ESSAYS OF THE VISUAL EFFECTS ON ONLINE HUMAN DECISION-MAKING AND DATA SCIENCE APPLICATIONS

by

Jian-Ren Hou

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy at
The University of Texas at Arlington
December, 2019

Arlington, Texas

Supervising Committee:

Dr. Jennifer Jie Zhang, Supervising Professor
Dr. Douglas Grisaffe
Dr. Jingguo Wang
Dr. Sridhar Nerur

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

LIST OF TABLES

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

ABSTRACT

Essays of The Visual Effects on Online Human Decision-Making and Data Science Applications

Jian-Ren Hou

The University of Texas at Arlington, 2019

Supervising Professor: Jennifer Jie Zhang

How human makes decisions online is an important research area in IS, many different techniques and online scenarios are used to probe into the related issues. In the first essay, this study investigates the crowdfunding platform. This study uses a deep neural network to extract the emotion metrics from the project images on the crowdfunding platform. The result shows that emotions in project images such as sadness and contentment can positively affect the performance of crowdfunding projects. This study also provides possible designs to change the emotions in the project images. In the second essay, this study investigates how to use different components of color to design effective warning messages to stop users and to study how malvertising attracts users to click on for cybersecurity issues. The result shows the different effects of red, green, blue, saturation, and lightness, which can be applied by researchers and practitioners in the future. In the third essay, this study applies data science techniques to improve the current SEM procedure. The essay provides new methods and new indexes to improve the SEM technique, which provides values for future SEM research.

Keywords: Crowdfunding, Deep Learning, Warning messages, Malvertising, Color, SEM, Data science, Cluster

CHAPTER ONE

Introduction

The Internet provides a unique online environment, in which users interact with others or interact with the environment itself. Observing and researching these interactions, this dissertation focuses on how to provide theoretical implications and managerial implications for researchers and practitioners.

To study online human decision-making, studies have observed various online environments and phenomena to develop research. Observing the crowdfunding platform, we find that each crowdfunding project requires a project image, and the project image can be one of the most salient parts of each project. The first essay uses the SOR (Stimulus, Organism, and Response) model to lay the foundation for this essay. For the first essay of this research, a deep neural network is used to extract emotion in project images of crowdfunding projects, and the first essay expects to see the results which show these emotions will affect the fundraising performance of crowdfunding projects in terms of how humans react to the project images and make funding decisions. On top of that, the first essay also employs computer vision techniques to extract the image attributes in each project image and suggest reasonable design in images to generate desired reactions.

Observing another online phenomenon regarding online human behavior, the second essay focuses the important issues in cybersecurity such as warning messages and malvertising (malicious advertising). The second essay develops research which concentrates on background color settings of different scenarios, warning message and malvertising. Based on the widely accepted color models such as the RGB and HSL models, the second essay aims to investigate how the color design affects human subjects' decision-making and provide elastic guidelines for researchers and practitioners to design effective warning message and to develop strategies to deal with the threat of malvertising.

Studying online human decision-making is a complex task, and how to investigate the psychological factors is a critical issue. SEM is a useful tool for researchers to probe into the systematic thinking process of humans. The third essay in the dissertation aims to use data science techniques to provide additional support for current SEM analysis. The third essay aims to employ word embedding and data science techniques, which allow us to detect possible low reliability constructs and catch possible common method bias before the actual data collection, and cluster techniques are used to support the current factor analysis results after data collection.

Coauthors help the author to improve and further complete this dissertation. Dr. Jennifer Jie Zhang improves the overall structure of all essays and keeps all essays pertinent to the respective research questions and contributions by her extraordinary knowledge. Dr. Kunpeng Zhang uses his outstanding computer science knowledge to help the author to extract the important metrics in the first essay. Dr. Jingguo Wang provides his valuable knowledge of cybersecurity and makes sure the second essay echoes the important needs in cybersecurity area. Dr. Ying-Feng Kuo addresses the culture difference issues to ensure the validity of the second essay. Dr. Sridhar Nerur makes sure the proper data science techniques are being employed appropriately in the third essay.

In short, this dissertation investigates issues and techniques related to or used on online human decision-making. The rest of the dissertation is organized as follows. In Chapter 2, essay 1 investigates the project image in the crowdfunding platform; In Chapter 3, essay 2 investigates the background color design’s effect on warning message and malvertising; In Chapter 4, essay 3 introduces the new methodology to support current SEM analysis; In Chapter 5, the overall results and implications of this dissertation will be provided and discussed.

CHAPTER TWO

HOW EMOTIONS IN PROJECT IMAGES DRIVE THE SUCCESS OF CROWDFUNDING CAMPAIGNS? AN IMAGE DESIGN PERSPECTIVE

Jian-Ren Hou

Jie Jennifer Zhang

Kunpeng Zhang

Abstract

Crowdfunding is a novel way of raising funds through the contributions of many individuals. However, most crowdfunding campaigns fail. For example, more than 60% of the projects on Kickstarter.com fail to reach the goals. Hence, it is imperative to study how to improve the successfulness of crowdfunding campaigns. In this paper, this study focuses on the design of project images that display the themes and contents of the projects on a crowdfunding website. This study conducts two separate studies to verify the Stimulus-Organism-Response (SOR) model, which includes the relationships between image attributes (Stimuli) and image emotions (Organism), and between emotions of project images (Organism) and campaign outcomes (Responses). This study first implements a deep neural network machine learning model to identify the emotions conveyed in the project images, such as, awe, anger, contentment, excitement, and sadness. Then, using the images and project data collected from a popular crowdfunding platform kickstarter.com, this study empirically demonstrates several emotions such as sadness, contentment, and fear, strongly affect the outcomes of crowdfunding projects, in terms of backer number, the amount raised, and percentage of goal achieved. From a design perspective, this study explores what image aesthetic attributes (e.g. composition, color, contrast, and content) evoke each emotion. In the second study, this study examines the above model from an experimental approach. That is, this study manipulates images with different levels of image attributes to show that image attributes can influence the emotions. Moreover, the experimental data also suggests that participants' emotions are related to their pledge intention. This research contributes to the crowdfunding literature from a novel perspective of emotions in project images. Our machine learning model classifies image emotions in an objective way, and achieves significantly better prediction than the state-of-the-art baselines. Also, the findings in this paper provide useful insights for seekers to design successful crowdfunding campaigns.

