding literature and the overall information systems discipline (e.g., NIMA, Scale-invariant feature transform, and Random Sample Consensus).

This study also details several practical implications. The approach undertaken by this research can be used as a working template for future efforts by organizations or researchers attempting to formulate a multimedia strategy. Integration of various AI services and using them to analyze existing datasets can yield valuable and personalized insights to users. We instantiate this concept with an example of a multimedia strategy for guiding Kickstarter campaign creators. The pipeline developed in this research can readily be applied to other contexts where multimedia strategy is required.

This study does have some limitations. Only data from Kickstarter's "game" project category was utilized for training our model. To develop a content design strategy, solely relying on game category data may not be adequate to develop a model that can represent content in other disciplines. However, this issue is easily solvable by following the exact procedures described in this study on additional data. Another limitation is that we operationalize emotion detection in this study by merely transcribing spoken audio in videos to text, and then performing text mining. Some emotion detection based on recognizing human facial expressions directly in videos was attempted, but the results were dubious for two reasons. First, 406 out of 948 videos do not have any humans in the video. Second, many videos contain animated characters that become detected by the facial recognition algorithm and lead to inaccurate measures of human facial expressions. Both of these techniques were implemented using the Amazon Rekognition API but failed to provide meaningful results. Overcoming these limitations can provide fruitful directions for future research.

CHAPTER 3

THE IMPACT OF INTERNET WATER ARMIES ON ONLINE AND OFFLINE MEDIA: AN ANALYSIS OF MOVIE INDUSTRY

Abstract

Motivated by the growing phenomenon of purchasing Internet Water Armies to promote on social media, this paper examines the impact of using the Internet Water Armies as a way to perform social media promotion. In particular, we study whether using the Internet Water Armies helps sales and under what conditions it helps. Using a dataset from the movie market of China, we employ the method of panel vector autoregression (PVAR) to examine the relationship among Internet Water Armies, product financial outcomes, and media performances on both traditional media channels and new media channels. We further apply a dynamic matching approach to address the identification issue. Through analysis at both the post level and fans level, we find that the Internet Water Armies helps product sales. The effect is largely reflected through changing the number of emotional fans. Furthermore, the earlier to purchase the Water Armies, more haters, likers, and neutral fans it can attract. This work contributes to both the literature and practice by empirically exploring the impact of purchasing the Internet Water Armies.

Keywords: Internet Water Armies, Social Media Promotion, PVAR, Dynamic Matching

Hiring Water Armies is Part of Every Public Relations Manager's toolkit.

- LA Times

Popularity has a price.

- NY Times

1. Introduction

Internet Water Armies are defined as a group of internet ghostwriters who get paid to post anything online with some particular content. When using the Internet Water Armies, a large number of well-organized people "flood" the Internet with purposeful posts, reviews, comments, and other actions. According to a survey conducted by Xinhua News, 74 percent of the people interviewed believe the Internet Water Armies are ubiquitous on Chinese social media websites and is widely used by public relations managers. According to a spokesman of Weibo, the top social media platform in China (also known as the Twitter of China), around 40 percent of the trending hashtags on Weibo are created by Water Armies every day[7].

The Internet Water Armies are not just the tool used by public relations managers, but it is becoming a culture in China. The battleground for promoting and branding has shifted to the Internet because according to the statistics generated by the China Internet Network Information Center (CNNIC), there were around 564 million Internet users in China in 2019. Therefore, it has incubated a new business under the table, hiring Internet Water Armies to instill social media with fake content. For example, a famous TV drama

[7] http://www.sixthtone.com/news/1001904/guns-for-hire-chinas-social-media-militia-engage-on-command

show in China called "Eternal Love," generated 1.4 billion online clicks within a single day. However, the entire population of China is about 1.3 billion and there are only 564 million Internet users. Therefore, the authenticity of the number shown in the internet video platform is questioned. The truth is in the Water Armies market; the corporations can spend as little as 100 RMB (around $15) to purchase 100,000 clicks. With the excessive amounts of clicks, this TV drama show has attracted the attention of many viewers causing it to be nominated for and awarded several famous TV play awards and enabled it to maintain the record of being the most viewed TV play currently online. In the offline market, where Water Armies cannot fabricate the record, out of the 30 days it was played on television, Eternal Love has been the number one consumed TV play for the last 19 days. The strategy of flooding the online video platforms with "fake clicks" at the beginning of the show generates the buzz and attracts viewers' attention to the show. However, using the Internet Water Armies are not free of drawbacks. Despite such a success of the TV play, audiences consistently question the quality of the TV play/movies by the same production team. Every time the leading actress has a new project released or appears on the hot search rank, audiences will comment "stop buying the Water Armies," or "how much Water Armies did you purchase?" and so on. The second sequel in the Eternal Love series also attracted many questions.

The term Internet Water Armies comes from China, but the phenomenon of having "fake followers" and "fake posts" is universal. Facebook claimed that five percent of their online users are fake accounts and removed three billion fake accounts over six months (Romm, 2019). Some of the fake accounts belong to an obscure American company called Devumi. This company has collected millions of dollars by selling social media followers and retweeting to celebrities, businesses, and anyone who would pay for it and wanted to become popular or exert influence online (Confessore et al., 2018).

Using the Internet Water Armies becomes a popular promotion approach not only because the average consumption time of online media is increasing over the years, but also because the cost-efficiency of using social media promotions has increased. Social media promotion is cheaper than any form of advertising available today as it is one of the only forms of media that can reach over 1,000 people for less than $3[8]. Therefore, purchasing social media fans and social media posts becomes more and more popular, and the Internet Water Armies becomes a culture all over the world. However, the actual impact of using the Internet Water Armies are yet to be determined. On one hand, massive posts from the Internet Water Armies can boost the awareness effect and then lead to more organic posts and, therefore, make the promoted product become a hit on social media networks and even on the online media press. On the other hand, the organic posts have declined over the years because too many "fake" posts occupied their social network and consumers have lost their trust in online media channels. In addition, once the audiences realize the buzz was created by the Water Armies, they will lose trust for the company and become a long-term hater. Consequently, the impact of using the Internet Water Armies becomes an empirical question. Our study attempts to fill the

[8] https://www.lyfemarketing.com/traditional-media-versus-social-media/

literature gap by answering whether ***using the Internet Water Armies affects sales and under what conditions it helps sales?***

In order to answer these questions, we need to consider several aspects of the Water Armies phenomenon. First, the Internet Water Armies attempts to flood the social media platform. The impact may carry over to other media formats as well. Several scholars studied the impact of different forms of media on the financial outcomes. Goh, Heng, and Lin (2013) found that user-generated contents impact consumers' purchase decisions more than marketer generated contents. The interactions between traditional media and social media attract the attention of many scholars as well. The concurrent effects of multiple forms of media have been found in the entrepreneur field (Greenwood and Gopal, 2015), microlending marketplace (Stephen and Galak, 2012), and music sales (Dewan et al, 2014). The Internet Water Armies are almost exclusively founded through social media. Limited spaces on traditional media makes it costly to buy out massive traditional media posts. Stephen and Galak (2012) found that the mechanisms by which social media influences product sales are the influences through traditional media. Even when the Internet Water Armies exists only in social media, the influence can be carried over to other media channels, e.g. traditional media.

Second, the influence of using the Water Armies may be based on the purchased formats. Water Armies can provide a variety of services. Unlike fake reviews or fake news, the Water Armies can implant anything fake to the platform. Water Armies can help inject fake clicks, fake fans, fake posts, and fake reviews. The marketing team can hire Water Armies to mass-produce positive comments, posts, and ratings for them, or negative ones for competitors. Water Armies are also capable of deleting the negative posts by repeatedly replying with unappropriated content to force the platform monitor to remove the thread. Further, zombie fans are also well applied. The marketing company can purchase zombie fans to follow a certain account or a certain topic. Buying Water Army fans and Water Armies posts is different in terms of weights and intensity. In other words, the difference between buying Fans and buying posts is that in Fans many people spread the same information versus in using posts, one person keeps repeating the same information. It is reasonable to contend that the first scenario of many people spreading the same information is more persuasive as there is an old saying “three liars make a tiger.” Besides, publishing multiple similar contents by one account is a major criterion used by the social media platform to detect Water Armies. The platform can filter these posts to avoid the flood of Internet Water Armies. Furthermore, when the audience sees many similar posts coming from a few accounts, they can easily link these contents with the Internet Water Armies and form negative impressions. On the other hand, Hunter and Zaman (2018) found that the key in shifting public opinions is the number of posts a bot generated, not the number of bots there were in a political setting. Whether or not the effect of massive posts dominates the effect of tremendous accounts is ultimately an empirical question, for which we hope to provide some insights through our studies.

Next, one critical concern arising from using the Internet Water Armies are how it will influence the sentiment of the organic social media users. Using the Water Armies may bolster the presence and awareness of certain products/celebrities in the short run, but the illusion does not necessarily help in the long run. Substantial posts appearing on social media within a short timeframe will create the buzz and keep the discussion going. However, once the consumers realize that the buzz is generated by the Internet Water Armies, the consumers will feel they were manipulated. Feeling manipulated will increase consumers' negative incentives in staying tunes about the product/celebrities (Rudinow, 1978).

Moreover, consumers' state of knowledge is dynamically changing, especially on information goods, such as movies. The nature of information goods is they are also experienced goods. The fact that movies are experienced goods makes online word of mouth extremely valuable to promote the movies. Previous studies have declared that both pre-release social media buzz (Xiong and Bharadwaj, 2014) and post-release eWOM (Duan et al., 2008; Zhu and Zhang 2010) are significantly associated with product sales. Therefore, it is important to disentangle the temporal influence of using the Internet Water Armies across different stages. Buying the Internet Water Armies in the early stage can influence the opinions and generate awareness effects, while in the late stage, it can keep the discussion warm and generate the reinforcement effect. When to purchase the Internet Water Armies during product life cycle becomes an empirical question.

To address these questions, we have cooperated with a major entertainment data consulting company in China and assembled data on movies released in in 2017 and 2018. We constructed a daily panel data to perform the panel vector autoregression (PVAR) model. The advantages of using PVAR are we can assess the nature of bidirectional relationships between all pairs of dependent variables as all the dependent variables, box office sales, offline media activities, and online media activities are treated jointly endogenous (Dewan and Ramprasad, 2014; Ahlfeldt et al., 2015). Furthermore, we can identify the lag effects and understand the dynamic relationships (Duan et al., 2008). Offline media activities capture traditional media activities including newspaper articles and magazine articles. The online media activities show the online media press activities, the internet Water Armies' posts, and organic social media posts. Our identification strategy depends on the purchasing habits, in that some movies distributors buy internet armies while some don't, and different movies distributors purchase Internet Water Armies at different stages. We are able to take the approach used from the literature on treatment effects to measure the impact of purchasing internet armies together with a dynamic matching strategy.

We found that the Internet Water Armies helps product sales at both post-level and fans-level. The Internet Water Armies impacts product sales both directly and indirectly by changing the number of emotional fans. For purchasing strategies, hiring the Water Armies one week before release would spread the movie information and bring more haters, likers, and neutral fans. However, the Water Armies will not work if the purchase occurs in the third week after the movie release.

This research makes several unique contributions to the literature. To begin with, little work has been reported on the empirical evidence of using the Internet Water Armies. There are a few challenges that exist in conducting research on the Internet Water Armies. First, there is no ground truth[9] about the Internet Water Armies and

---

[9] Ground truth refers to the true information source for the posts on social media. There is no ground truth about water armies means that nobody besides the water army companies knows which post is generated by the water army.

therefore, it is hard to construct the dataset about the use of the Internet Water Armies. Second, the Internet Water Armies does not only mean purchased posts but also includes purchased accounts and other services. Representing the Internet Water Armies only using the purchased posts is not adequate. In our paper, we take advantage of the collaboration with the data analytics company in the entertainment industry and differentiate the impact between purchased posts and purchased fans. Furthermore, our study accounts for the impact of online media press simultaneously with traditional media and social media. Prior studies mostly focus on the interplay between traditional media and social media. However, as "online" becomes ubiquitous in daily life, online media press quickly occupies the life of the audience. This study contributes to the literature by examining the impact of online media press together with traditional media channels and social media platforms.

This work also has salient practical implications. As the Internet Water Armies are well used in the industry, our work provides empirical evidence showing that injecting Internet Water Armies does help with product sales. Our study identifies the time dynamics of injecting Water Army. In general, a firm can expect biggest impact on box office revenues when it purchases water army before the movie goes on-show. Positive effects can also be expected when the firm purchases the water army on the early days after the movie release, thought effect size is smaller than when the water army is used before movie release. However, water army will not be able to boost movie sales when it is used after the third week of movie release. The different temporal effect of injecting Internet Water Army presents a great guidance for the social media platform.

Reducing the effect of injecting Internet Water Army costs resource. Our study provides guidance for the platform on when to effectively utilize the resource to avoid the impact of the Internet Water Army. As mentioned above, using the Water Army is not free of drawbacks. Once the public realized that the buzz is generated by the Water Army, they will keep questioning about the subsequent movies/TV plays. Furthermore, Internet Water Army can reduce the social welfare since after the public realized the social media platform is largely occupied by the Internet Water Army, they will trim the time spent on the social media platform. Therefore, to protect the audiences and improve social welfare, the platform should pay more attentions before the movie gets released and less attentions to monitor the Internet Water Army after three weeks. Lastly, our work provides insights into the influence of future AI-generated content.

The remainder of the paper is structured as follows. In the next section, we will review the related prior literature and propose our theoretical model. Next, we will describe our data and follow with the development of our identification strategy. After the method section, we will deliberate our results followed by the robustness checks. Last, we will present the concluding remarks.

2. Related Literature

Our study draws upon and contributes to two literature streams: (1) Internet Water Armies, (2) the impact of different types of media in the motion picture industry. We review the relevant literature to understand the current state of media studies. The key issue addressed in the related literature deals with the influence of consumers' opinions and media activities on product outcomes, especially motion picture sales. A review of

Internet Water Armies research can help us identify the gaps in the literature and interesting directions of research to pursue that can yield academic contributions.

2.1 Internet Water Armies

The existing literature mainly consists of studies that seek to identify Water Armies using different techniques. Some scholars argued that the classification methods based on statistical identifiers such as reply ratio, average interval posing time, active days, and active reports might not be accurate. Instead, the inclusion of semantic identifiers like the similarity between posts greatly improves the detection rate because paid posters tend to spam similar messages with slightly changed content (Chen et al., 2013). Later studies investigated using extreme distributions and anomalies in various features to detect the Water Armies can improve detection efficiency. For instance, Water Armies spammers tend to produce short but intense bursts of topics that could be identified with the reply- and registration-related indices (Lau et al., 2011; Xu et al., 2014). High retweet ratio is another example of detection criterion as spammers tend to set up many fraudulent accounts to continuously retweet the same content to make it into the trending list, thus reaching more audiences (Yu et al., 2015). In the same direction, Dai and Wang (2018) developed the Extreme rank anomalous collection (ERAC) model using three types of characteristics: relationship, behavior, and network, to inform their unsupervised machine learning algorithm to detect Water Armies's activities. Most recently, Lian et al. (2019) applied the supernetwork theory to train the classifier algorithm using the characteristics of supernetwork layers such as relationships subnetwork (e.g. connection (friends) clustering, posting time, follower/following ratio), information subnetwork (e.g.

dissemination breadth and depth), psychological subnetwork (e.g. psychological type, transformation intensity) and negative keyword subnetwork (e.g. proportion of negative keywords, content similarity).

Few studies have focused on the impacts of the Water Armies on economic or business outcomes. Prior research hypothesizes that strong competition in the market motivates companies to employ spammers to help them get into the trending lists and thereby gain visibility (Yu et al., 2015). Additionally, Yu et al. (2015) stated that Water Armies spammers use one of three tactics: (1) promote a specific product, company, or person; (2) smear/slander competitors; and (3) delete negative posts or comments." There are abundant opportunities for studies to investigate the effects of the Water Armies, as the Water Armies are different from fake reviews or fake news. Unlike fake reviews or fake news, the Water Armies can implant anything fake to the platform. Water Armies can help inject fake clicks, fake fans, fake posts, and fake reviews. In addition, understanding the actual impact of using the Internet Water Armies can provide a starting point for companies who want to implement the strategy.

2.2 Impact of Media in the Motion Picture Industry

In the motion picture industry, traditional media channels, online media press, and social media are extensively used to promote the movie. Although online, social media is more interactive and accessible, traditional channels are still important in driving movie box office sales. Traditional media exposure moderates the effect of online WOM on movie performance in a way that more exposure on traditional media leads to more revenues for highly rated movies, and it also helps sustain movie revenues overtime after the effect of a rating diminishes (Moon et al., 2010). From a signaling perspective, massive exposure to traditional media illustrated the high advertising expenditure, which serves as one of the quality signals to attract movie-goers (Basuroy et al., 2006).

Regarding the study of social media, the volume and valence are the major focal points of prior studies in predicting a future movie's sales. In constructing a predictive model for movie sales, the inclusion of online reviews metrics, besides other metrics such as pre-release marketing, theater availability, and professional critics, significantly increases prediction accuracy because the volume of reviews drives early movie sales (Dellarocas et al., 2007). Later works have explored the relationship between social media activities and box office revenue. Dellarocas (2017) found out that social media volume, measured as the total number of posts, has a dual causality and a mutually promoting relationship with movie sales. Specifically, having more reviews increases sales, which in turn leads to even more future WOM (Duan et al., 2008a,b). The effects of using social media on different life cycle stages may vary. The pre-release buzz can reflect the early interest of the consumers and can influence the following consumers' opinions (Houston et al., 2011). In the meantime, in the later stage, audiences can update their views about the products by assimilating and combining the information. Therefore, the temporal effects of online opinions catch much attention. Xiong & Bharadwaj (2014) uncovered that including an early social media buzz evolution curve before product release improves the product revenue prediction accuracy, while Huang et al., (2017) discovered that both online words of mouth volume and valence do not play an important role in early stage, but the volume significantly influences the box office performance in later phases.

There are several under-explored areas in prior studies about media use in the motion picture industry. While online media activities are prone to manipulation and fraudulent practices performed by the Internet Water Armies, the impacts of such actions have not been empirically assessed. Manipulating social media performance may reinforce the organic social media performance or may backfire because audiences do not want to be manipulated. This study attempts to fill this gap by empirically analyzing the simultaneous impact of using the Internet Water Armies on other types of media activities and motion picture box office revenue.

3. Data

To address the research questions, we cooperate with a major entertainment data analytics company in China. As mentioned above, hiring the water army for promotion on social media platforms becomes a cultural phenomenon in China. “Influencer promotion” also contributes to this cultural phenomenon because public relations companies choose influencers based on their social media performances. The influencers with more followers, more retweet ratio, and comment ratio will have a higher chance to be selected. Consequently, influencers are more likely to flood their social media platforms. According to Mediakix influencer marketing statistics, the companies spent $8.5 billion on hiring influencers. If the influencers’ social media platform is flooded with internet water armies, the company will waste their resources. As a result, many companies, including public relations companies and movie production teams, will count on the “de-water-army” index.

Our partner company generates the “de-water-army” index for these companies. The company monitors the “de-water-army” performance for the influencers all year round and for the movie/TV shows for a relevant time window. The leading search engine in China and major entertainment authorities all use the “de-water-army” index and data from the same company.

Our dataset includes 317 movies in 2017 and 2018 from the Chinese movie market. We assembled a panel dataset at the daily level. For each movie in our data set, we observe daily box office revenue, the number of daily posts and fans on different medias, offline media (refer to traditional media), organic social media, and online media press. We also observe the number of daily purchased posts and fans. Offline media activities capture traditional media activities including newspaper articles and magazine articles. The online media activities show the online media press activities, the internet water armies’ posts, and organic social media posts. For each media channel, we gather the data at both the post level and the account level. For example, for the online social media channel, we collected the daily number of posts and the daily number of fans because certain fans can post multiple threads on one day.

Table 15 presents the variable descriptions and the summary statistics for our dataset. There are several points to be noted. First, on average, each movie is released in theaters for 44.5 days. This number includes the pre-release time and theatrical release time. It is a little bit shorter compared with American movies. This is because in the Chinese movie market, for every 30 days the movie will be shown on theatrical channels, movie distributors need to apply for a key from National Radio and Television and

Administration (NRTA). The process is long and complicated. Only movies that have huge potential in box office revenue will apply for the second movie "on show key." Most of the movies will stay in theaters for less than 30 days. The Chinese movie market is very competitive; some movies are released in theaters for less than two weeks, and then the theaters will remove it because of limited screens. Second, the degree of freedom for box office revenue is less than the degree of freedom for media performance. The reason is that the company monitors the social media performance long before it releases in theaters. The company starts to monitor the performance when they see there is a buzz happening on social media platforms. To avoid the bias generated by an unbalanced panel dataset, we limited the sample accordingly. Moreover, the degree of freedom for traditional media performance is less than others. In most of the cases, the movie distributors only utilize traditional media promotion right before and right after the movie is released in the theater because of the cost.

Table 15. Variable Description and Summary Statistics

Variable	Description	Mean	Max	Std
BO_{it}	Daily box office for movie i at day t.	2,726,518	546,763,328	15,996,967
PP_{it}	Number of purchased social media posts for movie i at day t.	5,240	3,343,549	69,383.65
PF_{it}	Number of purchased social media fans for movie i at day t.	180.6	64,728	1,346.803
SMP_{it}	Number of organic social media posts for movie i at day t.	5,684	3,399,867	68,574.04
SMF_{it}	Number of organic social media fans for movie i at day t.	998.5	321,077	5,978.399
OMP_{it}	Number of online media press posts for movie i at day t.	236.1	136,707	1,370.134
OMR_{it}	Number of online media press sites for movie i at day t.	20.14	1,006	39.679

TMP_{it}	Number of daily traditional media posts (include magazines and newspaper) for movie i at day t.	144.7	10,533	431.764
TMR_{it}	Number of daily traditional media reporters (include magazine office and newspaper office) for movie i at day t.	69.88	4,215	184.072
$Haters_{it}$	Number of organic haters (express negative attitude towards the movies) for movie i at day t.	45.12	21,290	330.789
$Likers_{it}$	Number of organic likers (express positive attitude towards the movies) for movie i at day t.	474.5	185,241	3,443.899
$Neutral_{it}$	Number of neutral fans for movie i at day t.	730.7	131,004	3,611.461

*For social media related variables (OSM, SM, haters, likers and neutral fans), they are on the unit of 1000.

4. Method and Empirical Results

In this session, we study whether water army has an impact on box office revenue. We begin with the panel vector autoregression (PVAR) model to explore the relationship between box office revenue, the number of purchased posts and fans, and the number of posts and fans on different Medias. We then split organic social media fans into haters, likers, and neutral fans, and explore how the impact would differ with different percentage of emotional fans. To deal with the potential endogeneity, we further adopt the approach of treatment effect combined with dynamic matching, exploring whether the impact was led by the purchasing of water army and when the best time to purchase water army is.

4.1 Impacts of Water Army Posts and Accounts

We first constructed a daily panel data to perform the panel vector autoregression (PVAR) model. The advantages of using PVAR is that we can assess the nature of bidirectional causality between all pairs of dependent variables as all the dependent variables, box office sales, offline media activities, and online media activities are treated jointly endogenous (Dewan and Ramprasad, 2014). Furthermore, we can identify the lag effects and understand the dynamic relationships (Duan et al., 2008). Our PVAR model is specified in Equation 1, where BO, TM, OM, PP, and OSM denote daily box office revenue, number of daily offline media posts, number of daily online media press posts, number of daily internet water army posts and number of daily social media posts, in day t (t=1,....,T). J is the order of the model. For the analysis at the fans level, the variables $LogTM_t$, $LogOM_t$, $LogPP_t$, and $LogOSM_t$ are replaced by their fan-level counterparts. Our PVAR model is specified (for each movie) as follows:

$$\begin{bmatrix} LogBO_t \\ LogTM_t \\ LogOM_t \\ LogPP_t \\ LogOSM_t \end{bmatrix} = \Sigma_{j=1}^{J} \begin{bmatrix} \pi_{11}^{t-j} & \pi_{12}^{t-j} & \pi_{13}^{t-j} & \pi_{14}^{t-j} & \pi_{15}^{t-j} \\ \pi_{21}^{t-j} & \pi_{22}^{t-j} & \pi_{23}^{t-j} & \pi_{24}^{t-j} & \pi_{25}^{t-j} \\ \pi_{31}^{t-j} & \pi_{32}^{t-j} & \pi_{33}^{t-j} & \pi_{34}^{t-j} & \pi_{35}^{t-j} \\ \pi_{41}^{t-j} & \pi_{42}^{t-j} & \pi_{43}^{t-j} & \pi_{44}^{t-j} & \pi_{45}^{t-j} \\ \pi_{51}^{t-j} & \pi_{52}^{t-j} & \pi_{53}^{t-j} & \pi_{54}^{t-j} & \pi_{55}^{t-j} \end{bmatrix} \begin{bmatrix} LogBO_{t-1} \\ LogTM_{t-1} \\ LogOM_{t-1} \\ LogPP_{t-1} \\ LogOSM_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{LogBO,t} \\ \epsilon_{LogTM,t} \\ \epsilon_{LogOM,t} \\ \epsilon_{LogPP,t} \\ \epsilon_{LogOSM,t} \end{bmatrix} \quad [1]$$

The results from our PVAR analysis are reported in Table 3 and Table 4, at the post and fan levels, respectively. We first examine the results for the regressions with movie box office revenue as the dependent variable. Looking at the coefficient estimates on purchased posts and purchased fans, we see that the results are fairly consistent at the post and fan levels: the water army has a positive relationship with the movie box office revenue, as shown by the positive and significant coefficients. Results also show that organic social media posts/fans and online media posts/fans are positively related to box office revenue, while the number of traditional media posts is negatively related to box office revenue and the coefficient of traditional media reporters is not significant, indicating the number of traditional media reporters has no discernible association with box office revenue.

Now we turn to the relationships between water army, organic social media, online media and traditional media. When water army is the dependent variable, we see fairly consistent results. The coefficients of organic social media are not significant at both post and fan levels, indicating that the number of organic social media posts/fans has no any discernible association with the purchasing behavior of water army, possibly because the positive and negative effects of emotional organic social media fans on water army balance each other out. The coefficients of online media are positive and significant suggesting that the number of online media posts/fans is positively related to the water army, while the negative and significant coefficients of traditional media indicate that the number of traditional media posts/reporters is negatively related to the water army. When organic social media, online media and traditional media are dependent variables, we see purchased water army is positively related to organic social media posts/fans, negatively related to online media posts, with no association with online media fans and traditional media posts/reporters. We could also see that online media is negatively related to traditional media, and traditional media is negatively related to organic social media both at the fans level. Organic social media fans are not related to online and traditional media fans. Other than that, the number of organic social media, online media, and traditional media posts/fans are positively correlated with each other.

Together, results in Table 16 and Table 17 both support the effectiveness of the water army. Water army, as well as organic social media and online media posts and fans, have positive impacts on box office revenue. Since it would be hard for consumers to differentiate between water army posts/fans and organic social media posts/fans, we will focus on the relationship between water army, organic social media, and box office revenue. More water army posts and accounts would lead to more organic social media posts and social media fans, while it is not necessary that more organic social media posts and fans would be positively associated with the water army. As discussed, the positive effects of likers and negative effects of emotional organic social media fans may balance each other out. Next, we will explore how the water army is associated with emotional fans and box office revenue.

Table 16. PVAR Result Table_Post Level

	logBO	logPP	logOSM	logOM	logTM
L.logBO	0.469***	0.0286*	0.0791***	0.0921***	0.0781***
	(0.0438)	(0.0138)	(0.0156)	(0.0106)	(0.0134)
L.logPP	0.190***	0.781***	0.196***	-0.0761***	-0.0284
	(0.0512)	(0.0266)	(0.0313)	(0.0197)	(0.0230)
L.logOSM	0.248***	0.0510	0.511***	0.112***	0.0625**
	(0.0475)	(0.0311)	(0.0367)	(0.0197)	(0.0222)
L.logOM	1.089***	0.134*	0.127*	0.356***	0.00993
	(0.146)	(0.0521)	(0.0531)	(0.0464)	(0.0558)
L.logTM	-0.131*	-0.0577*	0.0308	0.0724***	0.633***
	(0.0516)	(0.0233)	(0.0248)	(0.0193)	(0.0262)

Movie FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
N=7796					

Table 17. PVAR Result Table_Fan Level

	LogBO	LogPF	LogOSM	LogOM	LogTM
L.LogBO	0.483***	0.0457***	0.0886***	0.0525***	0.0738***
	(0.0425)	(0.0115)	(0.0158)	(0.00631)	(0.0110)
L.LogPF	0.320***	0.732***	0.214***	-0.0101	0.0138
	(0.0634)	(0.0245)	(0.0305)	(0.0159)	(0.0224)
L.LogOSM	0.179***	0.0327	0.539***	0.0195	0.0229
	(0.0409)	(0.0213)	(0.0261)	(0.0117)	(0.0166)
L.LogOM	1.428***	0.182**	0.392***	0.185***	-0.156*
	(0.200)	(0.0643)	(0.0792)	(0.0384)	(0.0673)
L.LogTM	0.0588	-0.110***	-0.0600*	0.0300*	0.659***
	(0.0598)	(0.0233)	(0.0289)	(0.0142)	(0.0271)
Movie FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

N=7220

4.2 Impact of Water Army Posts on Emotional Fans

From the last section, we see that the water army could directly impact movie box office revenue. Besides, the water army could impact organic social media posts and fans, and impact the box office revenue indirectly. To further explore the relationships between the water army, different users on social media and box office revenue, we identify likers, haters, and neural accounts at the account level from the original dataset and conduct the PVAR analysis. The results are reported in Table 18.

Again, we could see that the water army was positively associated with box office revenue. And after we split organic social media fans into haters, likers, and neutral fans, we found that only haters are positively related to box office revenue, while the number of water army accounts is not only positively related to the number of organic haters but also positively related to the number of organic likers and neutral fans. And the coefficients are higher on likers and haters than on neutral accounts suggesting that water army would lead to more emotional accounts. If we consider the number of water army accounts as the dependent variable, we could see that the number of likers is negatively associated with the number of water army fans, while, on the contrary, the number of haters and neutral fans is positively related to the number of water army accounts. The results indicate when there are more likers, movie firms feel less need to purchase paid accounts, while when there are more haters and neutral fans, firms would buy more paid accounts. Those purchased accounts would lead to not only more likers, haters and neutral fans, and in general more emotional fans than neutral fans.

Table 18. PVAR Result Table_Emotional Fans

	LogBO	LogPF	LogLikers	LogHaters	LogNeutral	LogOM	LogTM
L.LogBO	0.456*** (0.0441)	0.0378** (0.0118)	0.0861*** (0.0195)	0.0988*** (0.0184)	0.0848*** (0.0149)	0.0542*** (0.00656)	0.0706*** (0.0114)
L.LogPF	0.303*** (0.0668)	0.728*** (0.0250)	0.253*** (0.0378)	0.230*** (0.0362)	0.193*** (0.0286)	-0.00807 (0.0163)	0.0146 (0.0228)
L.LogLikers	0.000907 (0.0342)	-0.0571** (0.0180)	0.268*** (0.0290)	0.0309 (0.0266)	0.0532* (0.0222)	0.00828 (0.0110)	0.00167 (0.0158)
L.LogHaters	0.131*** (0.0318)	0.0545*** (0.0149)	0.0794*** (0.0223)	0.390*** (0.0229)	0.0712*** (0.0173)	-0.0141 (0.00929)	0.00394 (0.0135)
L.LogNeutral	0.0767 (0.0494)	0.0577* (0.0252)	0.263*** (0.0414)	0.241*** (0.0395)	0.414*** (0.0341)	0.0246 (0.0169)	0.0233 (0.0230)
L.LogOM	1.493*** (0.214)	0.184** (0.0676)	0.391*** (0.101)	0.151 (0.0845)	0.371*** (0.0746)	0.180*** (0.0402)	-0.143* (0.0706)
L.LogTM	0.0318 (0.0629)	-0.113*** (0.0237)	-0.0252 (0.0353)	0.0692* (0.0307)	-0.0483 (0.0266)	0.0251 (0.0145)	0.656*** (0.0277)
Movie FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

N=7015

We further consider the percentage of different social media fans instead of the numbers. We conduct the analysis using the percentage of likers, haters, and square terms as exogenous variables. The results are reported in Table 19. We could still see the positive relationships between water army and box office revenue. Besides, we could see the relationship between organic social media fans and box office revenue would be moderated by the percentage of haters and likers. The positive and significant coefficient suggests that the percentage of likers is positively related to box office revenue, and the negative coefficient of the quartic term suggests that the impact would decrease after a certain percentage. Similarly, the percentage of haters also has a positive impact and the impact would decrease after a certain number. If we only consider likers, then we could get the numbers that the impact would be highest when the percentage of likers is 43 percent, and disappears when it reaches 86 percent. And if we consider only haters, the effect is highest when the percentage is 33 percent and disappears when reaches 67 percent. Different from previous results that the number of likers had no impact on box office revenue, the percentage of likers has a positive impact. The percentage of haters also has a positive impact. The results suggest that it would be better to have a higher percentage of emotional fans; however, the percentage should not be too high.

Table 19. PVAR Result Table_Percentage of Emotional Fans

	LogBO	LogPF	LogOSM	LogOM	LogTM
L.LogBO	0.439*** (0.0464)	0.0369** (0.0123)	0.0681*** (0.0165)	0.0461*** (0.00657)	0.0693*** (0.0118)
L.LogPF	0.433*** (0.0741)	0.761*** (0.0263)	0.248*** (0.0320)	-0.00796 (0.0163)	0.0251 (0.0238)
L.LogOSM	0.156** (0.0479)	0.0431 (0.0239)	0.552*** (0.0287)	0.0255* (0.0129)	0.0203 (0.0188)
L.LogOM	1.364*** (0.207)	0.153* (0.0645)	0.313*** (0.0770)	0.172*** (0.0389)	-0.147* (0.0680)
L.LogTM	0.138* (0.0631)	-0.0820*** (0.0229)	-0.00504 (0.0278)	0.0385** (0.0139)	0.667*** (0.0270)
l.pLikers	6.526*** (1.118)	0.858* (0.437)	3.751*** (0.581)	0.838** (0.297)	0.500 (0.428)
l.pHaters	6.621*** (1.181)	0.184 (0.512)	1.234* (0.611)	0.658* (0.323)	0.165 (0.490)
l.pLikers^2	-7.601*** (1.151)	-1.651*** (0.448)	-4.524*** (0.597)	-0.677* (0.286)	-0.514 (0.419)
l.pHaters^2	-9.919*** (1.939)	-0.822 (0.691)	-1.861* (0.749)	-0.120 (0.439)	0.311 (0.629)
Movie FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

N=6982

However, it is still not clear how the relationship between water army and box office revenue would be affected by emotional organic social media fans. To address this question, we divide data into four categories, corresponding to the ratio of haters and likers. Based on this classification: likers dominated, likers weakly dominated, haters dominated, and haters weakly dominated, we conducted another analysis using the interaction terms of different groups and water army as the exogenous variables to the PVAR analysis. The results show that the relationship between water army fans and box office revenue is moderated by the percentage of emotional fans. Compared to the group of likers dominated, where the number of haters is less than 1 percent of the number of likers, the impacts of the water army in other groups are higher. In general, it is a good strategy to purchase water army when the number of haters is higher than 1 percent of likers.

Table 20. PVAR Result Table_Domination of Emotional Fans

	LogBO	LogPF	LogOSM	LogOM	LogTM
L.LogBO	0.472*** (0.0431)	0.0326** (0.0115)	0.0674*** (0.0154)	0.0508*** (0.00685)	0.0771*** (0.0113)
L.LogPF	0.107 (0.0613)	0.635*** (0.0282)	0.0197 (0.0384)	-0.0461* (0.0189)	-0.0113 (0.0265)
L.LogOSM	0.186*** (0.0413)	0.0393 (0.0213)	0.539*** (0.0260)	0.0298* (0.0122)	0.0283 (0.0169)
L.LogOM	1.252*** (0.185)	0.155* (0.0605)	0.300*** (0.0716)	0.162*** (0.0376)	-0.191** (0.0648)
L.LogTM	0.0666 (0.0576)	-0.104*** (0.0223)	-0.0415 (0.0271)	0.0308* (0.0140)	0.660*** (0.0264)
L.PFHaterdomin	0.249*** (0.0432)	0.0916*** (0.0196)	0.203*** (0.0266)	0.0565*** (0.0127)	0.0260 (0.0181)
L.PFHaterdomin weak	0.258*** (0.0388)	0.105*** (0.0171)	0.212*** (0.0242)	0.0441*** (0.0113)	0.0239 (0.0159)
L.PFLikerdomin weak	0.207*** (0.0370)	0.0947*** (0.0157)	0.210*** (0.0227)	0.0328** (0.0103)	0.0135 (0.0145)
Movie FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes

N=7316

4.3 Robustness Check: Dynamic matching and Treatment Effect

In the previous session we have found water army is positively related to movie box office revenue using PVAR. To strengthen the causal inference, we adopt an empirical approach taken from the literature on treatment effects (Woodbridge, 2002) combined with dynamic matching. Previously we examined the impact using the exact numbers of how many water army accounts were purchased. In this session, we consider whether or not a movie has purchased water army. Because a firm's behavior to purchase the water army may change in different stages, the treatment group is formed dynamically. For instance, one firm may purchase water army in week 1, but not in week 2. Movies that purchased the water army in a specific week are considered as the "treatment" group this week. To deal with the potential endogeneity, that firms hire Internet water army for movies that are doing well, or movies that are about to do well, or movies that were recently doing well, or so on, we create "proper" control groups for movies that didn't make a purchase of water army in that week by using propensity score matching (PSM). Movie characteristics, such as movie genres, counties, how many stars are used, star power index, how many stars are in different categories in terms of star power, and so on, as well as the movie's performance in the past two weeks, and the movie's water army purchasing behavior in the past two weeks are used to match the movies exhibiting similar patterns. We ensure that the control and treated groups are comparable in terms of movie features, previous performance, and previous purchasing behaviors. Then, we run the regression exploring the effect using Equation [2]. The dependent variable Y_{it} is the movie i's performance on day t. $Treat_{is}$ denotes to whether movie i purchases water armies in week s. $Week_s$ is the dummy denoting week s. α_i is the movie fixed effects, θ_s is the week fixed effect. The coefficients of the interaction terms capture the impact of purchasing water army on the movie performance in week s.