Keywords: Crowdfunding, Emotions, Project Images, Deep Learning, Image Attributes

1. Introduction

Crowdfunding is an open call funding mechanism that solicits small funds from the general public through online digital platforms for the purpose of financing projects (Ahlers et al., 2015; Allison et al., 2015; Mollick, 2014; Ordanini et al., 2011). While crowdfunding is a relatively new phenomenon, it has quickly gained popularity. Worldwide, crowdfunding raised over $5.3 billion in 2018 (Statista 2019). As a leading crowdfunding platform that "helps bring creative projects to life," Kickstarter.com has raised over $4.6 billion for 170,334 projects (Kickstarter Stats, September 2019). Although many people turn to crowdfunding to support their projects financially, not every campaign is funded successfully. For example, more than 60% of Kickstarter campaigns fail to reach their goal (Kickstarter Stats, June 2019); on Indiegogo.com, another major crowdfunding platform, the unsuccessful rate is over 80% (Liu, 2018). Hence, it is imperative to study how to improve the successfulness of crowdfunding campaigns.

Echoing this need, recent crowdfunding studies have focused on identifying the key factors that impact the success of crowdfunding projects, and in particular, the project setting parameters such as duration, award amount, and preset goal (Ahlers et al., 2015; Allison et al., 2015; Burtch et al., 2013; Mollick, 2014; Zheng et al., 2014). Our study complements this literature by investigating how to improve the performance of crowdfunding projects through the design of *project images*[1], which are visual displays about the themes and contents of the funding projects on the platform website (Figure 1). The project image of a crowdfunding campaign is of great importance in attracting backers' attention and motivating them to pledge. Kickstarter (2019a), in its online tutorial, reminds project designers to consider it thoughtfully, as "it's the first part of your project people will see — you'll want to make a good first impression."

[1] This term is adopted from Kickstarter.com, and it is called "campaign card image" by Indiegogo.com.

Therefore, this paper investigates two questions in the context of crowdfunding. First, are emotions reflected by a project image a critical factor affecting crowdfunding performance? If so, what is the quantifiable effect of the project image's emotional content on the performance of a crowdfunding project? Second, what attributes of a project image, such as composition, color, content, and the relation between main body and background, are related to its emotions? To answer these questions, this study conducts both empirical analyses and lab experiments. This study expects to understand the relationships between emotions in project images and crowdfunding project success, and to provide fundraisers practical suggestions on how to design project images when launching crowdfunding projects.

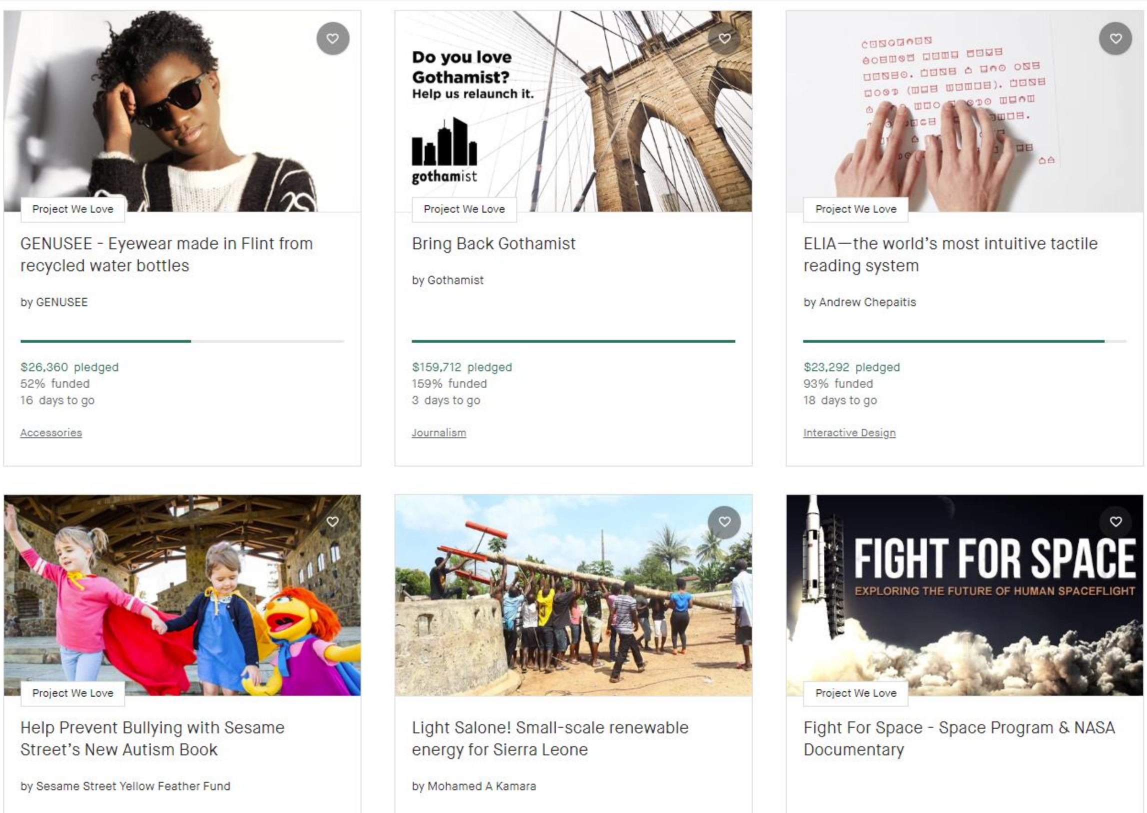


Figure 1. An Example of Some Project Images on Kickstarter.com.

Previous research has shown the significance of images in influencing business outcomes, but rarely focused on image emotions. In the e-commerce setting, Peck & Childers (2003) conducted lab experiments to show that product images can lead to higher perceived quality for certain consumers, and consumers' inability to touch or hold the real products makes this kind of visual cues especially important. In addition, Bland et al. (2007), Chung et al. (2012), Di et al. (2014), Goswami et al. (2011) and Kelley (2002) have shown that, product images do play important roles in influencing users' trust, risk perception, attitudes and purchase intentions in terms of click-through rate and conversion rate. A study closely related to ours is Small and Verrochi (2009), which used images with user confirmed emotions in experiments and showed that "sad" emotion could solicit a significantly larger donation amount than "happy" and neutral emotions. Following and extending the above literature, this study identifies emotions in project images with objective machine learning techniques, and conducts more thorough analyses based on empirical and experimental results to quantify the impacts of image emotions on the outcomes of crowdfunding campaigns. This study also associates different image design attributes with emotions.

This study develops our research model based on the Stimulus-Organism-Response (SOR) theory (Figure 2): the objective image attributes (Wang et al., 2013; Zhang et al., 2017) as the external stimulus, the emotions this study extracts from the project images as the organism (emotional state), and crowdfunding campaign outcomes as the response. Two hypotheses are developed based on the connections between the adjacent model components, e.g. stimulus and organism, and organism and response.

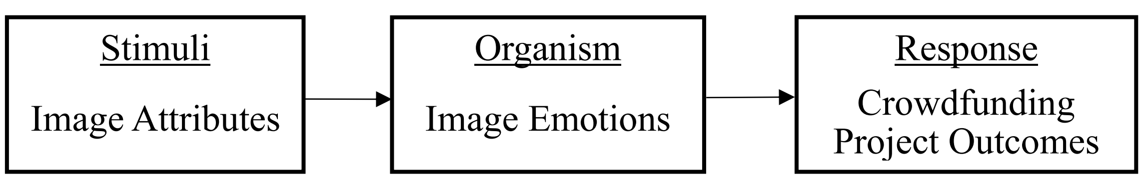


Figure 2 Research Model

This model is first verified through empirical analyses based on data collected from the "public benefit" category of Kickstarter.com, including project images and project-related information such as preset funding targets, participant numbers, the total amount raised, and textual descriptions. This study develops a state-of-the-art deep neural network-based classifier (Appendix B) to predict general viewers' emotional reactions (e.g. amusement, contentment, fear, sadness), which we refer to as "image emotions" to those project images. To understand the relationships between image characteristics and image emotions, this study also applies image analysis tools to measure several objective image attributes (such as composition, color, and the relationship between main figure and background).

This study also conducts a separate study to further verify the hypotheses and the research model with experiments. With manipulated images with different levels of image attributes, this study shows that manipulated image attributes can influence the emotions of participants. Moreover, based on the data collected from the experiments, our logistic regression model also suggests that participants' emotions are related to their pledge intention.

This study contributes to the literature and crowdfunding practice in the following ways. First, this study is among the first studies demonstrating emotions in project images play a critical role in the success of crowdfunding campaigns. Compared with the related literature, which mainly focuses on the impact of limited and self-reported facial emotions on traditional fundraising, this study identifies richer emotion types (e.g. contentment, fear, sadness) in broader topics of images and empirically examines their roles in a new crowdfunding context.

Second, this study utilizes deep neural network techniques to extract image emotion, and this study is one of the first to apply these image emotion measurements in a business context to provide empirical results. Compared with prior literature that analyzes emotions with self-reported surveys, our deep neural network techniques suggest a new way to conduct emotion-related studies. On top of that, since the measurement of image emotion is possible, we also show the relative importance and impact between the text emotions and image emotions accordingly.

Third, the various advanced methodologies and tools are employed in this paper provide new angles from which to study emotions. For example, this study implements machine learning techniques, (i.e., deep learning, text mining, and image analysis) to examine emotional impacts. This study also conducts empirical studies and experiments to understand the relationships between image emotions and crowdfunding outcomes.

Finally, this research implements the *objective* measurement of image attributes in order to understand better the connection between an image and the emotions it evokes. Our results suggest that specific attributes of the project images (e.g., composition, color, content) are related to the viewer's emotions. This finding creates a bridge from a more aesthetic and intuitive understanding of photographic imagery to a more quantifiable understanding of how emotions and fundraising performance play out in crowdfunding scenarios. Our results provide actionable and practical guidance to crowdfunding seekers and platform owners for improving project performance through optimal design of project images.

The rest of the paper is organized as follows. Section 2 summarizes existing literature and theories, and discusses our research model in light of earlier works. Section 3 describes our machine learning algorithms that identify emotions in the project images, and our empirical analyses based on data collected from Kickstarter.com. This empirical study helps us understand how image emotions are related to crowdfunding performance and how image attributes can explain image emotions. Section 4 introduces randomized experiments to further confirm our empirical findings; Section 5 discusses the implications, and finally, concluding the paper and discuss limitations and directions for future research in Section 6.

2. Literature Review and Hypotheses Development

This study reviews two relevant streams of literature, namely, crowdfunding and images, and the SOR model adopted as a basis for our research model.

2.1 Literature on Crowdfunding

Traditionally, individuals draw from limited funding sources in order to complete their projects. The Internet provides a unique platform for individuals or organizations to seek financial support from a considerable number of individuals, and accumulates the negligible singular contributions into a significant amount (Mollick, 2014; Ordanini et al., 2011).

Researchers have investigated several antecedents for successful crowdfunding projects. The preset goal, duration, featured status, geographic factors (project location), framing of the narrative, funder's social network, and even the social networks of crowdfunding projects have all been identified as factors affecting the success of crowdfunding projects (Allison et al., 2015; Hong et al., 2018; Lin & Viswanathan, 2015; Mollick, 2014). For example, Mollick (2014) found that project duration, funding goal, featured status (a designation displayed on the crowdfunding platform), and the project founder's social network is associated with crowdfunding success. Increasing the fundraising goal is negatively associated with success as well, and featured status on a platform such as Kickstarter is positively associated with success. Founders with bigger social networks also have a better chance of success.

Prior crowdfunding studies have also explored different metrics for measuring the success of crowdfunding projects. These metrics include successful fundraising, percentage of preset goal achieved, speed in achieving the funding goal, frequency of donations, and fundraising amount (Ahlers et al., 2015; Allison et al., 2015; Burtch et al., 2013; Hobbs et al., 2016; Kang et al., 2016; Mollick, 2014; Yuan et al., 2016; Zheng et al., 2014). This study uses the three most commonly adopted metrics: the total amount raised, the number of backers, and the percentage of goal achieved.