We also controlled for lagged movie performance, water army purchased, and organic social media fans. The results will not only show whether purchasing water army would lead to higher box office revenue, but also suggest when to make the purchase of the water army.

$$Y_{it} = \beta_0 + \beta_1 * Week_s + \beta_2 * Treat_{is} + \beta_3 * Week_s * Treat_{is} + \alpha_i + \theta_s + controls + \epsilon_{it} \tag{2}$$

The results are reported in Table 21. Columns 1 to 4 show the results when a firm purchases the water army from one week before the movie release to the third week after the movie release. Comparing the coefficients of the interaction term, all of them are positive and significant except the last column, the third week after the movie release. The results suggest that if the firm purchases the water army in the week before the movie release, we would see the positive impact of the water army on box office revenue. And if the firm purchases the water army during the week of the movie release, we could still see the positive impact; although smaller than the impact in the previous week. And the impact continues to decrease when the purchasing behavior happens in the second week after the movie release and disappears if the firm decides to buy the water army in the third week. It indicates that the water army does not work at this stage. The trend is also shown in Figure 4.

In general, it is effective to enter the market early and start to purchase water army from the week before the movie is on-show. The water army plays a role in increasing movie box office revenue. However, if the firm start to purchase the water army after the movie release, we could still see the effect, but the effect would decrease.

And if the firm enters the market too late such as in the third week after release, water army will not work.

Table 21. Dynamic Treatment

	One week before movie release	Movie release	Second week of movie release	Third week of movie release
week	-4.994** (-1.072	0.268 -1.768	2.213*** (0.392)	-1.391*** (0.217)
Treat*week	6.724** -1.418	4.170* -1.71	0.696* (0.294)	-0.102 (0.215)
L.logWeeklyBO	0.252* -0.102	0.177*** -0.0382	0.194*** (0.0382)	0.650*** (0.0370)
L.logWeeklyPF	-3.772** -1.16	-0.355* -0.176	0.0185 (0.0882)	-0.0621 (0.0385)
L.logWeeklyOSM	3.362** -0.892	0.333* -0.143	0.191** (0.0733)	0.0380* (0.0168)
Movie FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
N	55	1938	4669	5312

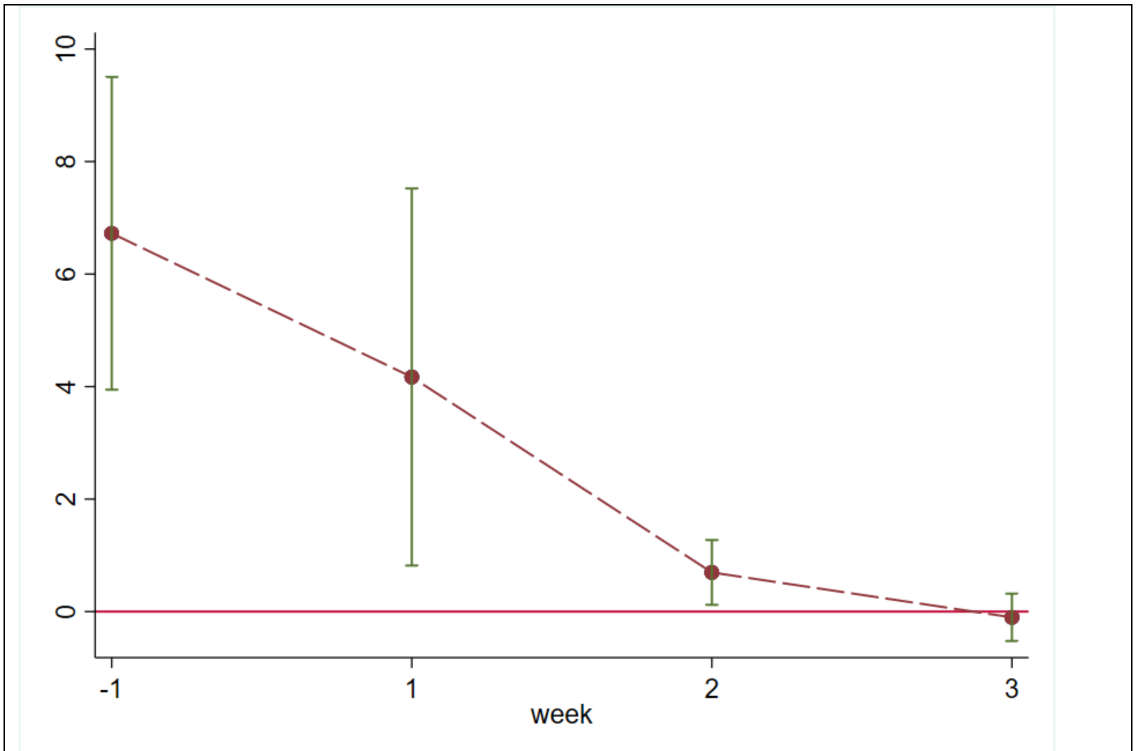


Figure 4. Dynamic Treatment

5. Concluding Remarks

In this paper, we investigate the impact of using Internet Water Army on product sales and the mechanisms and conditions of how Internet Water Army affects those sales. To address these questions, we have cooperated with a major entertainment data consulting company in China and assembled movie data in 2017 and 2018. We constructed a daily panel data to perform the PVAR. To address the identification issue, we estimate a treatment effect with dynamic matching techniques. We found that the Internet Water Army helps product sales at both post-level and fans-level. The Internet Water Army impacts product sales directly and indirectly by changing the number of emotional fans Social media platforms can also utilize our findings to improve social welfare by reducing the impact of the Internet Water Army. Since injecting the water army two weeks before release would spread the movie information and bring more haters, likers, and neutral fans. The platforms can spend more resources in monitoring and removing the Internet Water Army activities.

Our study yields a number of implications for both research and practice. This research is among the first to examine the economic and social impact of purchasing Internet Water Army. Previous studies on Internet Water Army majorly focus on the detection mechanisms from the technical views. We were very fortunate to collaborate with the entertainment data consulting company and construct this novel dataset. This novel dataset allows us to investigate the effect of purchasing the internet water army, a well-used but under-explored marketing approach.

In addition, previous literature on media performances is largely based on the quantity of the posts, such as number of posts on Twitter, number of reviews on IMDB and number of ratio plays. Number of posts and number of accounts may have different implications and simply using the number of posts does not fully capture the details. For example, the reach and awareness generated by one account posting ten times versus ten accounts each post one time can be different even though the representations of using the number of posts are the same for both scenarios. However, more accounts can be a better indication for a wider reach. Our study differentiates the number of accounts and the number of posts, and we found the effects are different. The effect is especially different in the traditional media channel. If one traditional reporter wrote multiple articles about the same movie, the box office revenue will be negatively impacted. Moreover, our study accounts for the impact of online media press simultaneously with traditional media and social media. Prior studies focus mostly on the interplay between traditional media and social media. However, as "online" becomes ubiquitous in daily life, online media press quickly occupies the life of the audience. This study contributes to the literature by examining the impact of online media press together with traditional media channels and social media platforms.

This study does have some limitations. First, we have data on only one type of traditional media--printed media--including magazines and newspapers. It may be interesting for future research to account for other traditional media channels such as TV shows. Many movie production teams will join TV shows as guests to promote their movies. Combining water armies and TV shows could be another effective way.

However, the underlying mechanisms and buying conditions could still stay the same. Next, in our study, we focus only on the impact of the volume of the water armies, not the valence. However, as the company told us most companies purchased water armies expressing positive opinions, that there is not much variance in the valence, specifically for movie contents. Future studies can take a step further and explore the content design of water army posts. Last, in our study we look only at the impact of purchasing posts and purchasing fans starting from later 2018, buying influencers' as the water armies becomes crucial. But since our data is largely from 2017 and early 2018, the influencer water armies are not a major concern in our study. Overcoming these limitations can provide blooming directions for future research. Overall, this work sheds new light on the impact of purchasing the internet water army on different media performance as well as the financial outcomes of the product. Our results provide valuable suggestions for both academic studies design and social media marketing promotion in practice.

CHAPTER 4

DO FIRMS PREFER LIKERS OR DOUBTERS?

Abstract

Abstracting from an example drawn from the movie industry, we build a game model to study the movie distributor's trade-off between honestly promote the movie according to their (subjective or objective) evaluation and catering the consumer's prior belief on the movie quality to stay on the market as long as possible. We provide insights on the optimum usage of promotion on social media and demonstrate how conventional wisdom about negative reviews will hurt business may be misleading in the presence of social media. In particular, we show that, contrary to conventional wisdom, only allowing positive reviews does not always yield the optimal strategy. We find that the information asymmetry about the movie quality could give movie distributor who wants to stay in the market in the long-run incentives to keep both positive and negative reviews on social media during the promotion. Moreover, in the extension, we find that mix-rating persistently exists even if the market is competitive in the sense that another can enter the market to compete with the monopoly distributor.

Keywords: Social Media, Information Asymmetry, Game Model

1 Introduction

Social media has become increasingly popular as an instrument for the promotion of products/services in many industries (Duan et al. (2008), Miller and Tucker (2013)). Abundant evidence shows that promotions of products using social media are effective in driving product sales (Miller and Tucker (2013), Luo et al. (2013), Chen et al. (2015)). For instance, Dell achieves three million revenue with Twitter-related sales. 10 See https://bits.blogs.nytimes.com/2009/06/12/dell-has-earned-3-million-from-twitter/ . For experienced goods such as movies, adopting social media is popular among firms in promoting their products. Consumers may seek signals of quality in advance of watching and social media platforms provide the best way for the audience to gather information. Bird box, a movie produced by Netflix, has caught attention by massive social media fans because they feel like nobody in their life talks about this movie but everyone on the Internet discusses it.[11] See https://www.metrotimes.com/the-scene/archives/ 2018/12/28/did-netflix-use-fake-twitter-accounts-to-make-bird-box-memes-is-anything-on-the-internet-real .The popularity of adopting social media raises consumers' concern about the quality of information delivered through the promotion. Whether the true quality of movie is reflected to market is also valued in addition to the movie quality per se. Therefore, the firm concerns the accuracy of promotion as well as creating awareness on social media.

[10] See https://bits.blogs.nytimes.com/2009/06/12/dellhasearned3millionfromtwitter/ .
[11] See https://www.metrotimes.com/thescene/archives/2018/12/28/didnetflixusefake-twitteraccountstomakebirdboxmemesis anythingontheinternetreal .

Companies use many ways to influence social media promotions such as giving away free products (Zhu and Furr (2016)) and hiring experts or opinion leaders ((1999)), Plucker et al. (2009)) to write reviews. Those mechanisms will lead to Holbrook a situation that both "Likers" and "Doubters" will appear on social media, which brings both positive reviews and negative reviews at the same time. Prior research presents mixed results about having negative reviews about your products. Berger et al. (2010) shows that negative reviews showing on social media can boost the sales. While Basuroy et al. (2003) point out that negative reviews hurt performance more than positive reviews help performance. The actual effect of having doubters remains unanswered. This phenomenon is intensified in social media promotion. In this paper, we aim to study the following research question: How would the movie distributor use social media to promote products? Specifically, how the movie distributor would use social media to promote the movies? What is the rationale for distributor to underrate the movie quality? Can information technology improve market efficiency?

To answer these questions, this paper focuses on the movie industry and develops a model to study the movie distributor's optimal promotion strategy for information goods while using social media. We start from a simple assumption: A distributor wants to be believed as a distributor who can precisely promote a movie according to its true quality. The quality of the movie is difficult to observe directly by consumers before watching it. However, consumers may form their belief about the movie quality based on observation of the comments or reviews on social media and decide whether to watch the movie. Knowing that consumers refer to social media to make their consumption decision, the distributors may have an incentive to shape these comments in whatever way, which will be most likely to improve their reputations. Thus, it increases their long-run profits by keeping consumers to watch their movies in the future.

Our most interesting result indicates that the distributor tends to not follow their observed signal on movie quality if the consumer's requisition on the quality is not extreme. In other words, it would allow both "good reviews" (or thumbs up) and "bad reviews" (or thumbs down) exist on the social media to confirm consumers' priors beliefs on movie quality. Consider the case where a noisy signal on movie quality is more likely to induce promotions that contradict the true quality. A consumer who has a strong prior belief on the true quality of the movie will expect inaccurate information (from advertising) to contradict that belief more often than an accurate signal. For example, Suppose Walt Disney promotes that Marvel's Avengers: Endgame is a movie that is of better quality than Star Wars: Episode IV –A New Hope. However, consumers believe this to be highly unlikely a priori, and they will infer that Walt Disney probably has exercised poor judgment and/or is overrating the movie, which might hurt Walt Disney's reputation. Therefore, Walt Disney will be reluctant to promote in line with the consumer's prior if the quality cannot be perfectly revealed to the market. The more the prior favor a given quality, the less likely the firm will promote contradicting that prior. Unlike the previous research related to reputation, which argues that the reputation effect gives marker players incentives to do good for others; however, the reputation effects in our study give players incentives to do the opposite.

Besides, we find that if consumers' requisition on movie quality is extremely high or extremely low, the distributor will promote according to their observed signal. Take Marvel's Avenger as an example. The professional evaluation team of Walt Disney indicates that the movie quality is low. Meanwhile, the consumer survey indicates that Marvel fans will not be the main audience, only low quality movie trackers and some middle-aged house husbands who want to kill time will watch the movie. Then, on social media, Walt Disney would increase bad reviews by deleting good comments or hire water army to thumbs down to match the observed signal. It will confirm with belief of these potential audience.

2 Related Literature

Our study draws from and contributes to the studies dealing with (1) online word of mouth research in Information System discipline, (2) fake reviews and (3) motion picture studies.

Online word of mouth in IS First, our work relates to the stream of social media research. Numerous studies have investigated that consumers' online opinions such as social media posts and reviews significantly influence product revenues, especially information goods, such as books (Chevalier and Mayzlin (2006)), TV shows (Godes and Mayzlin (2004)), music CDs (Morales-Arroyo and Pandey (2010); Abel et al. (2010); Dewan and Ramaprasad (2014)), and movies (Dellarocas et al. (2007); Duan et al. (2008), Moon et al. (2010); Chintagunta et al. (2010)). In terms of online words of mouth metrics, volume and valence are the focus of the studies. Some researchers found that the volume of online reviews drives product sales (Duan et al. (2008); Carson & Moore

2014). For example, Xiong and Bharadwaj (2014) found that using the volume of social media buzz, especially the prerelease buzz, increases the product revenue prediction accuracy. While other researchers contended that online word of mouth valence is more important than the volume (Dellarocas et al. (2007); Chintagunta et al. (2010)). In regard to valence, some study suggests that positive word of mouth is more likely to generate higher revenues (Wojnicki and Godes (2011); Berger and Milkman (2012)). Prior research presents mixed results about having negative reviews about your products. Berger et al. (2010) shows that negative reviews showing on social media can boost sales. While Basuroy et al. (2003) point out that negative reviews hurt performance more than positive reviews help performance. Our study focuses on incentives from the supply side, e.g., the distributor, to allow the negative reviews exist on social media. Our results show that such phenomena are the market equilibrium prediction of strategic interactions among rational market players.

Second, we contribute to the emerging stream of literature on fake reviews. The current state of the literature mainly consists of studies that seek to identify fake reviews using different techniques. Early studies found that classification using only statistical identifiers such as reply ratio, average interval posing time, active days and active reports might not be accurate. Instead, the inclusion of semantic identifiers like the similarity between posts greatly improves the detection rate because paid posters tend to spam similar messages with slightly changed content (Chen and Berger (2013)). Later studies examined extreme distributions and anomalies in various features to detect online fake reviews. For instance, fake review spammers have a tendency to produce short but intensive bursts of topics that could be identified with the reply- and registration-related indices (Xu et al. (2014)). High retweet ratio is another example of detection criterion as spammers tend to set up many fraudulent accounts to continuously retweet the same content to make it into the trending list thus reach more audiences (Yu et al. (2015)). In the same direction,Dai and Wang (2018) developed the Extreme rank anomalous collection (ERAC) model using three types of characteristics: relationship, behavior, and network, to inform their unsupervised machine learning algorithm to detect water army's activities. Most recently, Lian et al. (2019) applied supernetwork theory to train the classifier algorithm using the characteristics of supernetwork layers such as relationships subnetwork (e.g. connection (friends) clustering, posting time, follower/following ratio), information subnetwork (e.g. dissemination breadth and depth), psychological subnetwork (e.g. psychological type, transformation intensity) and negative keyword subnetwork (e.g. proportion of negative keywords, content similarity). However, all the above research is all from the angle of the technical view. They completely neglect the intrinsic mechanism which drives the usage of these technics. Our study deviates from these studies by providing a theory from the view of economics to demonstrate why the entities have incentives to adopt these technologies to allow the existence of fake reviews.

Third, our study relates to motion picture literature. It has been widely studied what made a movie successful in box office sales, such as star power (Ravid (1999), Elberse (2007)), genres and MPAA ratings (Austin and Gordon (1987)), media advertisement (Faber et al. (1984), Duan et al. (2008)), and market competition from others (Ainslie (2003)). Other studies about the motion picture industry are building the connection between movie box office revenue and social network (Duan et al. (2008); Brown et al. (2012); Lee et al. (2015)). Dellarocas et al. (2007) found that including online reviews metrics significantly increases prediction accuracy. Specifically, having more reviews increases sales, which in turn leads to even more future WOM (Duan et al. (2008)). However, it is still now clear why it is often seen that the movie distributor allows both positive and negative reviews that exist on social media, given that they can clear up all the negative reviews. Our study fills this gap by providing a game-theoretical model to facilitate our understanding of the coexistence of positive and negative reviews on the distributor's social media.

Lastly, our study sheds lights in the literature of advertisement. Advertisement serves as a signal to convey the information and indicate the quality of the promoted product (Kihlstrom,1984). Advertising provide noisy signals about the brand attributes (Anand and Sharchar, 2011). Especially for the experience goods, it is well established that advertising expenditure signal quality by conveying information about a firm’s suck cost (Riordan, 1984). The advertising expenditures at the initial stages indicate the quality of the company and increase long-run marginal revenue of the products (Milgrom and Roberts, 1986; Erdem et al., 2008). Advertisement presents a costly signal for the product since making advertisements costs resources[12]. Our study demonstrates a cost-free signal

[12] https://www.webfx.com/blog/business-advice/the-cost-of-advertising-nationally-broken-down-by-medium/

of advertising. In addition, our study is among the first to propose a theoretical framework in social media advertisement.

3 Distributor's Market Reputation and Consumer's Assessments

In this section, we provide evidence supporting key assumptions of our model: (i) Movie distributors try to build a reputation for truthful promotion according to the true quality of the movie. (ii) Consumers' assessments of the true quality of movies depend on prior beliefs (about movie quality and distributor's quality).

3.1 Reputation Concern of Distributor

The distributors' desire to maintain a reputation for providing accurate promotion on their products is at the heart of our model. Both descriptive stories and previous studies provide strong evidence of this assumption. For example, on February 16th, 2017, 21th Century Fox apologizes for creating fake news sites as part of the digital marketing campaign for the new film “A Cure for Wellness." The studio spokesperson said, “... In this case, we got it wrong. The digital campaign was inappropriate on every level, especially given the trust we work to build every day with our consumers." In the meantime, Ebbers et al. (2012) found that the commercial reputation of a film producer based on past box office performance has a positive effect on the size of the investment by distributors.

3.2 The Influence of Consumer Priors on Distributor’s Promotion

A large body of anecdotal evidence suggests a connection between consumers' prior beliefs and movie distributor's promotion strategies. For example, Joker's movie is using negative reviews for marketing. "Despite having an abundance of positive Joker reviews to choose from, WB selected a pull quote from a negative review for marketing." Many movies with big marketing budgets have low ratings on IMDB pages (for example, Justice League 2017, Spectre 2015, Star Wars: Episode VII - The Force Awakens, 2015, etc.) IMDB's best movies of 2015 have ratings of 1 as best reviews.

4 Baseline Model

In this section, we develop a model to study the optimal promotion strategies of a distributor (movie distribution firm or television network) in the presence of social media. As discussed previously, we use the movie industry to motivate the key features of our model, but the model is applicable to other industries in which firms seek to attract consumers in the long run and where distributors (firms) must make decisions on how to use social media to promote the information products. For example, it could be a publisher or an online retailer selling books.

Let us consider a movie market consists of one distributor and two groups of consumers. The distributor uses social media to promote a movie whose attributes are not perfectly/all revealed to the consumers. However, the distributor might observe all the attributes of the movie before the promotion. The first group of consumers always watch movie and we call them “hardcore moviegoers". The second group is composed of consumers who choose whether to watch the movie based on distributor's promotion and the discussion with hardcore moviegoers on social media. Their decisions are based on the promotion on the market and we call them “normal consumers". We use the Figure 5 to summarize the interactions among the three players.

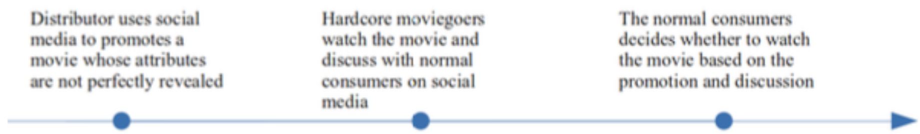


Figure 5. Interactions among market players

4.1 Distributor

At the beginning of the game, a (monopoly) distributor obtains a movie which is independently produced by a film production company or studio.[13] The “attribute" of the movie is not perfectly realized when the distributor obtains it. We use random variables $s \in h,l$ to summarize the movie's attribute in each period. Here, $s = h$ (respectively, $s = l$) indicates that the attribute is high (respectively, low).

In our model, we use “attribute" to represent an exogenous parameter that is part of the consumer's preference. In general, the attribute could represent quality, characteristic, or property of a product. In this study, we take the attribute as a scalar for simplicity, but results will not change if we extend to a multi-dimensional attribute setting.[14] The specific measurement of “attribute" might be different across industries. We use “misfit cost" which has been studied in the literature, e.g., Wattal et al. (2009), as a measure of movie attribute in this study. The movie with attractive attribute indicates a low misfit

[13] We will not address the interactions between the distributor and the studio in this paper. The profitsharing relationship between them is assumed to be determined when the distributor decides on the usage of social media promotion.
[14] We can think of the scalar s as the projection of a multidimensional movie attribute space onto a unidimensional space.

cost or fits perfectly to the normal consumer's taste, and the one with normal attribute indicates a high misfit cost or does not fit the normal consumer's taste.[15]

Additionally, we assume that there is an uncertainty on the attribute, i.e., the value of misfit cost. This means that, ex-ante, the market players cannot perfectly observe the attribute of a movie. It is motivated by the fact that the market size for a specific movie is not realized before normal consumers choose to watch. However, all the market players have an evaluation on it before any decisions. Essentially, consumers are not fully aware of their misfit cost, whereas the distributor's promotion strategy may influence consumers' ex ante perception of their misfit cost. For the convenience of expression, we will use “attribute" and “misfit cost" interchangeably in this study.

In particular, we assume that the probability of $s = h$ is $(s = h) = p > 1/2$, and the probability of $s = l$ is $(s = l) = 1 - p$. The distribution of the attribute is publicly known, i.e., p is the common prior of the attribute being high for all players. The assumption that $p > 1/2$ is without loss of generality, and we use it to capture the market (distributor and consumers) attitude (or sentiment) on the attribute. When $p > 1/2$, it means that the market is optimistic about the attribute. Our goal of this paper is to study how consumer's attitude about the attribute affects the distributor's promotion strategy through social media. Thus, we do not take a stand on whether the market is optimistic or pessimistic about the uncertainty, i.e., the attribute. We take the optimism as an example

[15] More explanation based on misfit cost will be illustrated when we introduce the preference of normal consumers.

of consumer's biased attitude against the uncertainty of the attribute. If we assume $p < 1/2$, i.e., pessimistic about the attribute, we will have a symmetric prediction.

After obtaining the movie, the distributor assesses the misfit cost with its expertise. Then based on its assessment of the misfit cost s, it strategically chooses how to promote the movie on the social media. More specifically, during the process of assessment, the distributor receives a signal $\omega \in h,l$ on whether the movie fits the taste of the representative of normal consumers. The distribution of the signal depends on the distributor's "type" $\theta \in \theta_H,\theta_L$, e.g, expertise. The type θ is private information and can only be observed by the distributor itself. With probability λ, the distributor is "honest" type, i.e., $\theta = \theta_H$, and observe the signal perfectly reveal the true attribute of the movie. With probability $1 - \lambda$, it is "strategic" type, i.e., $\theta = \theta_L$, and observes a noisy signal distributed according to

$$\Pr(\omega = h|s = h) = (\omega = l|s = l) = q \in (p,1).$$

This reads as: the probability of observing h (respectively, l) when the misfit cost is h (respectively, l) is q. Here, $q \in (p,1)$ is to assume that the signal observed by the strategic type is noisy but informative. In other words, the signal is not perfect about the value of misfit cost, i.e., $q < 1$; however, conditional on observing a signal of h (respectively, l), the value of misfit cost is more likely to be h (respectively, l), i.e., $(s = h|\omega = h) > (s = h)$ (respectively, $(s = l|\omega = l) > (s = l)$).

Based on the signal observed, the distributor chooses the social medial promotion strategy $\sigma(\omega,\theta) = (x = G|\omega,\theta) \in [0,1]$. Here, $x \in G,B$ represents the "messages" posted on social media[16]. When $x = G$, the messages are all positive, or any messages that indicate the movie will perfectly fit the normal consumer's taste; whereas $x = B$, the messages are all negative. $\sigma(\omega,\theta)$ specifies the proportion of positive messages, i.e., $x = G$, that the type θ distributor wants to keep/use on the social media platform when the observed signal is ω.

Without loss of generality, we assume that high-quality distributor always promotes according to the observed signal (the true attribute of the movie), i.e., $\sigma(\omega = h,\theta_H) = 1$ and $\sigma(\omega = l,\theta_H) = 0$. Allowing both types of distributors to promote strategically only complicates our analysis and lead our model to have multiple equilibria, but without any new insights. In fact, in this general set up, the high-quality distributors promote according to the observed signal is the unique equilibrium after applying refinement of intuitive criterion. From now on, to simplify notation, we further let $\sigma(\omega) \equiv \sigma(\omega,\theta_L) \in [0,1]$. Additionally, to catch the reality and simplify analysis, we assume the distributor is the price-taker in the market, i.e., it cannot strategically set the ticket price.

In the movie industry, $\sigma(\omega)$ could represent the proportion of "good reviews" or "thumbs up" shown on the social medial. For instance, in the distributor's YouTube

[16] The word "message" can be broadly explained. It could be promotions posted by the distributor in all formats, text, video, and audio, as long as they are informative. In addition, the symbols, such as thumbs up or thumbs down, also fit our definition of the message.

channel, we might see 9,000 thumbs up and 1,000 thumbs down, then we would explain $\sigma(\omega) = 0.9$. For simplicity, we restrict our attention to the case under which $\sigma(h) \geq \sigma(l)$. Cases where this assumption does not hold are equivalent to a relabeling of the promotion (action). Under this specification, the situation where a distributor allows both “good reviews" and “bad reviews" exist on social media corresponds to the cases of $\sigma(\omega) \in (0,1)$, where $\omega \in h,l$.

Discussion In this paragraph, we explain more about the “attribute". In reality, it could relate to many factors, including the budget, the characters of directors, writers, actors, quality of pictures and its genre. Normal consumers might observe some of them before decisions, such as the directors, actors. However, we assume that the relation between these observed factors and whether they can fit normal consumers' tastes is still uncertain ex-ante. In the meantime, normal consumers could have an evaluation of the relation. For instance, with a high chance, a high budget would generate movies with a high-quality picture. A sophisticated director usually bring a breathtaking storyline with high likelihood.

In the above specification, we implicitly assume that all the factors share the following features. First, they are exogenously given, which cannot be influenced by any players in our model. Second, normal consumers care about these features in the sense that their purchase decisions depend on these features. Third, normal consumers may not perfectly observe them before watching or discussing the movie. For instance, most movies will not publicly reveal their budget information before the release[17]. Then, it will be difficult for normal consumers to evaluate the “attribute" based on the budget. Fourth, the normal consumers may not understand the idea that the movie director wants to deliver before their in-depth discussion after watching the movie[18].In the U.S market, most major studios do not disclose the full budgets for their films, such as production, development, marketing, and advertising. For example, the production budget of Marvel's "The Avengers" is recorded as $220 million. However, after incorporating marketing and advertising costs, the budget is far more than this number.

4.2 Consumers

Each normal consumer watches the promotion of the movie and participates the discussion on social media, and then decides whether or not to watch it. Rather than specify the pricing structure in detail, furthermore, we make the following assumption: the the first group of consumers, i.e,. “hardcore moviegoers", always purchase and watch the movie[19]. This assumption allows us to focus on how normal consumers update beliefs about distributor's type and attribute based on the promotion and the discussion on social

[17] In the U.S market, most major studios do not disclose the full budgets for their films, such as production, development, marketing, and advertising. For example, the production budget of Marvel’s "The Avengers" is recorded as $220 million. However, after incorporating marketing and advertising costs, the budget is far more than this number.

[18] Quentin Tarantino is famous for possessing the ability to stir controversy, and his career has been marked by it at almost every turn

[19] It is trivial to give conditions for these consumers to watch in equilibrium. We only need to normalize the value of watching a movie with low attribute to be zero; then, it is easy to find conditions for these consumers to watch in equilibrium. Moreover, we are only interested in the equilibrium strategies in which all consumers on the market watch movies.

media and decide whether to watch movies from the distributor in the future. We assume that the normal consumers always prefer movies with low misfit cost, i.e., movies that are closer to their ideal taste.

After watching the movie, we assume that the attribute is not immediately revealed to the normal consumers. However, with probability $\rho \in [0,1]$, the normal consumers can learn the true attribute of the movie through a "discussion system" without any cost[20]. In reality, the discussion system could be explained broadly, such as bulletin board system (BBS), discussion forum, the social media or other online communities; and the normal consumers might learn the true attribute of the movie by participating the discussion using their spare time. In this specification, we use ρ to measure the information accuracy from the discussion system. The more the ρ is close 1, the more accurate the information is. We further let $A \in H,L,\emptyset$ denote the attribute received through the discussion system, with $A = \emptyset$ indicating the case of learning nothing, whereas H indicating high attribute and L indicating low attribute. We slightly abuse notation and denote by $\mu(x,A) \equiv \theta = \theta_H|x,A \in [0,1]$ the normal consumer's posterior belief that the distributor is "high-quality", θ_H, given a promotion, x, and result of discussion, A. In addition, we let $\varphi(x,A) \equiv s = h|x,A \in [0,1]$ to denote the normal consumer's posterior belief that the attribute is high. After then, the consumer decides whether to watch the movie from the distributor in the future.

[20] In the baseline model, ρ is exogeneous given. In the extension, we will let ρ be a function of discussion time and further discuss how social media engagement would affect the distributor's promotion strategy.

4.3 Preferences

The payoff of the distributor is given

$$u_d(x,d,s) = \begin{cases} R & \text{If } (x,d,s) \in \{(G,1,h),(B,1,l)\} \\ 0 & \text{If } d=0, \text{ for any } x \in \{G,B\}, s \in \{h,l\} \\ -W & \text{Otherwise} \end{cases}.$$

Here, $d \in 1{,}0$ is the indicator of whether the normal consumer watches the movie from the distributor. $d = 1$ means to watch, and $d = 0$ means not. $R > 0$ is the net profit. It comes from the box-office revenue and all the other market returns (direct and indirect) that generated by correctly promoting the attribute through social media promotion[21]. However, if the promotion does not match the true attribute, the market will punish the distributor and induce a negative payoff (penalty) $-W$.

The rationale for above specification is as follows. First, the market is always efficient when there is no information asymmetry, meaning that the distributor always maximizes its net profit by promoting according to the perfectly observed attribute. Second, our most interesting result is that the distributor chooses to allow both positive and negative reviews to exist on social media. If we assume that the distributor prefers the payoff from not matching the true attribute, then the result is not interesting because it is could completely be driven by the preference, but not the strategic interactions among

[21] In general, the net payoff in the case of $(x, d, t) = (G, 1, \text{h})$ might be different from the cases of $(x, d, t) = (B, 1, l)$. However, the absolute value of the payoffs is not the key to drive our main finding. Any positive payoffs will sustain our results. Therefore, to enhance the argument that our results are driven by information uncertainty and to simplify the notation, we let the payoffs be the same for the two cases.

all the players in the market. To avoid this, we make this specification to ensure that the distributor always reaches its best when the promotion matches the true attribute, given that the normal consumers watch it. In the meantime, we will prove, even if this benefit exists, the distributor still has an incentive not to promote according to true attribute. This result is driven by the information structure and the strategic interactions among the players in the market.

For each normal consumer, the payoff is, for $\theta \in \theta_H, \theta_L$,

$$u_c(d,s) = \begin{cases} Q & \text{If } (d,s) = (1,h) \\ Q - \Delta & \text{If } (d,s) = (1,l) \\ 0 & \text{Otherwise} \end{cases} .$$

Here, $Q > 0$ is the normal consumer's net payoff, i.e., gross consumer surplus from watching the movie. If he watches a movie with high attribute, the misfit cost is zero. If normal consumers watch a movie with low attribute, they will suffer the misfit cost $\Delta > 0$. If the normal consumers do not watch the movie, regardless of the attribute, they get the reservation value, which is normalized to zero[22].

[22] Implicitly, we assume the following specification to support this preference. Let us consider a linear city model with one distributor and a continuum of normal consumers who are distributed on an interval [0, 1]. Each point on the interval represents the evaluation of the movie based on the normal consumer's taste. We assume that the true value of the taste on the distributor's movie is located at point 0. However, the taste of the representative consumer is not realized before watching the movie. We model it as a random variable $s \in \{0, \Delta\}$, where $\Delta \in (0, 1)$. When $s = 0$, the movie perfectly fits the representative consumer's taste, then the misfit cost is zero, and we call it a high attribute movie for normal consumers. This corresponds to the case when s = h. When $\theta = \Delta$, the movie does not fit the representative's taste, then he needs to pay a misfit cost Δ to watch the movie, and we call it a low attribute movie for normal consumers. This corresponds to the case when $s = l$.

We use above specification to catch two features. First, the consumer values attribute and distributor's characteristic, i.e, type. Second, the expected payoff based on this specification will show that the distributor's promotion strategies provide informative signals about the unknown attribute of the movie, and consumers value this information because they face some decisions, e.g., whether to watch the movie by the distributor. Additionally, in the baseline model, we do not put any special behavior assumption on normal consumer's preferences, such as the behavioral pattern found in Yin et al. (2016)[23]. The above payoff structure enhances that our results are driven by the strategical interactions among the rational market players, not the preference per se. However, in the extension, we will consider a payoff structure studied in Yin et al. (2016).

4.4 Timing

The timing of the game is as follows.

1. Nature determines the attribute of the movie and the type of distributor.
2. The distributor observes a signal related to the true attribute of the movie, and then chooses how to promote on social media. In the meantime, the “hardcore moviegoers" watch the movie and post discussion on the social media regardless of the distributor's promotion.

[23] They found that confirmation bias exists in consumers’ mindset. Consumers have a tendency to perceive reviews that confirm (versus disconfirm) their initial beliefs as more helpful.

3. The normal consumers discuss with “hardcore moviegoers" on the social media. With some probability, the normal consumers learn the true attribute of the movie through the discussion;
4. The normal consumers update belief about distributor's type and the attribute based on the promotion and the discussion. Then they decide whether to watch movies from the distributor.
5. At the end of the game, payoffs are realized.

5 Main Results

We use the movie industry to illustrate our main equilibrium predictions:

1. If, with a high probability, the movie quality will be revealed to normal consumers in the social media discussion before their purchase decision, then the distributor would promote according to their observed signal, i.e., $\sigma(h) = 1$ and $\sigma(l) = 0$.
2. If the probability of getting the true quality of a movie through social media discussion is not high, then the distributor still might promotes according to their observed signal, i.e., $\sigma(h) = 1$ and $\sigma(l) = 0$. It will be true if the consumer's requisition on distributor being “high-quality" is extremely high or extremely low.
3. However, if the probability of getting the true quality of a movie through social media discussion is not high and the consumer's requisition on distributor's being `high-quality" is moderate, then the distributor's promotion

strategy will not follow the observed signal. It will move towards consumers' prior by allowing both “good reviews" and “bad reviews" exist on the social media when an observed signal indicating the movie quality is low, i.e., $\sigma(h) = 1$ and $\sigma(l) \in (0,1]$.

As the first step to analyze the prediction of the game, we study how a normal consumer's posterior belief about distributor's type depends on the normal consumer's prior on the movie quality, p, the distributor's promotion strategy, $\sigma(\omega)$, and the expertise of the distributor, q.

Lemma 1. *The posterior* $\mu(G,\emptyset)$ *is strictly increasing with the likelihood ratio:* $\frac{(G|\theta_H)}{(G|\theta_L)}$.

Proof. All the proofs are in the Appendix.

This result induces that the more the normal consumers' prior beliefs favor the high quality, $s = h$, the higher they will value the quality of a distributor that reports G. In other words, when normal consumers believe that the movie quality is high, then a distributor who promotes it as a good movie will attract these consumers to watch.

The intuition is as follows. As p, the probability that the movie is high-quality, increases, the probability that an honest distributor promotes G will increase faster than the probability that a strategic distributor promotes G, because the strategic distributor's promotion is less likely related to the true quality. Then the promotion G is a better indicator of the movie quality and the type of the distributor. Furthermore, the strategic distributor might have incentives to allow more positive reviews to exist on social media to persuade the normal consumer that the distributor is honest and the promoted movie is high quality.

Apply a similar analysis to the normal consumer's belief after promotion B, we have the following results.

Proposition 1. *Suppose the normal distributor chooses G and B with positive probability, then the posterior* $\mu(G,\emptyset)$ *is strictly increasing in* p*, strictly decreasing in* $\sigma(h)$ *and* $\sigma(l)$*, and decreasing in* q*;* $\mu(B,\emptyset)$ *is strictly decreasing in* p*, strictly increasing in* $\sigma(h)$ *and* $\sigma(l)$*, and increasing in q.*

Formally, we have the following proposition to support this prediction.