This study differs from and expands the scope of the existing crowdfunding literature by taking the unique approach of examining the impact of emotions evoked by project images on crowdfunding performance.

2.2 Literature on Images

Construal Level Theory (Trope & Liberman, 2010) states that people's perceived psychological distance affects how they think about and describe objects. Psychological distance has four dimensions, including time (now vs. past and future), space (far vs. near), social distance (acquaintance vs. not an acquaintance), and hypotheticality (high possibility vs. low possibility). People constitute their psychological distance using a reference point, i.e., themselves (Trope & Liberman, 2010). For instance, in the time dimension of psychological distance, people consider the future and the past to have more psychological distance than the present. When people try to describe an object at a great psychological distance, they usually prefer a broad and inexplicit mode of description (high-level construal). However, when they try to describe an object at a close psychological distance, they prefer a detailed and explicit mode (low-level construal). Pictures are considered as examples of low-level construal and disseminate a sense of closeness (Amit et al., 2009). Thus, although a crowdfunding project itself exists at a great psychological distance from the viewer in all four dimensions, project images are presumably essential because their low psychological distance creates that sense of closeness. The effect of project images, therefore, needs to be explored in the context of crowdfunding. Due to the limited research in the existing literature, this study believes that this paper can contribute to the crowdfunding theory and help funders achieve better crowdfunding performance.

In the web-based e-commerce setting, images have been shown to be an important and influential determinant to the success of e-commerce. For example, in the early stages of e-commerce research, Gilkeson & Reynolds (2003) proposed that images on auction pages should be considered in future e-commerce research to reveal further more information about the determinants of success. Further research investigated the effect of the presence of real photos or stock photos (existing product-related images created for users to purchase and use directly), and the results indicated that the presence of photos would positively influence the outcomes of e-commerce (Bland et al., 2007). Researchers have further considered not only the presence of images but also the features of those images, such as brightness and contrast, and regional features, such as background and foreground settings (Chung et al., 2012; Goswami et al., 2011). These studies concluded that these attributes are important components related to the click-through rate (Chung et al., 2012; Goswami et al., 2011). Besides, Di et al. (2014) showed that the display of photos, and the count and quality of images on an e-commerce website are positively related to online shoppers' attention, trust, and conversion rate. As in the e-commerce setting, the crowdfunding platforms also use websites as an interface or medium to communicate with users, who do not see the real products or outcomes before payment and shipment. Some crowdfunding projects are even considered to be the equivalent of pre-ordering products (Belleflamme et al., 2014). Thus, this study expects that the prior research on the role of images in e-commerce success can be related and applied to the crowdfunding context.

Images are commonly considered in the literature to cause emotional arousal (Carroll, 2003). Existing literature (Carroll, 2003) mentioned that emotional responses are crucial in the viewer's response to artworks. Other studies have stated that there is an intimate relationship between art and emotions (Silvia, 2005), and that emotions should be involved in aesthetic appreciation (Barry, 2006). Further, different ways of presenting the same object may cause different emotional feelings for the viewers (Baberini et al., 2015). Instead of focusing on only images or only emotions, this study combines the findings of prior research to employ image attributes as antecedences of emotions, and propose hypothesis 1:

Hypothesis 1: ***Image attributes of project image affect the emotions in project images.***

Visual presentations can transmit ideas to the viewers and further make viewers more willing to try the focal product (Chaudhuri & Micu, 2014). Choi et al. (2011) also showed that emotions have a significant effect on users' acceptance. In the case of products that consumers have not experienced before, emotions play an especially important role in a potential buyer's evaluation of the products (Wood & Moreau, 2006). In the context of charity fundraising, Small and Verrochi (2009) suggested that emotion plays an important role in the donation amount, since they find that viewers who see sad photographs donate more. Thus, this study proposes our hypothesis that the image emotions have an effect on crowdfunding performance.

Hypothesis 2: ***Emotions of the project image affect the project's performance.***

2.3 Stimulus-Organism-Response (SOR) Model

The Stimulus-Organism-Response (SOR) model is a way of describing how individuals react to external environmental stimuli. The model describes human reactions in three steps: after receiving an external stimulus (S); individuals will generate an affective reaction and internal emotional state (O), depending on the stimulus, an emotional state (O) is generated in their mind internally; which in turn affects the individuals' actual behavior (R) (Mehrabian & Russell, 1974). The SOR model is widely used in discovering consumer behavior in different settings that include both offline and online environments. For offline environments, previous studies have employed the SOR model to reveal how atmospheric factors in physical retail stores, such as store design, affect consumer behavior in the store (Singh et al., 2014). In the online setting, one study of an auction website showed that online stimulus, such as technological services, network effect, and product diversity, affect consumers' internal organisms and their loyalty (Cui & Lai, 2013). Similarly, in online shopping scenarios, studies utilized the SOR model to illustrate how customers' intention to repurchase or revisit is formed from different stimuli such as the color scheme of the online store (Ettis, 2017; Peng & Kim, 2014). The SOR model has also been adopted in social network scenarios to explore why users discontinue their usage of Facebook (Luqman et al., 2017). In an online virtual environment, the SOR model was used to examine how the virtual environment affects users' purchase intention to virtual products (Animesh et al., 2011). Overall, the SOR model has shown its generalizability to explore user behavior in different offline and online environments. Furthermore, the generalizability of the SOR model to probe into consumer behaviors has been supported by a meta-analytical study, which also supported the strong association among stimulus, emotion, and response (Vieira, 2013). Thus, we employ the SOR model to explore how backers react to stimuli, namely, the project images in crowdfunding. Our research model has three different parts. The objective image attributes (Wang et al., 2013; Zhang et al., 2017) serve as the external stimulus in the SOR model, and the image emotions we extract from the project images are the organism (emotional state), which in turn affects whether viewers participate in the crowdfunding campaign or not (response). The overall research model based on the SOR is summarized in Figure 3.