Proposition 2. *Given* $\rho > \rho^*$*, there is an equilibrium under which* $\sigma(h) = 1$ *and* $\sigma(l) = 0$*. Given* $\rho < \rho^*$*, if* $\mu(\theta_H|B,\emptyset) < \mu^* < \mu(\theta_H|G,\emptyset)$*, there is an equilibrium under which* $\sigma(h) = 1$ *and* $\sigma(l) \in (0,1]$*; if* $\mu^* < \mu(\theta_H|B,\emptyset)$ *or* $\mu^* > \mu(\theta_H|G,\emptyset)$*, there is an equilibrium with* $\sigma(h) = 1$ *and* $\sigma(l) = 0$*. Here* ρ^* *is the cutoff value for the probability of uncertainty resolution; and* μ^* *is the cut-off value of consumer's purchase decision.*

This Proposition demonstrates that there are two crucial conditions to induce the distributor to allow both “Doubter" and “Liker" exist on social media in equilibrium: first, consumer's cut-off value μ^*, second, the probability that the normal consumers could learn the movie quality through the discussion on the social media, ρ. If the quality is revealed through the discussion and the distributor reports B, i.e., buy “Doubter", the consumer's posterior on facing a high-quality distributor turns to be $\mu(\theta_H|B,\emptyset)$. If it is less than the cut-off value, the normal consumer would not watch. That is because the consumer prefers a high-quality distributor which always generates high expected payoff. However, if the belief is bounded above, say, $\mu^* < \mu(\theta_H|G,\emptyset)$, then the distributor still has a chance to get the normal consumer as long as it reports G. This potential benefit gives a strategic distributor, which thinks the quality of the movie is more likely to be high, an incentive to deviate from truthful reporting B. Promote as G would increase its chance of staying on th e market if the uncertainty is not resolved.

Table 22. Summary of Results

$\rho > \rho^*$	$\rho < \rho^*$		
	$\mu^* < \mu(\theta_H\|B,\emptyset)$	$\mu(\theta_H\|B,\emptyset) < \mu^* < \mu(\theta_H\|G,\emptyset)$	$\mu^* > \mu(\theta_H\|H,\emptyset)$
Honest promote	Honest promote	Strategic distributor chooses both “Doubter" and “Liker" when observing L	Honest promote

Nevertheless, choosing both “Doubter" and “Liker" has two costs. First, the distributor always receives positive payoff directly from truthfully reporting the observed signal. Therefore, if, after imperfectly observing the signal, the strategic distributor believes that the chance of facing a low-quality movie is large, then choosing both “Doubter" and “Liker" might be too costly for the strategic distributor. The incentive to mix would be reduced. Second, if the quality is revealed through the social media discussion, the distributor would stay on the market if and only if it truthfully promotes what has been observed. This cost of choosing both “Doubter" and “Liker" is not affected by ρ, the probability that the movie quality is revealed through social media discussion.

The second cost is the loss of reputation return which would depend on ρ. Given ρ is high, if the distributor chooses to mix, it is likely to be caught and fail to get reputation return from future. So promoting according to observed signal is sustained. However, if ρ is low, the distributor is unlikely to be caught if it mixed. So promoting according to the observed signal cannot be sustained in equilibrium.

Next, let us discuss the cutoff value ρ^*. This value is the upper bound below which the normal consumers watch the movie from the distributor. It depends on the posterior belief when it observes l, on the probability q that a strategic distributor observes the true state. It is easy to check it is increasing with ψ and q. The idea is as follows. If the strategic distributor believes that the probability of the movie quality being h increases, then the strategic distributor would have more incentives to promote G after observing l. Therefore, to reduce the incentive of choosing both “Doubter" and “Liker", the probability of revealing the quality through social media discussion needs to increase. Similarly, when q increases, the incentive of choosing both “Doubter" and “Liker" also increases.

6. Conclusion

In this paper, we present a new model to understand how the movie distributor's promotion strategies on social media would be affected by the consumer's prior beliefs. Interestingly, we find that when uncertainty, about movie quality and distributor's quality, is significant, mix-rating on social media persistently exists in equilibrium, even when the competition is introduced. This result does not arise from the movie distributor's platforms' own preference, but from a rational distributor's desire to stay in the market for long-term returns. Our model sheds light on policy implications for the regulation of the motion picture industry because it identifies conditions that may improve market efficiency.

Nevertheless, our model still has limitations and leaves several questions open. First, our model assumes consumers can only decide whether or not to watch movies and have no bargaining power to influence the distributors' decisions in other ways. While in reality, normal consumers can form a coalition online and influence the distributor's promotions in other ways. The bargaining power on the consumer side might affect the distributor's incentive to promote according to the observed information.

Second, our main results rely on the assumption that the movie distributors care about their future reputation return from the market. In our model, we endogenize this market return in a reduced-form so that we can use the most parsimonious model to highlight the channel of demand-side mix-rating. Future research could examine a more complex market structure that could determine the reputation return and how it would affect distributors' decisions.

CHAPTER 5

OVERALL CONCLUSION

This dissertation presents three studies regarding digital media analytics, specifically, social media marketing from two perspectives: effective content design and optimal promotion strategy using the Internet Water Army. Together, the three essays have numerous contributions for both research and practice in digital media analytics. Our studies provide a working template for future efforts by organizations or researchers attempting to formulate a multimedia strategy and offer a guideline for organizations that want to use the Internet Water Army to promote.

The first essay proposes a systematic, theory-driven approach for assessing and guiding the creation of multimedia content strategy and then contextualizes within a study to examine the funding success of Kickstarter campaigns. By drawing upon theories from Aristotelian reasoning and utilizing Kickstarter campaign data that incorporates many distinct data formats including videos, images, text descriptions, and associated project creators' information, our proposed method lays a foundation for designing an effective multimedia strategy with means to evaluate a vast range of data types. We found that both video aesthetics and image aesthetics are crucial to an effective content design by directly influencing the project funding outcomes and indirectly increasing consumer engagement. Stronger emotions such as including emotion joy and sadness can benefit the funding outcomes. For multimedia content design, having content reinforcement and emotion sad reinforcement will increase the total money pledged.

This study carries several implications for both research and practice. This research is among the first to attempt theory-driven consolidation of distinct data sources spanning unstructured text, image, and video data. Furthermore, this study contributes to crowdfunding literature. Our study is among the first to evaluate the impact of video content on crowdfunding project success. This study introduces numerous recent developments from computer science to crowdfunding literature and the overall information systems discipline (e.g., NIMA, Scale-invariant feature transform, and Random Sample Consensus). This study also details several practical implications. The approach undertaken by this research can be used as a working template for future efforts by organizations or researchers attempting to formulate a multimedia strategy. Integration of various AI services and using them to analyze existing datasets can yield valuable and personalized insights to users. We instantiate this concept with an example of a multimedia strategy for guiding Kickstarter campaign creators. The pipeline developed in this research can readily be applied to other contexts where multimedia strategy is required.

The second essay examines the actual impact of using the Internet Water Army. We found that the Internet Water Armies helps product sales at both post-level and fans-level. The Internet Water Armies impacts product sales both directly and indirectly by changing the number of emotional fans. For purchasing strategies, hiring the Water Armies one week before release would spread the movie information and bring more haters, likers, and neutral fans. However, the Water Armies will not work if the purchase occurs in the third week after the movie release.

The second study yields a number of implications for both research and practice. This research is among the first to examine the economic and social impact of purchasing Internet Water Armies. Previous studies on Internet Water Armies majorly focus on the detection mechanisms from the technical views. By constructing a novel dataset, we are able to investigate the effect of purchasing the internet Water Armies, a well-used but under-explored marketing approach. In addition, previous literature on media performances is largely based on the number of posts, such as the number of posts on Twitter, the number of reviews on IMDB, and the number of radio plays. The number of posts and the number of accounts may have different implications and simply using the number of posts does not fully capture the details.

Essay 3 presents a new model to understand how the movie distributor's promotion strategies on social media would be affected by the consumer's prior beliefs. Interestingly, we find that when uncertainty, about movie quality and distributor's quality, is significant, mix-rating on social media persistently exists in equilibrium, even when the competition is introduced. This result does not arise from the movie distributor's platforms' preference, but from a rational distributor's desire to stay in the market for long-term returns.

Our model sheds light on policy implications for the regulation of the motion picture industry because it identifies conditions that may improve market efficiency. First of all, this study focuses on incentives from the supply side, e.g., the distributor, to allow the negative reviews exist on social media. Our results show that such phenomena are the market equilibrium prediction of strategic interactions among rational market players. In addition, our study deviates from prior studies that mostly studying the fake contents from the technical view by providing a theory from the view of economics to demonstrate why the entities have incentives to adopt these technologies to allow the existence of fake reviews. Moreover, it is still now clear why it is often seen that the movie distributor allows both positive and negative reviews that exist on social media, given that they can clear up all the negative reviews. This study fills this gap by providing a game-theoretical model to facilitate our understanding of the coexistence of positive and negative reviews on the distributor's social media.

REFERENCES

Abel, F., DiazAviles, E., Henze, N., Krause, D., and Siehndel, P. (2010). Analyzing the blogosphere for predicting the success of music and movie products. In *2010 International Conference on Advances in Social Networks Analysis and Mining*, pages 276–280. IEEE.

Agrawal A, Gans J, Goldfarb A (2018) Exploring the Impact of Artificial Intelligence: Prediction versus Judgment. NBER Working Paper No. 24626, *National Bureau of Economic Research*, Cambridge, MA.

Agrawal, A., Catalini, C., & Goldfarb, A. (2015). Crowdfunding: Geography, social networks, and the timing of investment decisions. *Journal of Economics & Management Strategy*, 24(2), 253-274.

Ahlfeldt, G. M., Moeller, K., & Wendland, N. (2015). Chicken or egg? The PVAR econometrics of transportation. *Journal of Economic Geography*, 15(6), 1169-1193.

Ainslie, A., Dreze, X., & Zufryden, F. (2003). Modeling Movie Lifecycles and Market Share. *Marketing Science,* 24, 508–517.

Ainslie, G. (2003). Uncertainty as wealth. *Behavioural Processes*, 64(3):369–385.

Austin, B. A. and Gordon, T. F. (1987). Movie genres: Toward a conceptualized model and standardized defini

Akpinar, Ezgi, and Jonah Berger. “Valuable Virality.” *Journal of Marketing Research*, vol. 54, no. 2, 2017, pp. 318–330., doi:10.1509/jmr.13.0350.

Albuquerque, P., Pavlidis, P., Chatow, U., Chen, K.-Y., & Jamal, Z. (2012). Evaluating Promotional Activities in an Online Two-Sided Market of User-Generated Content. Marketing Science, 31(3), 406–432.

Anand, B. N., & Shachar, R. (2011). Advertising, the matchmaker. *The RAND Journal of Economics*, 42(2), 205-245.

Anderson, M., and Magruder, J. 2012. “Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database* Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database,” *Economic Journal* (122:563), pp. 957–989. (https://doi.org/10.1111/j.1468-0297.2012.02512.x).

Anmarkrud, Ø., Andresen, A. and Bråten, I., 2019. Cognitive load and working memory in multimedia learning: Conceptual and measurement issues. *Educational Psychologist*, 54(2), pp.61-83.

Aristotle (1991) On Rhetoric. Kennedy GA, trans. (Oxford University Press, New York).

Arnott, D., Lizama, F., & Song, Y. (2017). Patterns of business intelligence systems use in organizations. *Decision Support Systems*, *97*, 58–68. doi: 10.1016/j.dss.2017.03.005

Austin, B., & Gordon, T. (1987). Movie Genres: Toward a Conceptualized Model and Standardized Definition. In B. Austin (Ed.), Current Research in Film: Audience, *Economics, and Law*, Vol. 3. Norwood, NJ: Ablex Publishing Co.

Barbado, R., Araque, O., and Iglesias, C. A. 2019. "A Framework for Fake Review Detection in Online Consumer Electronics Retailers," *Information Processing & Management* (56:4), pp. 1234–1244. (https://doi.org/10.1016/j.ipm.2019.03.002).

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of marketing*, 67(4), 103-117.

Basuroy, S., Desai, K. K., and Talukdar, D. 2006. "An Empirical Investigation of Signaling in the Motion Picture Industry," *Journal of Marketing Research (JMR)* (43:2), pp. 287–295. (https://doi.org/10.1509/jmkr.43.2.287).

Berger, J. and Milkman, K. L. (2012). What makes online content viral? Journal of marketing research, 49(2):192– 205.

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive effects of negative publicity: When negative reviews increase sales. Marketing Science, 29(5), 815-827.

Binder, J., Howes, A., & Sutcliffe, A. (2009, April). The problem of conflicting social spheres: effects of network structure on experienced tension in social network sites. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 965-974).

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

Blok, J., Dijkhuis, T., & Dol, A. (2017). Toward a generic personalized virtual coach for self-management: a proposal for an architecture. In *9th International Conference on eHealth, Telemedicine, and Social Medicine 2017*.

Bostrom, R. N., Humphreys, R. J., & Roloff, M. E. (1981). Communication and helping behavior: The effects of information, reinforcement, and sex on helping responses. *Communication Quarterly*, *29*(3), 147–155. doi: 10.1080/01463378109369400

Branco, F., Sun, M., & Villas-Boas, J. M. (2016). Too Much Information? Information Provision and Search Costs. *Marketing Science*, *35*(4), 605–618. doi: 10.1287/mksc.2015.0959

Brown, A. L., Camerer, C. F., and Lovallo, D. 2013. "Estimating Structural Models of Equilibrium and Cognitive Hierarchy Thinking in the Field: The Case of Withheld Movie Critic Reviews," *Management Science* (59:3), pp. 733–747. (https://doi.org/10.1287/mnsc.1120.1563).

Brown, A. L., Camerer, C. F., & Lovallo, D. (2012). To review or not to review? Limited strategic thinking at the movie box office. *American Economic Journal: Microeconomics*, 4(2), 1-26.

Burtch, G., Ghose, A., & Wattal, S. (2013). An Empirical Examination of the Antecedents and Consequences of Contribution Patterns in Crowd-Funded Markets. *Information Systems Research*, *24*(3), 499–519. doi: 10.1287/isre.1120.0468

Burtch, G., Ghose, A., & Wattal, S. (2015). The hidden cost of accommodating crowdfunder privacy preferences: A randomized field experiment. *Management Science*, 61(5), 949-962.

Burton GO (2007) Silva rhetoricae.Accessed August 26,2019,http:// rhetoric.byu.edu/.

Butler, D. (2013). When Google got flu wrong: US outbreak foxes a leading web-based method for tracking seasonal flu. *Nature*, 494(7436), 155-157.

Cacioppo, J. T., & Petty, R. E. (1979). Effects of message repetition and position on cognitive response, recall, and persuasion. *Journal of Personality and Social Psychology*, *37*(1), 97–109. doi: 10.1037//0022-3514.37.1.97

Chen, C., Wu, K., Srinivasan, V., and Zhang, X. 2013. "Battling the Internet Water Army: Detection of Hidden Paid Posters," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13*, Niagara, Ontario, Canada: ACM Press, pp. 116–120. (https://doi.org/10.1145/2492517.2492637).

Chen, DQ., Mocker, M., Preston, DS. (2010). Information Systems Strategy: Reconceptualization, Measurement, and Implications. *MIS Quarterly,* 34(2), 233-259.

Chen, H., De, P., & Hu, Y. J. (2015). IT-enabled broadcasting in social media: An empirical study of artists' activities and music sales. *Information Systems Research*, 26(3), 513-531.

Chen, Y., Liu, Y., and Zhang, J. 2012. "When Do Third-Party Product Reviews Affect Firm Value and What Can Firms Do?? The Case of Media Critics and Professional Movie Reviews," *Journal of Marketing* (76:2), pp. 116–134.

Chen, Z., & Berger, J. (2013). When, why, and how controversy causes conversation. *Journal of Consumer Research*, 40(3), 580-593.

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

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing science*, 29(5):944–957.

Cornelius, P. B., & Gokpinar, B. (2019). The role of customer investor involvement in crowdfunding success. *Management Science.*

Dafoe, A. (2018). AI governance: A research agenda. Governance of AI Program, *Future of Humanity Institute, University of Oxford: Oxford, UK*.

Dai, H., Zhu, F., Lim, E.-P., and Pang, H. 2015. "Detecting Anomaly Collections Using Extreme Feature Ranks," *Data Mining and Knowledge Discovery* (29:3), pp. 689–731. (https://doi.org/10.1007/s10618-014-0360-3).

Dai, J. and Wang, X. (2018). An identification model of water army based on data analysis. In *Journal of Physics: Conference Series*, volume 1069, page 012085. IOP Publishing.

Dai, J., and Wang, X. 2018. "An Identification Model of Water Army Based on Data Analysis," *Journal of Physics: Conference Series* (1069), p. 012085. (https://doi.org/10.1088/1742-6596/1069/1/012085).

Darke, P. R., and R. J. Ritchie. 2007. "The Defensive Consumer: Advertising Deception, Defensive Processing, and Distrust." *Journal of Marketing Research* 44 (1): 114–127.

Del Vicario, M., Quattrociocchi, W., Scala, A., and Zollo, F. 2019. "Polarization and Fake News: Early Warning of Potential Misinformation Targets," *ACM Transactions on the Web* (13:2), p. 10. (https://doi.org/10.1145/3316809).

Dellarocas, C. 2006. “Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms,” *Management Science* (52:10), pp. 1577–1593. (https://doi.org/10.1287/mnsc.1060.0567).

Dellarocas, C., Zhang, X., and Awad, N. F. 2007. “Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures,” *Journal of Interactive Marketing* (21:4), pp. 23–45. (https://doi.org/10.1002/dir.20087).

Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, *55*(1), 412–421. doi: 10.1016/j.dss.2012.05.048

Desai, N., Gupta, R., & Truong, K. (2015). Plead or pitch? The role of language in kickstarter project success.

Dewan, S. and Ramaprasad, J. (2014). Social media, traditional media, and music sales. *Mis Quarterly*, 38(1):101– 122.

Dong, B., Li, M., and Sivakumar, K. 2019. “Online Review Characteristics and Trust: A Cross-Country Examination,” *Decision Sciences* (50:3), pp. 537–566. (https://doi.org/10.1111/deci.12339).

Dow, J. (2015). Passions and persuasion in Aristotles rhetoric. New York, NY: Oxford University Press.

Duan, W., Gu, B., and Whinston, A. B. (2008). The dynamics of online wordofmouth and product sales—an empirical investigation of the movie industry. *Journal of retailing*, 84(2):233–242.

Elberse, A. (2007). The Power of Stars: Do Star Actors Drive the Success of Movies? *Journal of Marketing*, 71(4), 102-120.

Erdem, T., Keane, M. P., & Sun, B. (2008). A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science,* 27(6), 1111-1125.

Erkama, N., & Vaara, E. (2010). Struggles Over Legitimacy in Global Organizational Restructuring: A Rhetorical Perspective on Legitimation Strategies and Dynamics in a Shutdown Case. *Organization Studies*, *31*(7), 813–839. doi: 10.1177/0170840609346924

Etter, V., Grossglauser, M., & Thiran, P. (2013). Launch hard or go home! *Proceedings of the First ACM Conference on Online Social Networks - COSN 13*.

Faber, R. J., Meyer, T. P., and Miller, M. M. (1984). The effectiveness of health disclosures within children's television commercials. *Journal of Broadcasting & Electronic Media*, 28(4):463–476.