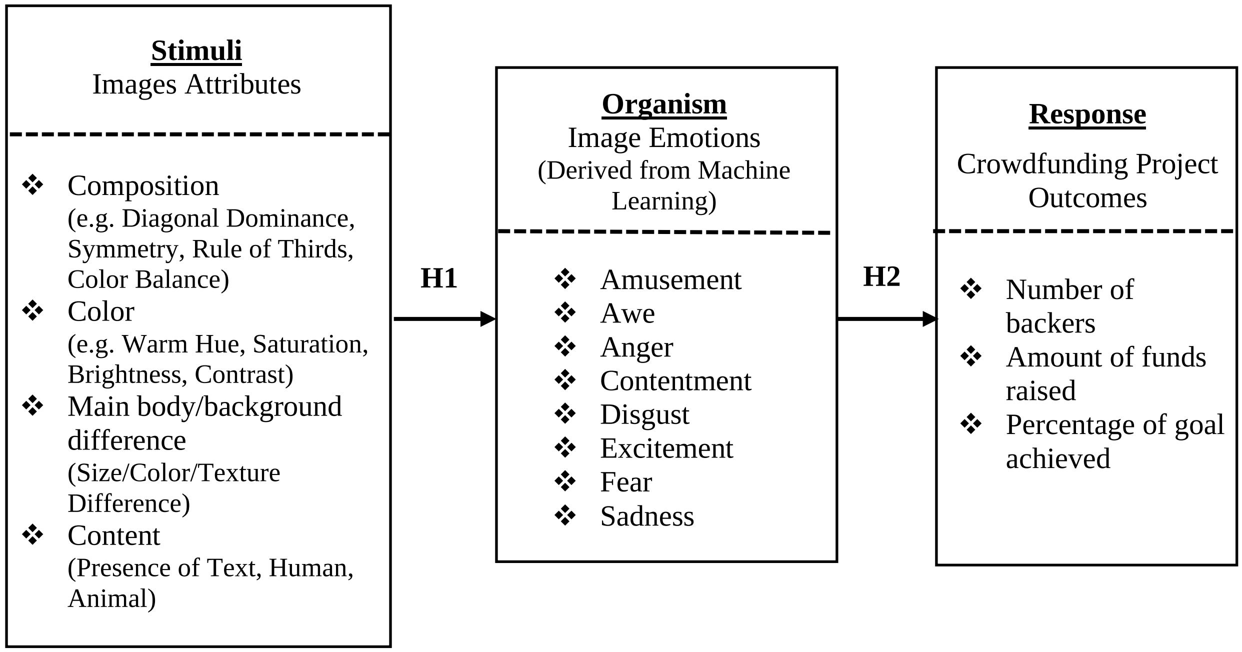


Figure 3 The Proposed Research Model

Our work builds on this literature of image and emotion and expands on other innovative developments in machine learning to analyze images. Instead of identifying image emotions through human labeling, we develop machine learning algorithms to recognize the emotions in project images for the benefit of efficiency, accuracy, and objectivity, and for the sake of avoiding such issues as response bias and noises in understanding images that arise from the individual characteristics of the human labeler (culture, gender, etc.). The foundations for this work lie in many domains, particularly computer vision, that make an extensive study of images as one of the most common and important sources of unstructured data. This has led to the development of various algorithms to analyze and process images, with capabilities ranging from very basic feature extraction (Lowe, 1999) to advanced object detection (Dalal & Triggs, 2005). As artificial intelligence has advanced and deep learning has achieved great success in almost every domain, both academic and industrial researchers have proposed and implemented a variety of deep neural network-based models to understand image semantics, such as objects embedded in images (Google Vision API), image memorability (Khosla et al., 2015), image aesthetics (Kong et al., 2016), and emotions of images with human faces (Emotion API, Azure, Amazon Rekognition, Google Vision API). These models, including VGG (Simonyan & Zisserman, 2014), ResNet (He et al., 2016), Xception (Chollet, 2017), and many custom models created by researchers, have proven themselves effective in many downstream business applications. For example, Zhang et al. (2017) leveraged large-scale image analytics and developed convolutional neural network models to estimate Airbnb property demand. Liu et al. (2018) developed deep convolutional neural networks to measure brand attributes (glamorous, rugged, healthy, fun) from images and then apply the classifiers to brand-related images posted on social media to measure what consumers are visually communicating about brands. Our work is similar to most existing work regarding the classification model in that we leverage the deep learning framework and tailor it for our specific context – emotion detection. However, it is different in that we do not use pixel-level features as inputs; rather, we use mid-level features (see details in Appendix B). Therefore, our deep learning model can predict image emotions with higher accuracy than state-of-the-art baselines, such as those reported in Machajdik and Hanbury (2010) and You et al. (2016).

To the best of our knowledge, this study is the first study that examines how image emotions, derived from machine learning tools, affect the performance of crowdfunding projects. To highlight the contribution of this study, Table 1 summarizes and contrasts this study with the prior works that have empirically or experimentally studied the relationships among image, emotion, and fundraising or e-commerce.

Table 1. Summary of Major Literature.

Articles	Image	Image Emotion	Text Emotion	Fundraising Performance	E-tailing performance	Crowdfunding
Carroll (2003)	X	X				
Machajdik & Hanbury (2010)	X	X				
Bland et al. (2007)	X				X	
Chung et al. (2012)	X				X	
Di et al. (2014)	X				X	
Goswami et al. (2011)	X				X	
Kelley (2002)	X				X	
Peck & Childers (2003)	X				X	
Chowdhury et al. (2008)	X	X			X	
Small & Verrochi (2009)	X	X		X		
Liang et al. (2016)			X	X		
Our study	X	X		X		X

In the next two sections, this study will test the hypotheses through two studies based on two different approaches: empirical analyses and experiments.

3. Study 1: Empirical Studies

Based on the SOR model, internal emotions (O) affect human behavior (R). Thus, this study would like to show the effect of image emotions in project images on backers' behaviors, namely, pledge behavior (H2), and this study further shows the relationship between image attributes and emotions in the project images (H1). This study collected data from one of the most popular crowdfunding platforms, Kickstarter.com, to empirically test the relationships between image emotions and the performance of a crowdfunding project. Note that our study finds its foundation in charity fundraising (Baberini et al., 2015; Liang et al., 2016; Small & Verrochi 2009), which is close to projects in the public benefit category on Kickstarter.com.

3.1 Scope

This study focuses on how crowdfunding projects attract and seek funding from potential backers who browse the crowdfunding platform via categories or by searching keywords on the listing page. This study chooses project images on the listing pages as our main focus. A listing page (See the example in Figure 1), serving as a portal, is the first and central place where viewers can find potential crowdfunding projects that may align with their interests. This listing page displays simple snapshots of projects, including information from fundraisers such as project images and short textual descriptions, as well as platform provided data such as the amount raised so far, how many backers have supported the project, and when the fundraising timeline ends. Backers do not see project details until they click to enter the individual project page. Therefore, the listing page serves as an important gateway between backers and the actual pledge, and as such is worth researching. Prior studies provide strong evidence that images displayed and listed on the search result page significantly relate to the click-through rate (CTR) in the e-commerce context (Goswami et al., 2011). Similarly, project images on the listing page of a crowdfunding platform are also likely to play a dominant role in crowdfunding success. It is worth mentioning that this study does not limit our study to information presented on the listing pages. Important information displayed on the project pages is also included as control variables (see details in our model).

3.2 Data Collection

This study has collected the project data in the public benefit category on Kickstarter.com during the research period. Public benefit is a preset category that can be found on Kickstarter's Explore page (Appendix A). This study first uses a web scraper to parse and extract project images, textual descriptions, and several other project-related attributes, such as the number of backers, the preset fundraising goal, and the raised amount. This study obtains a total of 842 projects available in that category by August 2017. The last two projects were removed because they were still ongoing and not yet closed at the end of collection time. Thus, 840 projects remain in our final dataset.

3.3 Emotions of Project Images

This study adopts the well-defined set of emotions from Machajdik and Hanbury (2010), which includes amusement, anger, awe, contentment, disgust, excitement, fear, and sadness. Machajdik and Hanbury (2010) developed this set of emotions based on the work of Mikels et al. (2005), whose emotion set is developed in the context of the study of image and emotion. On top of that, this emotion set also considers the basic emotion set proposed by (Ekman et al., 1987). In this work, this study builds a deep neural network-based model to predict emotions for project images. The training dataset is from a set of human-labeled images created by You et al. (2016) (Figure B1 in Appendix B). The architecture of our prediction framework is shown in Figure 4. Our model differs from most previous models (Machajdik and Hanbury 2010), this study primarily uses medium-level features, instead of low-level features (i.e., color, saturation, hue). The medium-level features include adjective noun pairs (ANP) and tags/content (i.e., objects contained in images) extracted via Google Vision API. Many image cognition services are proved, such as Amazon AWS Rekognition, Google Vision, IBM Watson, and Microsoft Azure Computer Vision. Comparing the performance of above image recognition services, reports show that Google Vision provides the highest accuracy in tagging content of images, and the tags generated Google Vision is more similar to the tags generated by human (Ali, 2019; Enge, 2019).

This study evaluates and compares our model using 10-fold cross validation. Results indicate that our model achieves better accuracy than the baselines in You et al., (2016). This study then applies this learned optimal classifier to our crowdfunding project images to predict their emotions.

For each of the eight emotions for every project image, the classifier provides a degree ranging from 0 to 100 (the higher the degree, the more likely the image is associated with that emotion). Note that our emotion measurements are not derived from any lower-level image attributes (i.e., color, saturation, hue) that will be used as variables in our later model, but solely from semantics of images (i.e., objects, ANPs). Therefore, our measurements of image attributes are independent of the emotion measurements, making this study more rigorous. This study presents the details of our emotion detection via deep learning in Appendix B.

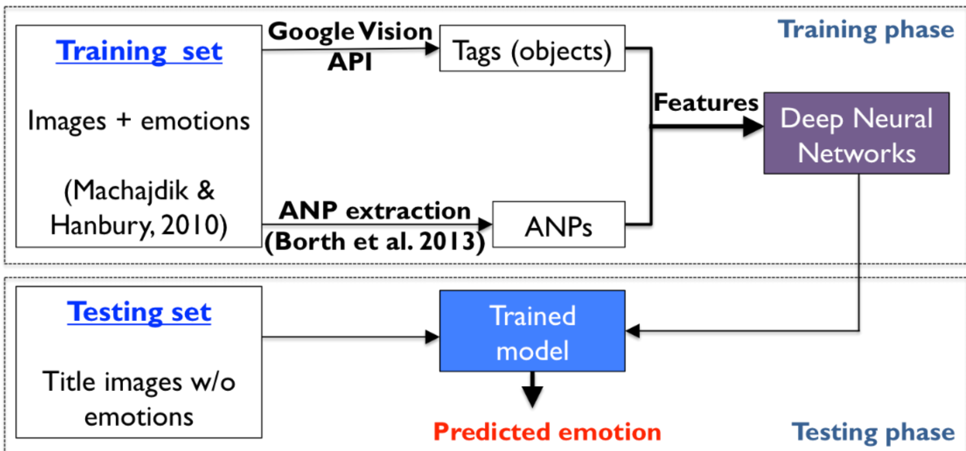


Figure 4 Architecture of our deep neural network-based image emotion classifier.

Figure 5 shows a few examples of predicted emotions of images in our project image dataset. It is worth mentioning that our deep neural network-based model can detect emotion even without the presence of a human in the image. For instance, the model is able to detect the awe emotion in the image of a collapsed building in Figure 5.

	
Amusement: 0.17 Anger: 0.04 **Awe: 98.12** Contentment: 0.67 Disgust: 0.16 Excitement: 0.1 Fear: 0.49 Sadness: 0.26	Amusement: 0.71 Anger: 1.15 Awe: 0.53 **Contentment: 91.46** Disgust: 0.23 Excitement: 2.3 Fear: 1.30 Sadness: 2.31
	
Amusement: 0.06 Anger: 9.66 Awe: 0.14 Contentment: 3.13 Disgust: 1.2 Excitement: 0.31 Fear: 2.12 **Sadness: 83.38**	Amusement: 0.35 Anger: 3.94 Awe: 0.24 Contentment: 0.22 Disgust: 13.29 Excitement: 1.64 **Fear: 66.81** Sadness: 12.06

Figure 5 Examples of project images from Kickstarter.com and their predicted emotions.