Faber, R., & O'Guinn, T. (1984). Effect of Media Advertising and Other Sources on Movie Selection. *Journalism Quarterly*, 61, 371–377.

Ferraro, R., Bettman, J. R., & Chartrand, T. L. (2009). The Power of Strangers: The Effect of Incidental Consumer Brand Encounters on Brand Choice. *Journal of Consumer Research*, *35*(5), 729–741. doi: 10.1086/592944

Fogg BJ, Eckles D (2007) The behavior chain for online participation: How successful web services structure persuasion. Proc. Second Internat. Conf. Persuasive Tech., Lecture Notes in Computer Science, Vol. 4744 (Springer-Verlag, Berlin), 199–209.

Galak, J., Kruger, J., & Loewenstein, G. (2013). Slow Down! Insensitivity to Rate of Consumption Leads to Avoidable Satiation. *Journal of Consumer Research*, *39*(5), 993–1009. doi: 10.1086/666597

Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing, 21*(6), 58-62.

Gaziano, C., and McGrath, K. 1986. "Measuring the Concept of Credibility," *Journalism Quarterly* (63:3), pp. 451–462. (https://doi.org/10.1177/107769908606300301).

Gazley, A., Clark, G., and Sinha, A. 2011. "Understanding Preferences for Motion Pictures," *Journal of Business Research* (64:8), pp. 854–861. (https://doi.org/10.1016/j.jbusres.2010.09.012).

Ghose A, Ipeirotis PG (2011) Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Trans. Knowledge Data Engrg*. 23(10):1498–1512.

Ghosh, B., & Stock, A. (2010). Advertising Effectiveness, Digital Video Recorders, and Product Market Competition. *Marketing Science*, *29*(4), 639–649. doi: 10.1287/mksc.1090.0544

Godes, D. and Mayzlin, D. (2004). Using online conversations to study wordofmouth communication. *Marketing science*, 23(4):545–560.

Greenberg, M. D., Pardo, B., Hariharan, K., and Gerber, E. Crowdfunding support tools: predicting success &failure. CHI EA '13 (2013).

Greenwald, A. G., & Leavitt, C. (1984). Audience involvement in advertising: Four levels. *Journal of Consumer Research*, 11(1), 581-592.

Greenwood, B. N., & Wattal, S. (2017). Show Me the Way to Go Home: An Empirical Investigation of Ride Sharing and Alcohol Related Motor Vehicle Homicide. *MIS Quarterly*, *41*(1), 163–187.

Guo, X., Liu, J., & Chen, Y. (2017, May). FitCoach: Virtual fitness coach empowered by wearable mobile devices. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications* (pp. 1-9). IEEE.

Guzdial, M., & Riedl, M. (2019). An Interaction Framework for Studying Co-Creative AI. *arXiv preprint arXiv:1903.09709*

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd - An anger superiority effect. *Journal of Personality and Social Psychology*, 54(6), 917 - 924.

Hargittai E (2007) Whose space? Differences among users and nonusers of social network sites. J. *Computer Mediated Comm*. 13(1): 276–297.

Hawkins, S. A., & Hoch, S. J. (1992). Low-Involvement Learning: Memory without Evaluation. *Journal of Consumer Research*, *19*(2), 212–225. doi: 10.1086/209297

Hendricks, K., & Sorensen, A. (2009). Information and the skewness of music sales. *Journal of political Economy*, 117(2), 324-369.

Ho, S. Y., & Bodoff, D. (2014). The Effects of Web Personalization on User Attitude and Behavior: An Integration of the Elaboration Likelihood Model and Consumer Search Theory. *MIS Quarterly*, 38(2), 497–520. doi: 10.25300/misq/2014/38.2.08

Holbrook, M. B. (1999). Popular appeal versus expert judgments of motion pictures. *Journal of consumer research*, 26(2):144–155.

Hoque, M. E., Courgeon, M., Martin, J. C., Mutlu, B., & Picard, R. W. (2013, September). Mach: My automated conversation coach. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (pp. 697-706). ACM.

Hou, J., Zhang, J., & Zhang, K. (2019). Can Title Images Predict the Emotions and the Performance of Crowdfunding Projects? *Proceedings of the 52nd Hawaii International Conference on System Sciences* | 2019,4439-4448.

Hu, N., Bose, I., Koh, N. S., and Liu, L. 2012. “Manipulation of Online Reviews: An Analysis of Ratings, Readability, and Sentiments,” *Decision Support Systems* (52:3), pp. 674–684. (https://doi.org/10.1016/j.dss.2011.11.002).

Hu, N., Liu, L., and Sambamurthy, V. 2011. “Fraud Detection in Online Consumer Reviews,” *Decision Support Systems* (50:3), On Quantitative Methods for Detection of Financial Fraud, pp. 614–626. (https://doi.org/10.1016/j.dss.2010.08.012).

Huang, J., Boh, W. F., and Goh, K. H. 2017. “A Temporal Study of the Effects of Online Opinions: Information Sources Matter,” *Journal of Management Information Systems* (34:4), pp. 1169–1202. (https://doi.org/10.1080/07421222.2017.1394079).

Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009a). Path data in marketing: An integrative framework and prospectus for model building. *Marketing Science*, 28(2), 320-335.

Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009b). Research note—the traveling salesman goes shopping: The systematic deviations of grocery paths from TSP optimality. *Marketing science*, 28(3), 566-572.

Hui, S. K., Huang, Y., Suher, J., & Inman, J. J. (2013). Deconstructing the “first moment of truth”: Understanding unplanned consideration and purchase conversion using in-store video tracking. *Journal of Marketing Research*, 50(4), 445-462.

Hunter, D. S., & Zaman, T. (2018). Opinion dynamics with stubborn agents. *arXiv preprint arXiv*:1806.11253.

Jagarlamudi, J., Daumé III, H., & Udupa, R. (2012, April). Incorporating lexical priors into topic models. *In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 204-213)*. Association for Computational Linguistics.

Jiang, Z. (J., Wang, W., Tan, B. C., & Yu, J. (2016). The Determinants and Impacts of Aesthetics in Users’ First Interaction with Websites. *Journal of Management Information Systems*, *33*(1), 229–259. doi: 10.1080/07421222.2016.1172443

Johar, G. V., & Roggeveen, A. L. (2007). Changing False Beliefs from Repeated Advertising: The Role of Claim-Refutation Alignment. *Journal of Consumer Psychology*, *17*(2), 118–127. doi: 10.1016/s1057-7408(07)70018-9

Kihlstrom, R. E., & Riordan, M. H. (1984). Advertising as a Signal. *Journal of Political Economy*, 92(3), 427-450.

Kim, A., and Dennis, A. R. 2019. “Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media,” *MIS Quarterly* (43:3), pp. 1025–1039. (https://doi.org/10.25300/MISQ/2019/15188).

Kim, A., Moravec, P. L., and Dennis, A. R. 2019. “Combating Fake News on Social Media with Source Ratings: The Effects of User and Expert Reputation Ratings,” *Journal of Management Information Systems* (36:3), pp. 931–968. (https://doi.org/10.1080/07421222.2019.1628921).

Kumar, N., Venugopal, D., Qiu, L., and Kumar, S. 2018. “Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Learning,” *Journal of Management Information Systems* (35:1), pp. 350–380. (https://doi.org/10.1080/07421222.2018.1440758).

Kuppuswamy, V., & Bayus, B. L. (2017). Does my contribution to your crowdfunding project matter?. *Journal of Business Venturing*, 32(1), 72-89.

Lappas, T., Sabnis, G., and Valkanas, G. 2016. “The Impact of Fake Reviews on Online Visibility: A Vulnerability Assessment of the Hotel Industry,” *Information Systems Research* (27:4), pp. 940–961. (https://doi.org/10.1287/isre.2016.0674).

Lau, R.Y.K.; Liao, S.Y.; Kwok, R.C.W.; Xu, K.; Xia, Y.; and Li, Y. Text mining and probabilistic language modeling for online review spam detection. *Transactions on Management Information Systems, ACM*, 2,4(2011), 1–30. doi:10.1145/2070710.2070716

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science*, 64(11), 5105-5131.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5), 881-894.

Lee, Y.J., Hosanagar, K., and Tan, Y. (2015). Do i follow my friends or the crowd? information cascades in online movie ratings. *Management Science*, 61(9):2241–2258.

Lian, Y., Dong, X., Chi, Y., Tang, X., and Liu, Y. (2019). An internet water army detection supernetwork model. *IEEE Access*, 7:55108–55120.

Lin, M., & Viswanathan, S. (2016). Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science*, 62(5), 1393-1414.

Liu, X., Shi, S. W., Teixeira, T., & Wedel, M. (2018). Video content marketing: The making of clips. *Journal of Marketing*, 82(4), 86-101.

Liu, X., Zhang, B., Susarla, A., & Padman, R. (2019). Go to YouTube and Call Me in the Morning: Use of Social Media for Chronic Conditions. Forthcoming, *MIS Quarterly*.

Lu, S., Xiao, L., & Ding, M. (2016). A video-based automated recommender (VAR) system for garments. *Marketing Science,* 35(3), 484-510.

Luca, M., and Zervas, G. 2016. "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud," *Management Science*, p. 36.

Luo, X., Zhang, J., and Duan, W. (2013). Social media and firm equity value. *Information Systems Research*, 24(1):146–163.

Manchanda, P., Dubé, J.-P., Goh, K. Y., & Chintagunta, P. K. (2006). The Effect of Banner Advertising on Internet Purchasing. *Journal of Marketing Research*, *43*(1), 98–108. doi: 10.1509/jmkr.43.1.98

Matthes, J., Wonneberger, A., & Schmuck, D. (2014). Consumers green involvement and the persuasive effects of emotional versus functional ads. *Journal of Business Research*, *67*(9), 1885–1893. doi: 10.1016/j.jbusres.2013.11.054

Mayzlin, D., Dover, Y., and Chevalier, J. 2014. "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *The American Economic Review* (104:8), p. 2421. (https://doi.org/10.1257/aer.104.8.2421).

Milgrom, P., & Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, 94(4), 796-821.

Miller, A. R. and Tucker, C. (2013). Active social media management: the case of health care. *Information Systems Research*, 24(1):52–70.

Mitra, T., & Gilbert, E. (2014, February). The language that gets people to give: Phrases that predict success on kickstarter. *In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 49-61).

Mohan, S., Venkatakrishnan, A., Silva, M., & Pirolli, P. (2017, February). On designing a social coach to promote regular aerobic exercise. In *Twenty-Ninth IAAI Conference*.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, 29(1), 1-16.

Mollick, E., & Nanda, R. (2016). Wisdom or madness? Comparing crowds with expert evaluation in funding the arts. *Management Science*, 62(6), 1533-1553.

Moon, S., Bergey, P. K., and Iacobucci, D. (2010). Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. *Journal of marketing*, 74(1):108–121.

Moon, S., Bergey, P. K., and Iacobucci, D. 2010. "Dynamic Effects Among Movie Ratings, Movie Revenues, and Viewer Satisfaction," *Journal of Marketing* (74:1), pp. 108–121. (https://doi.org/10.1509/jmkg.74.1.108).

MoralesArroyo, M. and Pandey, T. (2010). Identification of critical ewom dimensions for music albums. In *2010 IEEE International Conference on Management of Innovation & Technology*, pages 1230–1235. IEEE.

Muller, M., Geyer, W., Soule, T., Daniels, S., and Cheng, L.-T. Crowdfunding inside the enterprise: employee-initiatives for innovation and collaboration. Proc. CSCW'13, ACM (2013), 503–512.

Munzel, A. 2016. "Assisting Consumers in Detecting Fake Reviews: The Role of Identity Information Disclosure and Consensus," *Journal of Retailing and Consumer Services* (32), pp. 96–108. (https://doi.org/10.1016/j.jretconser.2016.06.002).

Nielsen, R. K., & Schrøder, K. C. (2014). The relative importance of social media for accessing, finding, and engaging with news: An eight-country cross-media comparison. *Digital journalism*, 2(4), 472-489.

Oreilly, C. A. (1980). Individuals and Information Overload in Organizations: Is More Necessarily Better? *Academy of Management Journal*, *23*(4), 684–696. doi: 10.5465/255556

Plucker, J. A., Kaufman, J. C., Temple, J. S., and Qian, M. (2009). Do experts and novices evaluate movies the same way? *Psychology & Marketing*, 26(5):470–478.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1), 51-59.

Ravid, S. A. (1999). Information, Blockbusters, and Stars: A Study of the Film Industry. *Journal of Business*, 72, 463–492.

Ravid, S. A. (1999). Information, blockbusters, and stars: A study of the film industry. *The Journal of Business*, 72(4):463–492.

Ross, B., Pilz, L., Cabrera, B., Brachten, F., Neubaum, G., and Stieglitz, S. 2019. "Are Social Bots a Real Threat? An Agent-Based Model of the Spiral of Silence to Analyse the Impact of Manipulative Actors in Social Networks," *European Journal of Information Systems* (28:4), pp. 394–412. (https://doi.org/10.1080/0960085X.2018.1560920).

Rudinow, J. (1978). Manipulation. *Ethics,* 88(4), 338-347.

Sauermann, H., Franzoni, C., & Shafi, K. (2019). Crowdfunding scientific research: Descriptive insights and correlates of funding success. *PloS one*, 14(1).

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., and Liu, Y. 2019. "Combating Fake News: A Survey on Identification and Mitigation Techniques," *ACM Transactions on Intelligent Systems and Technology* (10:3), p. 21. (https://doi.org/10.1145/3305260).

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K. C. (2019). Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach. *MIS Quarterly,* forthcoming.

Siering, M., Koch, J. A., & Deokar, A. V. (2016). Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems*, 33(2), 421-455.

Sinclaire, J. K., & Vogus, C. E. (2011). Adoption of social networking sites: an exploratory adaptive structuration perspective for global organizations. *Information Technology and Management*, 12(4), 293-314.

Smith, G. F., & Dorfman, D. D. (1975). The effect of stimulus uncertainty on the relationship between frequency of exposure and liking. *Journal of Personality and Social Psychology*, 31(1), 150-155.doi: http://dx.doi.org/10.1037/h0076174

Spenner P, Freeman K (2012) To keep your customers, keep it simple. *Harvard Bus. Rev.* 90:108–114.

Steigenberger, N., & Wilhelm, H. (2018). Extending Signaling Theory to Rhetorical Signals: Evidence from Crowdfunding. *Organization Science*, *29*(3), 529–546. doi: 10.1287/orsc.2017.1195

Strebe, R. (2016). Aesthetics on the web: effects on approach and avoidance *behaviour. Behaviour & Information Technology, 35*(1), 4-20.

Suddaby, R., & Greenwood, R. (2005). Rhetorical Strategies of Legitimacy. *Administrative Science Quarterly*, 50(1), 35–67. doi: 10.2189/asqu.2005.50.1.35

Sundar, S. S. (2000). Multimedia effects on processing and perception of online news: A study of picture, audio, and video downloads. *Journalism & Mass Communication Quarterly*, 77(3), 480-499.

Sundblad, W. (2018, October 18). Data Is The Foundation For Artificial Intelligence And Machine Learning. Retrieved from https://www.forbes.com/sites/willemsundbladeurope/2018/10/18/data-is-the-foundation-for-artificial-intelligence-and-machine-learning/

Susarla, A., Oh, J. H., & Tan, Y. (2012). Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research*, 23(1), 23-41.

Tam, K. Y., & Ho, S. Y. (2006). Understanding the impact of web personalization on user information processing and decision outcomes. *MIS quarterly*, 865-890.

Tanner, R. J., & Maeng, A. (2012). A Tiger and a President: Imperceptible Celebrity Facial Cues Influence Trust and Preference. *Journal of Consumer Research*,39(4), 769-783.

Teixeira, T., Wedel, M., & Pieters, R. (2012). Emotion-Induced Engagement in Internet Video Advertisements. *Journal of Marketing Research*, *49*(2), 144–159. doi: 10.1509/jmr.10.0207

Trafton, A. (2014). In the blink of an eye. MIT News. Retrieved from http://news.mit.edu/2014/in-the-blink-of-an-eye-0116 on 2019.11.12

Tucker, C. E. (2015). The Reach and Persuasiveness of Viral Video Ads. *Marketing Science*, *34*(2), 281–296. doi: 10.1287/mksc.2014.0874

Valkenburg PM, Peter J, Schouten AP (2006) Friend networking sites and their relationship to adolescents' well-being and social self esteem. *Cyberpsych. Behavior* 9(5):584–90.

Visentin, M., Pizzi, G., and Pichierri, M. 2019. "Fake News, Real Problems for Brands: The Impact of Content Truthfulness and Source Credibility on Consumers' Behavioral Intentions toward the Advertised Brands," *Journal of Interactive Marketing* (45), pp. 99–112. (https://doi.org/10.1016/j.intmar.2018.09.001).

Vosoughi, S., Roy, D., and Aral, S. 2018. "The Spread of True and False News Online," *Science* (359:6380), pp. 1146–1151. (https://doi.org/10.1126/science.aap9559).

Wang Y, Meister DB, Gray PH (2013) Social influence and knowledge management systems use: Evidence from panel data. *MIS Quarterly.* 37(1):299–313.

Wang, C., Fang, T. and Gu, Y., 2020. Learning performance and behavioral patterns of online collaborative learning: Impact of cognitive load and affordances of different multimedia. *Computers & Education*, 143, p.103683.

Wang, K., Xiao, Y., and Xiao, Z. 2014. “Detection of Internet Water Army in Social Network,” *in Proceedings of the 2014 International Conference on Computer, Communications and Information Technology*, Beijing, China: Atlantis Press. (https://doi.org/10.2991/ccit-14.2014.50).

Waters, R. D., Burnett, E., Lamm, A., & Lucas, J. (2009). Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. *Public Relations Review*, 35(2), 102-106.

Wattal, S., Telang, R., and Mukhopadhyay, T. (2009). Information personalization in a twodimensional product differentiation model. *Journal of Management Information Systems*, 26(2):69–95.

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration, 42*(7), 596-615

Wojnicki, A. C. and Godes, D. (2011). Signaling success: Strategically positive word of mouth. *University of Maryland, College Park Working Paper*.

Xiong, G. and Bharadwaj, S. (2014). Prerelease buzz evolution patterns and new product performance. *Marketing Science*, 33(3):401–421.

Xiong, G., & Bharadwaj, S. (2014). Prerelease buzz evolution patterns and new product performance. *Marketing Science*, 33(3), 401-421.

Xu, A., Yang, X.,Rao, H., Fu, W., Huang, S., and Bailey, B.P,. 2014a. Show me the money!: an analysis of project updates during crowdfunding campaigns. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA,

Xu, H., Cai, W., and Chen, G. (2014). oriented research on water army detection for bursty topics. In *2014 9th IEEE International Conference on Networking, Architecture, and Storage*, pages 78–82. IEEE.

Xu, H., Cai, W., and Chen, G. 2014. “Forum-Oriented Research on Water Army Detection for Bursty Topics,” in 2014 9th *IEEE International Conference on Networking, Architecture, and Storage*, China: IEEE, August, pp. 78–82. (https://doi.org/10.1109/NAS.2014.18).