3.4 Metrics for Crowdfunding Performance

There exist many ways to measure the success of a crowdfunding project. This study chooses three metrics that are most commonly used in crowdfunding literature and easy to interpret as our dependent variables.

- **Number of backers:** The number of individuals (registered accounts on the platform, to be precise) who contribute financial support to the crowdfunding project.
- **Amount raised:** The total amount of money, in the U.S. dollar, raised by the finished crowdfunding project.

- **Percentage of goal achieved**: Every crowdfunding project has a preset funding target. Projects can achieve various percentages of this target within the valid fundraising period. Some projects may raise up to 10 or more times of the target amount, while others may fail to reach 100%. Thus, the percentage of fundraising goal achieved is a good performance index to measure the success of a crowdfunding project.

3.5 Control Variables

In our model, this study controls various variables, which are listed below, along with a brief rationale for each.

- **Emotions of text descriptions:** Emotions in textual descriptions can also be important in a crowdfunding project and may affect a backer's decision and eventually, the project performance. To investigate whether observing a similar effect of text emotions on crowdfunding project performance, this study controls for text emotions in our analysis. Since the image emotion is the focus, text emotions are treated as control variables. The text emotion set this study uses is chosen by a standard of high reliability and is processed by LIWC software (Pennebaker et al., 2015), which this study employs to analyze the emotions of text descriptions for each crowdfunding project. The LIWC tool has been widely adopted by researchers due to its reliability and validity for measuring psychological components, such as emotion, from text content (Bantum & Owen, 2009; Yin et al., 2014). It measures the degree of anxiety, anger, and sadness in text content using a built-in dictionary where, for each emotion, there is a set of related words that are identified by a series of well-designed psychological and social studies (Pennebaker et al., 2015). LIWC calculates emotion scores based on the frequency of emotion words in its dictionary that occur in the text.

- **Project popularity:** Some crowdfunding projects are very popular while others are less so, due to the following facts: 1) when choosing a crowdfunding project, individuals have different topic preferences; and 2) herding behavior suggests that people are more likely to follow crowds. To control the variation in the attractiveness of topics, this study includes project popularity as a control variable in our model. The project popularity is calculated via a simple procedure. This study first extracts the ten most frequent nouns in text descriptions across all crowdfunding projects in our dataset, as shown in the word cloud in Figure 6. Then this study uses Google Trends to obtain the relative popularity score for each word. A final popularity score for each project is computed based on popularity scores of the top ten words appearing in the description.

Figure 6 Word cloud of top popular terms in crowdfunding project descriptions.

- **Preset goal:** The preset funding target for each crowdfunding project.
- **Duration:** The length of time that the crowdfunding campaign lasts.
- **Length of full text description:** The number of words in the full text description.
- **Number of images in project description:** The number of embedded images in the project.
- **Number of videos in project description:** Video is not necessary for Kickstarter projects; some projects do not even include any video on their fundraising page (in our dataset, this is 22.7% of projects). However, the inclusion or non-inclusion of videos may impact

project performance. Thus, this study includes the number of videos embedded in the project description as another control variable.

This study summarizes the descriptive statistics of our data regarding variables in Table 2.

Table 2 Descriptive statistics of variables.

Type of Variable	Variable	Attributes	Min	Max	Mean	SD
Dependent Variable	Outcome of projects	# of Backers	0	94770	319.02	637.53
		Amount ($)	0	815601.5	27671.69	59810.14
		Percentage (%)	0	3534.61	141.795	240.064
Independent Variable	Emotion in project images	Amusement	0.022	85.822	11.930	15.584
		Awe	0.015	98.117	7.362	17.821
		Contentment	0.030	91.264	6.914	11.456
		Disgust	0.023	98.181	11.802	18.192
		Excitement	0.048	95.099	16.182	17.792
		Fear	0.187	68.037	11.332	11.015
		Sadness	0.046	83.381	10.185	11.583
Control Variable	Emotion in text description	Anxiety	0	9.09	0.082	0.650
		Anger	0	9.38	0.186	0.979
		Sadness	0	10	0.202	1.030
	Goal	Preset Goal ($)	15	1500000	27600.91	69145.689
	Project popularity	Degree of Interest	0	213.027	32.710	38.743
	Length of full description	Number of words in the full textual description	73	4565	946	632.19
	Number of Images	Number of images in full description	0	72	7.593	10.19
	Number of Videos	Number of videos in full description	0	7	0.848	0.643
	Duration	How long the campaign lasts	5	91	33.48	10.99
N	840	Successful	670			
		Failed	170			

3.6 Hypothesis 2 Testing: Empirical Results

This study performs a separate empirical analysis for each dependent variable, the number of backers, the amount of raised, and the percentage of fundraising goal achieved, respectively. For independent variables, this study uses image emotions. Note that anger[2] is removed from the model due to high multicollinearity. All results are summarized in Table 3.

Table 3 Effects of image emotions on crowdfunding performance

IV \ DV		Backers (#)	Amount (1000 USD)	Percentage of goal achieved	Backers (#)	Amount (1000 USD)	Percentage of goal achieved
Emotions in Project Images	Amusement	-0.928	-0.161	0.152			
	Awe	-1.126	0.101	-0.214			
	Contentment	**4.295****	**0.319***	**1.987****			
	Disgust	-1.595	-0.11	-0.130			
	Excitement	-0.366	-0.01	-0.355			
	Fear	**-4.541****	-0.267	-0.973			
	Sadness	**6.789*****	**0.391****	0.753			
Control Variables	Anxiety in text description	0.912	0.256	1.866	-2.335	0.074	0.737
	Anger in text description	15.075	0.767	-6.405	14.644	1.068	-5.362
	Sadness in text description	**38.406****	0.722	7.858	**35.447***	0.428	7.296
	Preset Goal	**0.002*****	**0.000*****	**0.000****	**0.002*****	**0.000*****	**0.000****
	Project popularity	**0.904***	0.03	0.107	0.770	0.022	0.068
	Length of full description	**0.180*****	**0.011*****	0.006	**0.176*****	**0.011*****	0.006
	Number of Images	**7.517*****	**1.116*****	**2.554*****	**7.82*****	**1.132*****	**2.529*****
	Number of Videos	**181.987*****	**18.382*****	**40.493*****	**189.996*****	**18.964*****	**44.166*****
	Duration	0.206	0.077	-1.230	0.109	0.082	-1.227
Constant		**-172.289***	**-19.323****	**123.556*****	**-166.364****	**-19.260*****	**125.315*****
R^2		23.7%	28.9%	5.8%	20.6%	27.3%	4.4%

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

The results in Table 3 suggest that:

[2] For every project image, the output of the deep neural network is a distribution on 8 different emotions.