Xu, J. (D., Benbasat, I., & Cenfetelli, R. T. (2014b). The Nature and Consequences of Trade-Off Transparency in the Context of Recommendation Agents. *MIS Quarterly*, 38(2), 379–406. doi: 10.25300/misq/2014/38.2.03

Yin, D., Mitra, S., and Zhang, H. (2016). Research note—when do consumers value positive vs. negative re views? an empirical investigation of confirmation bias in online word of mouth. *Information Systems Research*, 27(1):131–144.

Yu, L. L., Asur, S., and Huberman, B. A. (2015). Trend dynamics and attention in chinese social media. *American Behavioral Scientist*, 59(9):1142–1156.

Yu, L. L., Asur, S., and Huberman, B. A. 2015. "Trend Dynamics and Attention in Chinese Social Media," *American Behavioral Scientist* (59:9), pp. 1142–1156.

Zajonc, R. B. (1968). Attitudinal effect of mere exposure. *Journal of Personality and Social Psychology*, *9*(2), 1–27.

Zajonc, R. B., & Rajecki, D. W. (1969). Exposure and affect: A field experiment. *Psychonomic Science*, *17*(4), 216–217. doi: 10.3758/bf03329178

Zeng, X., & Wei, L. (2013). Social ties and user content generation: Evidence from Flickr. *Information Systems Research*, 24(1), 71-87.

Zhang, D., Zhou, L., Kehoe, J. L., and Kilic, I. Y. 2016. "What Online Reviewer Behaviors Really Matter? Effects of Verbal and Nonverbal Behaviors on Detection of Fake Online Reviews," *Journal of Management Information Systems* (33:2), pp. 456–481. (https://doi.org/10.1080/07421222.2016.1205907).

Zhang, J., & Liu, P. (2012). Rational Herding in Microloan Markets. *Management Science*, 58(5), 892–912. doi: 10.1287/mnsc.1110.1459

Zhang, X., Li, S., Burke, R. R., & Leykin, A. (2014). An examination of social influence on shopper behavior using video tracking data. *Journal of Marketing*, 78(5), 24-41.

Zhao, Yi, Yang, S., Narayan, V., and Zhao, Ying. 2013. "Modeling Consumer Learning from Online Product Reviews," *Marketing Science* (32:1), pp. 153–169. (https://www.jstor.org/stable/23361439).

Zhou, Mi, George H. Chen, Pedro Ferreira, Michael D. Smith. 2019. Disrupting Class: Using Video Analytics and Machine Learning to Understand Student Behavior Online. *Conference on Information Systems and Technology (CIST)*, Seattle, WA, Oct 2019

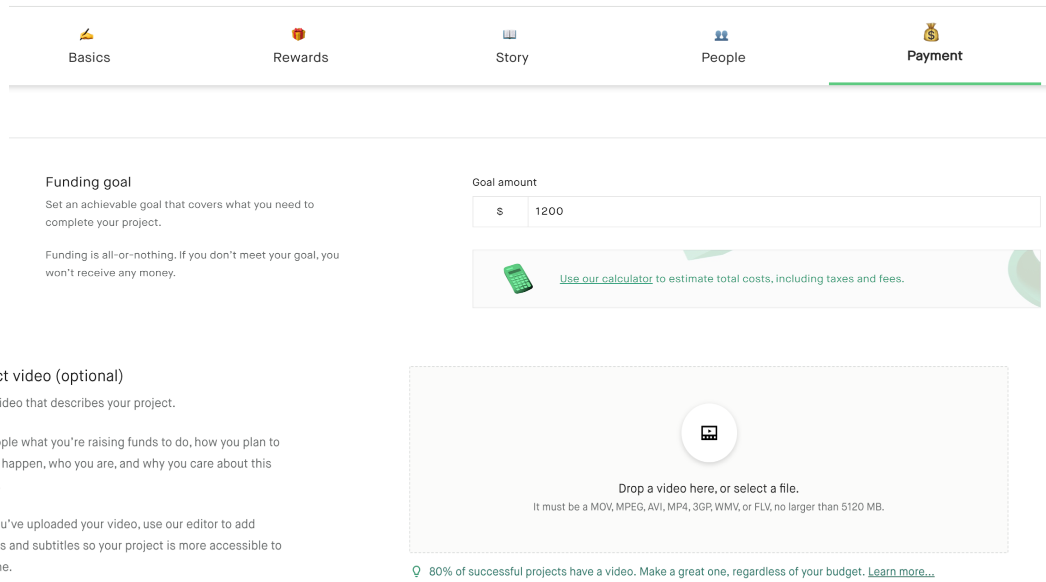
Zhu, F. and Furr, N. (2016). Products to platforms: Making the leap. *Harvard business review*, 94(4):72–78.

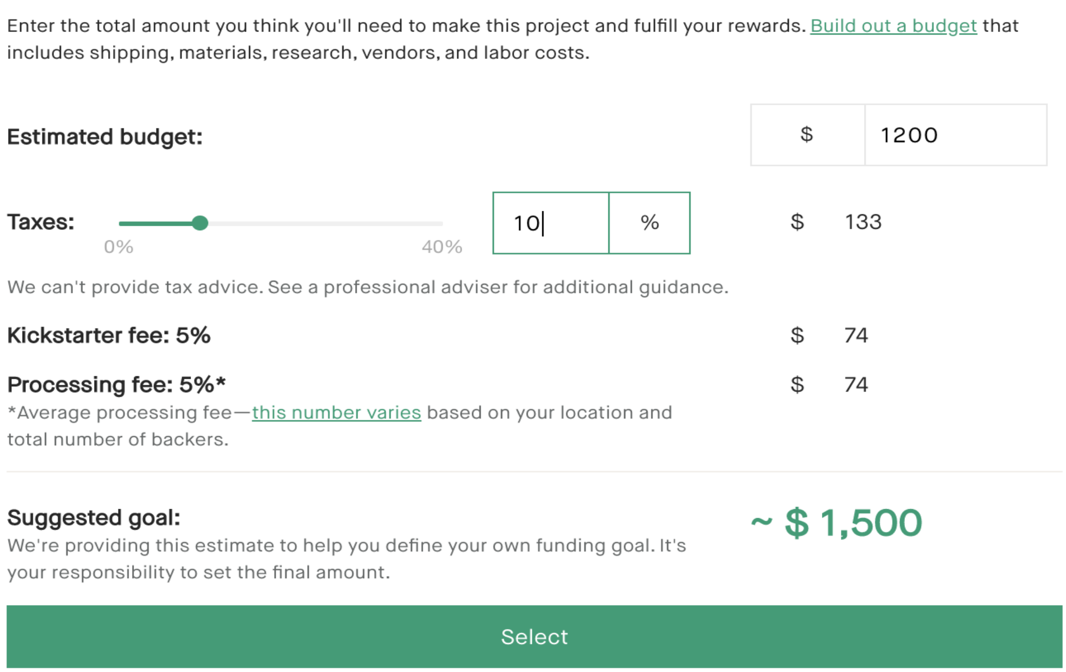
Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, 74(2), 133-148.

APPENDIX A

PROJECT CREATION SCREENSHOT

Appendix A. Project Creation Screenshot





Project description

Describe what you're raising funds to do, why you care about it, how you plan to make it happen, and who you are. Your description should tell backers everything they need to know. If possible, include images to show them what your project is all about and what rewards look like. Read more about telling your story

APPENDIX B

FEATURE SELECTION

Appendix B. Feature Selection

B.1 Appeal to Logic

Informativeness provides more information about the project because the main narrative is directly related to the project. The explicit presentation about the project is persuasive in nature (Akpinar & Berger 2017). One of the most important parts in the video to convey information and help the audience to make the decision is the narratives (Herzenstein et al., 2011). The narratives should make the consumers feel they have a better understanding toward the project because the narratives usually help to show the project creators' past experience, current situation and how they conceptualize their project and themselves (Wong & King, 2008; Herzenstein et al., 2011). Prior studies have found that the informativeness significantly influence consumers' purchase decisions. Akpinar and Berger (2017) found out that the informative appeals in advertising can boost brand evaluations and purchase intentions. Herzenstein et al., (2011) claimed that the narratives from the project owner can be associated with the project credibility and hence influence the crowdfunding project funding outcomes. Therefore, in our paper we will first include whether the project creators use narratives in the video as one variable of interest because narratives are long believed to provide more information.

In addition to the narratives, we identify the information categories. Different kinds of information serve different purposes. The project creators may introduce themselves in the video to increase the project credibility and trustworthiness. Introducing projects can give consumers a concrete understanding about the projects.

Crowdfunding projects are more like a value-exchange economics process (Herzenstein et al., 2017). In value-exchange economic activities, the backers expect that they could receive rewards after they invested. Including rewards in the video may attract more consumers to pledge in the projects.

B.2 Appeal to Emotion

Product aesthetics can change consumers' attitudes and is believed to improve consumers' product evaluation (Bloch, 1995, Page and Herr, 2002, Strebe 2016). Landwehr et al (2013) found out that consumers will always choose the attractive-looking product when they are given a choice between two products of comparable functions and prices and suggested that product aesthetics should be one of the most important market strategies. Jiang et al. (2016) found out that the users' perception of website aesthetics has a larger impact on their attitude towards the website. Prior studies mostly focus on the aesthetic score detection for image. Number of work uses small scale experiments Shin (et al., 2016) leverages a model built on Yahoo as a pipeline and uses a deep CNN model to classify the images and range the aesthetic score from 0 to 1 for a given image. To the best of our knowledge, our paper is the first few studies attempt to assign aesthetic scores to a video.

Plentiful studies largely use one scene as their basic level analysis of the videos. The video creators can control the way that the audiences experience the video through the length and the number of scenes included in the video (Liu et al., 2018). Previous studies in marketing have found out that the length of the scenes and the number of scenes significantly impact consumers consumption experience such as purchase intentions or post-purchase happiness. Galak et al., (2013) found out that the long and fast-paced scenes lead to a decrease in the overall enjoyment of the video game trailer because of the accelerated satiation. Liu et al., (2018) discovered that the fast-paced movie trailer with a larger number of scenes lead to a lower level of happiness and lower watching intentions. The audiences' happiness level generally increases as the sequence of scenes but so does the satiation. The authors found out that the optimal number of scenes in the movie trailer is shorter than the original movie trailer. The project campaign video serves the same purpose of the trailers that they both intend to introduce the products and attempt to take the consumers to the second level of participation, not just watching the video, but funding the project or watching the movie. We hence believe that the number of scenes will be a prime component.

Prior studies in marketing literature have found mixed results for having background music along with your advertisement. Park and Young (2005) explained that the effect of having background music on brand attitude is complex. Music has a positive impact on brand attitude at the beginning but music can distract audiences when they are interested in thinking about and learning information to make the decisions. Liu et al., (2018) discovered that silent movie trailers lead to a higher watching intentions compared with original trailers. But music can make the advertisement more appealing and increase the memorability to the advertisement. A beautiful song can help create a long-lasting impression and lead to a positive founding decision. We thereby suggest that including music in the video could be an important feature to impact the project funding outcomes.

The experience of emotion is one of the most fundamental and encompassing aspects of human existence. People often rely on their emotions to shape a wide variety of judgements including social, political, personal and economic decisions. Due to their importance, emotions have been studied in the domain of persuasion. Expressing emotional statements is considered as a compelling strategy to engage consumers in advertising

B.3 Appeal to Credibility

Human faces are one of the most important visual stimuli we grasp in our daily life (Tanner & Maeng, 2012). Kanwisher and Yovel (2006) claimed that human faces not only informed the identity of the person, but also disclose their mood, sex, age and direction of maze. From the marketing perspective, consumers appear to automatically process and classify the human faces in terms of their perceived trustworthiness (Engell et al., 2007; Tanner & Maeng, 2012). Tanner and Maeng 2012 found out that the perception of trustworthiness can be influenced by incorporating subtle facial cues. Therefore, we propose that including humans in the campaign video can significantly influence the audiences’ trust and then influence their funding decisions.

APPENDIX C

SEED WORDS LIST

Appendix C. Seed Words Lists

Topic	Seed Words List
Team	crew, seriously, experienced, collaborator, programmer, marketing, brothers, leadership, leader, developer, designer, post, artist, cinematography, work, compose, form, writer, director, research, develop
Product	characteristics, theme, role, evolution, feature, story, storytelling, attributes, expansion, design, interactive, aspect, language, gameplay, gears, ammunition, potions, experience, RPG, level, venture, activity, adventure, character, system, jobs, pack, adventure
Motivation	reason, inspired, suggest, fun, support, exciting, opportunity, ideas, excited, confident, feedback, assured, delivery, on-track
Reward	goal, pledge, access, rewards, tier, appreciate, appreciation, offer, bonus, promotion, price, share, bundle

APPENDIX D

DETAILED RESULTS

Appendix D. Detailed Results.

Table D-1. Content Allocation Results

	Product	Team	Motivation	Reward
Amount Pledged				
DesYesOrNo	0.216 (0.137)	-0.300 (0.185)	0.058 (0.145)	1.013*** (0.153)
DesLoading	-0.950 (0.738)	-3.270*** (0.574)	-1.131 (0.740)	2.121*** (0.572)
NarYesOrNo	0.051 (0.176)	-0.221 (0.189)	0.357 (0.192)	-0.147 (0.209)
NarLoading	-0.137 (0.729)	-0.982 (0.550)	1.430* (0.710)	0.850 (1.511)
Project Funded or Not				
DesYesOrNo	0.136 (0.230)	0.379 (0.307)	0.129 (0.257)	0.591** (0.249)
DesLoading	-0.754 (1.244)	-1.283 (1.007)	1.359 (1.283)	1.986 (1.124)
NarYesOrNo	-0.080 (0.297)	0.033 (0.320)	0.233 (0.330)	-0.359 (0.356)
NarLoading	-0.606 (1.208)	-0.196 (0.928)	0.986 (1.180)	0.442 (2.554)
Total Comments				
DesYesOrNo	0.209* (0.111)	-0.154 (0.149)	-0.010 (0.117)	0.611*** (0.124)

DesLoading	1.165* (0.593)	-1.713*** (0.461)	-0.527 (0.594)	2.101*** (0.460)
NarYesOrNo	0.044 (0.143)	-0.207 (0.153)	-0.044 (0.156)	0.204 (0.170)
NarLoading	0.453 (0.590)	-0.759* (0.446)	-0.294 (0.575)	2.529** (1.224)
First Day Pledged Percent				
DesYesOrNo	0.037 (0.144)	-0.301 (0.197)	0.103 (0.147)	0.660*** (0.169)
DesLoading	-1.852* (0.746)	-2.101*** (0.610)	-0.336 (0.778)	2.704*** (0.588)
NarYesOrNo	-0.192 (0.176)	-0.276 (0.182)	0.306 (0.181)	0.088 (0.202)
NarLoading	-0.821 (0.695)	-0.436 (0.543)	0.780 (0.712)	1.009 (1.379)
First Three - Day Pledged Percent				
DesYesOrNo	0.111 (0.142)	-0.291 (0.197)	0.153 (0.144)	0.560*** (0.169)
DesLoading	-1.411 (0.734)	-1.969*** (0.605)	-0.327 (0.798)	2.450*** (0.612)
NarYesOrNo	-0.158 (0.170)	-0.208 (0.176)	0.217 (0.175)	0.048 (0.196)
NarLoading	-0.683 (0.663)	-0.391 (0.535)	0.236 (0.687)	0.925 (1.314)

First Day Comments				
DesYesOrNo	0.231* (0.121)	0.003 (0.142)	0.087 (0.123)	0.315* (0.142)
DesLoading	0.433 (0.637)	-0.562 (0.522)	0.216 (0.665)	2.116*** (0.502)
NarYesOrNo	-0.249 (0.147)	-0.316* (0.152)	0.084 (0.151)	0.307 (0.169)
NarLoading	-0.880 (0.582)	-0.633 (0.455)	-0.217 (0.597)	2.311* (1.155)
First Three-Day Comments				
DesYesOrNo	0.220 (0.131)	0.094 (0.182)	-0.022 (0.133)	0.396** (0.156)
DesLoading	1.005 (0.680)	-0.590 (0.561)	-0.517 (0.739)	2.378*** (0.568)
NarYesOrNo	-0.215 (0.161)	-0.298 (0.167)	0.044 (0.165)	0.243 (0.185)
NarLoading	-0.542 (0.627)	-0.624 (0.505)	-0.435 (0.649)	2.042 (1.242)

Table D-2. Content Emotion Results

	Total Amount Pledged -OLS-	Funded or Not -Logistic-	Total Comments -OLS-	First Day Pledged Percent -OLS-	Three Day Pledged percent -OLS-	First Day Comments -OLS-	Three Day Comments -OLS-
Video - Joy	0.349 (0.221)	-0.524 (0.740)	0.338* (0.169)	0.104 (0.201)	0.009 (0.169)	0.060 (0.198)	0.026 (0.187)
Video - Sad	0.236 (0.249)	1.041* (0.450)	0.186 (0.202)	0.099 (0.246)	0.086 (0.206)	0.044 (0.239)	0.055 (0.226)
Video - Fear	-0.148 (0.487)	-0.417 (0.755)	-0.435 (0.395)	-0.160 (0.446)	-0.313 (0.374)	-0.071 (0.446)	-0.172 (0.420)
Video - Anger	-0.354 (0.457)	-0.524 (0.740)	0.251 (0.371)	-0.517 (0.497)	0.381 (0.417)	-0.082 (0.479)	0.512 (0.451)
Text - Joy	-0.043 (0.131)	0.198 (0.221)	-0.104 (0.103)	-0.038 (0.139)	-0.222 (0.115)	-0.020 (0.137)	-0.206 (0.125)
Text - Sad	0.534 (0.421)	0.486 (0.631)	-0.416 (0.331)	0.132 (0.495)	-0.579 (0.410)	0.367 (0.501)	-0.121 (0.457)
Text - Fear	-0.150 (0.476)	-0.262 (0.869)	0.215 (0.374)	0.098 (0.452)	0.233 (0.375)	0.064 (0.452)	0.261 (0.413)
Text - Anger	-0.315 (0.493)	-1.744* (0.823)	-0.491 (0.388)	-0.277 (0.569)	0.027 (0.471)	-0.421 (0.545)	0.056 (0.498)

Emotion Loadings							
Video - Joy	0.551 (0.332)	-0.242 (0.563)	0.532* (0.269)	0.165 (0.303)	0.075 (0.254)	0.057 (0.297)	0.033 (0.280)
Video - Sad	0.464 (0.437)	1.833* (0.777)	0.392 (0.354)	0.169 (0.430)	0.103 (0.361)	0.082 (0.417)	0.082 (0.393)
Video - Fear	-0.081 (0.822)	-0.692 (1.255)	-0.566 (0.666)	-0.223 (0.764)	-0.486 (0.640)	-0.003 (0.769)	-0.161 (0.725)
Video - Anger	-0.618 (0.797)	-1.061 (1.290)	0.442 (0.646)	-0.783 (0.875)	0.702 (0.733)	-0.084 (0.846)	0.917 (0.767)
Text - Joy	-0.086 (0.185)	0.249 (0.311)	-0.198 (0.145)	-0.125 (0.193)	-0.298 (0.160)	-0.102 (0.191)	-0.315 (0.174)
Text - Sad	0.908 (0.730)	0.853 (1.137)	-0.710 (0.574)	0.170 (0.885)	-1.067 (0.733)	0.612 (0.896)	-0.273 (0.818)
Text - Fear	-0.306 (0.702)	-0.534 (1.294)	0.259 (0.551)	-0.035 (0.702)	0.314 (0.582)	-0.082 (0.715)	0.344 (0.653)
Text - Anger	-0.580 (0.763)	-2.845* (1.300)	-0.808 (0.600)	-0.596 (0.954)	0.020 (0.790)	-0.804 (0.913	0.032 (0.833)

Table D-3. Reinforcement vs. Compliment Effect Result

	Total Amount Pledged -OLS-	Funded or Not -Logistic-	Total Comments -OLS-	First Day Pledged Percent -OLS-	Three Day Pledged percent -OLS-	First Day Comments -OLS-	Three Day Comments -OLS-
Repeat_Dummy	0.122 (0.168)	-0.026 (0.321)	0.004 (0.146)	0.088 (0.176)	0.082 (0.171)	-0.159 (0.153)	-0.106 (0.167)
Repeat_Loading	1.145*** (0.381)	0.524 (0.706)	0.212 (0.332)	0.814* (0.405)	0.720 (0.398)	-0.275 (0.356)	-0.194 (0.393)
New_Dummy	-0.192 (0.174)	-0.090 (0.331)	-0.127 (0.152)	-0.246 (0.181)	-0.172 (0.175)	0.017 (0.158)	-0.058 (0.172)
New_Loading	-1.046* (0.417)	-0.180 (0.769)	-0.147 (0.362)	-0.676 (0.451)	-1.229 (0.641)	0.407 (0.395)	0.270 (0.437)
Emotion_Joy	-0.533 (0.338)	-0.751 (0.572)	-0.176 (0.276)	0.042 (0.331)	-0.172 (0.322)	0.033 (0.279)	0.080 (0.307)
Emotion_Sad	1.498 (1.156)	15.071 (721.356)	0.703 (0.941)	2.945* (1.430)	1.927 (1.468)	0.140 (1.293)	0.345 (1.409)
Emotion_Anger	-0.792 (1.420)	0.489 (2.194)	-0.744 (1.150)	-0.906 (1.507)	-1.000 (1.432)	0.905 (1.271)	-0.560 (1.364)
Emotion_Fear	2.060 (1.924)	13.728 (882.744)	2.516 (1.562)	2.149 (1.572)	1.888 (1.500)	0.106 (1.329)	1.037 (1.435)

Significance code: *** 0.001, ** 0.01, * 0.05

APPENDIX E

MODEL COMPARISON

Appendix E1. Variable descriptions

E-1. Variable Descriptions

Project Variable (include appeal to credibility)	Include project duration, project creator information, amount asked, project reward levels and number of updates during the project durations.
Text Topic Dummy	Four binary variables indicating whether the project textual description includes the topic Team, Product, Reward and Motivation.
Text Topic Loading	The loading for topic Team, Product, Reward and Motivation in project textual description.
Text Emotion Dummy	Four binary variables indicate whether the description includes joy, sadness, anger and fear in project textual description.
Text Emotion Loading	The loading for emotion joy, sadness, anger and fear in project textual description.
Image Data	Number of content images.
Video Features	The aesthetic score and video feature generated by EFA using all other detected features, such as min, max, average and standard deviation of aesthetic score, music, narrative, including human, percentage of human showing in the video, video duration, and number of scenes.
Video Topic Dummy	Four binary variables indicating whether the campaign video includes the topic Team, Product, Reward and Motivation.
Video Topic Loading	The loading for topic Team, Product, Reward and Motivation in the campaign video.
Video Emotion Dummy	Four binary variables indicate whether the campaign video includes joy, sadness, anger and fear.
Video Emotion Loading	The loadings for joy, sadness, anger, and fear in the campaign video.
Content Comparison	The content comparison between campaign video and project textual description.

Emotion Comparison	The emotion comparison between campaign video and project textual description.

Appendix E-2. Model Comparisons.

E-2.1. Model Comparison 1
(Project Funded or Not)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13
Project Variable (include appeal to credibility)	√	√	√	√	√	√	√	√	√	√	√	√	√
Text Topic Dummy		√	√									√	
Text Topic Loading				√	√								√
Text Emotion Dummy			√									√	
Text Emotion Loading					√								√
Image Data						√	√	√	√	√	√	√	√
Video Features							√	√	√	√	√	√	√
Video Topic Dummy								√	√			√	
Video Topic Loading										√	√		√
Video Emotion Dummy									√			√	
Video Emotion Loading											√		√
Content Comparison													
Emotion Comparison													
Log Likelihood	-303.170	-283.430	-279.270	-284.398	-279.033	-298.346	-171.565	-170.540	-167.644	-170.576	-167.569	-152.483	-152.376
AIC	620.340	590.860	590.541	592.796	590.065	614.692	365.130	371.080	373.277	371.152	373.139	360.753	360.966
BIC	652.928	646.569	664.821	648.506	664.345	656.591	410.856	433.434	452.271	433.506	452.122	476.883	476.670

	Model 14	Model 15	Model 16	Model 17	Model 18	Model 19	Model 20	Model 21	Model 22
Project Variable (include appeal to credibility)	√	√	√	√	√	√	√	√	√
Text Topic Dummy		√			√			√	
Text Topic Loading			√			√			√
Text Emotion Dummy		√			√			√	
Text Emotion Loading			√			√			√
Image Data	√	√	√	√	√	√	√	√	√
Video Features	√	√	√	√	√	√	√	√	√
Video Topic Dummy		√			√			√	
Video Topic Loading			√			√			√
Video Emotion Dummy		√			√			√	
Video Emotion Loading			√			√			√
Content Comparison	√	√	√				√	√	√
Emotion Comparison				√	√	√	√	√	√
Log Likelihood	-171.483	-152.336	-152.289	-164.510	-150.695	-148.885	-164.307	-150.629	-148.814
AIC	366.966	362.673	362.577	375.021	365.391	377.769	376.613	367.259	379.628
BIC	416.850	482.730	482.634	470.631	497.867	543.363	476.381	503.875	549.363

E-2.2. Model Comparison
(Total Amount Pledged)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13
Project Variable (include appeal to credibility)	√	√	√	√	√	√	√	√	√	√	√	√	√
Text Topic Dummy		√	√									√	
Text Topic Loading				√	√								√
Text Emotion Dummy			√									√	
Text Emotion Loading					√								√
Image Data						√	√	√	√	√	√	√	√
Video Features							√	√	√	√	√	√	√
Video Topic Dummy								√	√			√	
Video Topic Loading										√	√		√
Video Emotion Dummy									√			√	
Video Emotion Loading											√		√
Content Comparison													
R-Square	0.483	0.553	0.554	0.561	0.564	0.565	0.565	0.570	0.575	0.574	0.578	0.620	0.628
Adjusted R-Square	0.479	0.546	0.545	0.555	0.555	0.561	0.556	0.557	0.558	0.561	0.561	0.596	0.605

	Model 14	Model 15	Model 16	Model 17	Model 18	Model 19	Model 20	Model 21	Model 22
Project Variable (include appeal to credibility)	√	√	√	√	√	√	√	√	√
Text Topic Dummy		√			√			√	
Text Topic Loading			√			√			√
Text Emotion Dummy		√			√			√	
Text Emotion Loading			√			√			√
Image Data	√	√	√	√	√	√	√	√	√
Video Features	√	√	√	√	√	√	√	√	√
Video Topic Dummy		√			√			√	
Video Topic Loading			√			√			√
Video Emotion Dummy		√			√			√	
Video Emotion Loading			√			√			√
Content Comparison	√	√	√				√	√	√
Emotion Comparison				√	√	√	√	√	√
Log Likelihood	-171.483	-152.336	-152.289	-164.510	-150.695	-148.885	-164.307	-150.629	-148.814
AIC	366.966	362.673	362.577	375.021	365.391	377.769	376.613	367.259	379.628
BIC	416.850	482.730	482.634	470.631	497.867	543.363	476.381	503.875	549.363

APPENDIX F

CONTENT GAP GRAPH FOR THE EXAMPLE

Appendix F. Content Gap Graph for the example

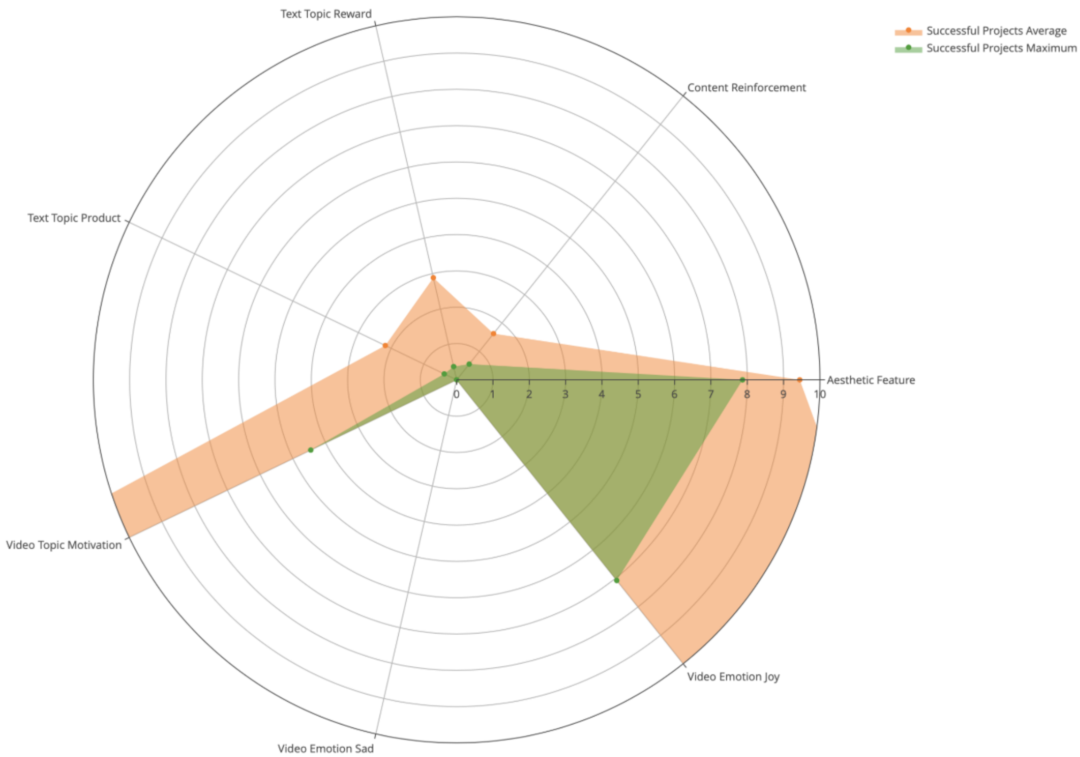


Figure F-1. Content Gap Illustration Example

Figure 4 illustrates the content gap between the focal product and all previously successful projects. The gap is illustrated with the blank area. The green area represents the comparison between the focal project and the maximum value of previously successful projects on selected features, and the orange area represents the comparison between the focal project and the average value of previously successful projects on selected features. Seven features that are positively connected with project outcomes are selected to be included in the content gap graph. Figure 4 clearly showed that the project owner should express some sadness in the campaign video, as on a scale of 10, the project creators include zero sadness in the video. The creators did a good job of having a high video aesthetic feature and expressing joy in the video. However, the creators need to deliberate more about the product and rewards in the project textual descriptions and discuss their motivations for starting this project in the campaign video. Further, the project creators can reinforce the information they provide more between campaign video and project textual description.

APPENDIX G

INTERNET WATER ARMIES EXAMPLES

Appendix G. Internet Water Armies Examples

On Jan. 7th, 2018, “Zi Guang Ge Gutter Oil” was shown in the Weibo trending rank. Gutter Oil means illegally recycled cooking oil in China. Zi Guang Ge is a magazine owned by the Chinese political party. How is a magazine associated with cooking oil on a Chinese social network? Magazine Zi Guang Ge wrote an article to criticize one rapper that he brought negative influences on younger fans by including the benefits of drugs in his lyrics. Those young fans hired the “Internet Water Army” to abuse Zi Guang Ge. However, because “Zi Guang Ge” sounds more like the name of a restaurant, the young fans falsely regard “Zi Guang Ge” as a restaurant and ask the water armies to link “Zi Guang Ge” with a food safety scandal. The fans paid an army company to tarnish the magazine that they believe it insults their “role model”.

The Internet Water Army is not just the tool being used by public relations managers, but it is becoming a culture in China. According to the statistics generated by the China Internet Network Information Center (CNNIC), there are around 564 million Internet users in China in 2019. The battleground for promoting and branding has shifted to the Internet. Therefore, it has incubated a new business under the table, Internet Water Army, which is defined as a group of internet ghostwriters, who get paid to post anything online with some particular contents. When using the Internet Water Army, a large number of people who are well organized to “flood” the Internet with purposeful posts, reviews, comments, and other actions. According to a survey conducted by Xinhua News, 74% of people believe the Internet Water Army is ubiquitous on Chinese social media websites and is widely used by public relations managers. According to a spokesman of

Weibo, the top social media platform in China (also known as Chinese twitter), around 40 percent of the trending hashtags on Weibo are created by water armies every day.

Internet Water Armies can be applied in numerous contexts. First of all, many corporations use Water Armies to implement marketing strategies such as creating buzz and generating awareness effects for certain products. For example, a famous TV drama show in China called "Eternal Love", generated 1.4 billion online click rates within a single day. However, the entire population of China is about 1.3 billion and there are only 564 million Internet users. I guess it is possible that every Internet user clicks Eternal love more than twice within a day but the chance is rare. Therefore, the authenticity of the number shown in the internet video platform is questioned. The truth is in the Water Army market, the corporations can spend as little as 100 RMB (around $15) to purchase 100,000 clicks. With the excessive amounts of clicks, this TV drama show has attracted many viewers' attention that it has been nominated and awarded for several famous TV play awards and keep the records of being the most viewed TV play online until today. In the offline market, where Water Armies can't fabricate the record, out of 30 days it is played on Television, Eternal Love has been the number one consumed TV play for the last 19 days. The strategy of flooding the online video platforms with "fake clicks" at the beginning of the show generates the buzz and attracts viewers' attention to the show. Actually, distributing firms of the entertainment industry have become the major players of hiring Water armies. Besides buying click rate, the firms hire water armies to flood movie reviews websites such as Douban and to post both positive and negative posts/comments on social media platforms to make the movie/show stay in the heat.

In addition, celebrities use the Internet Water Army to build and maintain their reputations online. Chen Yao, who used to be the "Weibo Queen" (the celebrity who has the most fans on Weibo) for a long time was suspected of buying "Zombie Fans", which means the fake followers injected by the Internet Water Army. Chen has more than 83 million followers on Weibo, but if you check her Weibo performance you can easily see that for one Weibo she posted, there are only 416 reposts, 422 comments, and 8,800 likes. The performance of her Weibo can't represent her number of followers. Many celebrities started on the foundation of water armies. Rumor said that the preeminent boy band TFBoys have benefited from the services. When they first enter the boy band market, they "live" on trending Weibo rank, which are the posts having the most repost/comments/likes on Weibo platforms. The company takes advantage of the Internet Water Army to make TFBoys look like a famous band. If the reader doesn't know them, it is his/her fault.

Water army can provide a variety of services. Unlike fake reviews or fake news, the Water Army can flood the platform with anything fake. Water Army can help inject fake clicks, fake fans, fake posts, and fake reviews. The marketing team can hire Water Army to mass-produce positive comments, posts, and ratings for them, or negative ones for competitors. Water Army also is capable of deleting the negative posts by repeatedly replying with illegal content to force the platform monitor to remove the thread. Further, Zombie Fans are also well applied that the marketing company can purchase Zombie Fans to follow a certain account or certain topic. Buying Zombie Fans and Water Army posts is different in terms of weights and intensity. In other words, the difference between buying Fans and buying posts is that many people spread the same information versus one person keeps repeating the same information. It is reasonable to contend that the first scenario of many people spreading the same information is more persuasive as there is an old saying “three liars makes a tiger”. Besides, publishing multiple similar contents by one account is a major criterion used by the social media platform to detect Water Army. The platform can filter these posts to avoid the flood of Internet Water Army. Furthermore, when the audience sees many similar posts coming from a few accounts, they are easy to link these contents with the Internet Water Army and form negative impressions.

The term Internet Water Army comes from China but the phenomenon of having “fake followers” and “fake posts” is universal. Facebook claimed that 5% of their online users are fake accounts and removed 3 billion fake accounts over six months (Romm, 2019). Some of the fake accounts belong to an obscure American company called Devumi. This company has collected millions of dollars by selling social media followers and retweeting to celebrities, businesses and anyone who paid it and wanted to become popular or exert influence online (Confessore et al., 2018).

APPENDIX REFERENCE

Akpinar, Ezgi, and Jonah Berger. "Valuable Virality." *Journal of Marketing Research*, vol. 54, no. 2, 2017, pp. 318–330., doi:10.1509/jmr.13.0350.

Bloch, P. H. (1995). Seeking the Ideal Form: Product Design and Consumer Response. *Journal of Marketing*, *59*(3), 16. doi: 10.2307/1252116

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala. *Journal of Cognitive Neuroscience*, *19*(9), 1508–1519. doi: 10.1162/jocn.2007.19.9.1508

Galak, J., Kruger, J., & Loewenstein, G. (2013). Slow Down! Insensitivity to Rate of Consumption Leads to Avoidable Satiation. *Journal of Consumer Research*, *39*(5), 993–1009. doi: 10.1086/666597

Herzenstein, M., Sonenshein, S., & Dholakia, U. (2011). Tell Me a Good Story and I May Lend You Money: The Role of Narratives in Peer-to-Peer Lending Decisions. *Journal of Marketing Research*, 48, S138-S149.

Jiang, Z. (J., Wang, W., Tan, B. C., & Yu, J. (2016). The Determinants and Impacts of Aesthetics in Users' First Interaction with Websites. *Journal of Management Information Systems*, *33*(1), 229–259. doi: 10.1080/07421222.2016.1172443

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361(1476), 2109-2128.

Landwehr, J.R.; Wentzel, D.; and Herrmann, A. Product design for the long run: Consumer responses to typical and atypical designs at different stages of exposure. *Journal of Marketing,* 77, 5 (2013), 92–107.

Liu, X., Shi, S. W., Teixeira, T., & Wedel, M. (2018). Video content marketing: The making of clips. *Journal of Marketing*, 82(4), 86-101.

Page, C., & Herr, P. M. (2002). An Investigation of the Processes by Which Product Design and Brand Strength Interact to Determine Initial Affect and Quality Judgments. *Journal of Consumer Psychology*, *12*(2), 133–147. doi: 10.1207/153276602760078668

Park, C., & Young, S. (1986). Consumer Response to Television Commercials: The Impact of Involvement and Background Music on Brand Attitude Formation. JMR, *Journal of Marketing* Research, 23(1), 11-24.

Strebe, R. (2016). Aesthetics on the web: effects on approach and avoidance *behaviour. Behaviour & Information Technology, 35*(1), 4-20.

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K. C. (2019). Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach. *MIS Quarterly*, forthcoming.

Tanner, R. J., & Maeng, A. (2012). A Tiger and a President: Imperceptible Celebrity Facial Cues Influence Trust and Preference. *Journal of Consumer Research*,39(4), 769-783.

Wong, N., & King, T. (2007). The cultural construction of risk understandings through illness narratives. *Journal of Consumer Research*, 34(5), 579-594.