1) Contentment has a statistically significant positive effect on all three outcome metrics. If the contentment increases by 1 unit, the average number of backers is expected to increase by approximately four people, the average amount raised is expected to increase by $319, and the average percentage of goal achieved is expected to increase by 1.99%. A similar correlation between contentment and donation has been shown in Krishna (2011), where authors find that people who donate are more contented.
2) Sadness exhibited in images is significantly associated with the number of backer and the amount raised with positive coefficients. This finding is consistent with the previous literature, where the emotion of sadness has a positive effect on the success of fundraising (Baberini et al., 2015; Small & Verrochi, 2009). If the sadness increases by 1 unit, the average number of backers is expected to increase by approximately seven people, and the average amount raised is expected to increase by $391.
3) Fear in project images leads to a lower number of backers. If fear increases by 1 unit, the average number of backers is expected to decrease by about five people. This funding aligns with studies in the medical literature showing that fear is a critical factor that inhibits people from donating their blood (France & France, 2018). Fear may be associated with distrust, which makes people reluctant to accept new technology (Hsiao, 2003). Consistent with the literature, this study does observe that fear makes backers balk. According to the above empirical results, H2 is supported.

On top of the above results, we can also see that the text emotions are included in the model. Although the estimated coefficients of text emotion are not comparable to image emotion's coefficients, we still can see that only sadness textual emotion is significant for increasing number of backers. On the other hand, three different image emotions are significantly related to the three outcome metrics. This might infer that image emotions can provide more insights to predict the outcomes of crowdfunding projects.

3.6.1 Robustness Checks

As a robustness check, this study builds alternative models that use image attributes to explain crowdfunding outcomes directly (Appendix C). To see how robust our proposed model is, this study compares to these alternative models concerning several standard metrics, such as adj. R^2, MAE, and RMSE. Results in Table 4 show that our proposed model can achieve fairly consistent performance with alternative models in terms of model fitness.

Table 4 Model comparisons on MAE, RMSE, and adj. R^2 for different dependent variables.

Dependent Variables	Number of backers			Amount raised			Percentage of goal achieved		
	MAE	RMSE	Adj. R^2	MAE	RMSE	Adj. R^2	MAE	RMSE	Adj. R^2
Proposed	262.63	**556.74**	**22.4%**	**22.41**	50.42	**27.6%**	**82.57**	**223.26**	**3.8%**
Alternative	**254.10**	563.20	19.8%	22.74	**50.26**	27.4%	83.38	232.75	3.2%

This study also conducts split-sample tests by randomly selecting the 75% of 840 crowdfunding projects (i.e., 630 projects) to fit the model. This operation is repeated three times and all give us results that are consistent with the full samples (Appendix D), which shows the robustness and consistency of our model.

Moreover, various text analysis tools can be employed to analyze the emotion in the text content. Thus, this study also uses IBM Tone Analyzer and LIWC with additional positive and negative emotions to replace current text emotions results from LIWC. Even with different text analysis tools, the effects of image emotions are still consistent with our current results in Table 3. The results are also presented in Appendix D.

3.7 Hypotheses 1 testing: Image Emotion Manipulation

The results from empirical analysis support the idea that image emotions in project images do affect the performance of the crowdfunding projects based on the SOR model. They are consistent with the previous findings that emotions are key triggers for the success of fundraising projects (Baberini et al., 2015; Small & Verrochi, 2009). However, the question of how to explain emotions in images remains. To further comprehensively utilize the SOR model to reveal backers' behavior and to provide additional practical guides, this study leverages the image attribute variables proposed by Wang et al. (2013) and Zhang et al. (2017) to investigate what crucial image attributes can explain emotions to test the H1.

3.7.1 Image Attributes

This study follows the instruments used in Wang et al. (2013) and Zhang et al. (2017), which provide a guideline for researchers and practitioners to judge image aesthetics. This study focuses on the objective attributes of images for the following two reasons: 1) these are standard attributes to rate image emotions, although meeting these standards does not mean that the image meets an individual viewer's aesthetic expectation and vice versa; and 2) these objective attributes can be very informative in explaining image emotions (as demonstrated by the findings in 5.3). Following Wang et al. (2013) and Zhang et al. (2017), this study, therefore, proposes the following four major components of image attributes, summarized in Table 5.

Table 5 Image attributes and definitions.

Component	Attributes	Operational Definitions
Composition	Diagonal Dominance	Distance between main figure and two diagonal lines (lower is better)
	Symmetry	Is main figure distributed evenly on left and right (lower is better)
	Color Balance	Is color distributed evenly on left and right
	Rule of Thirds	Distance between main figure and two equally spaced vertical lines (lower is better)

Table5 Image attributes and definitions. (Cont.)

Component	Attributes	Operational Definitions
Color	Warm Hue	Pixels that color hue are smaller than 30 or bigger than 110 under HSV color space
	Saturation	Saturation
	Brightness	Brightness
	Contrast of Brightness	Contrast of Brightness
	Image Clarity	Pixels that color brightness are between 0.7 and 1 on a 0-1 scale under HSV color space (High collinearity with brightness)
Main element -background relationship	Size Difference	Proportion of main element in whole image
	Color Difference	Color difference between main element and background
	Texture Difference	Texture difference between main element and background
Image content	Text	Are there any textual contents in the images?
	Human	Are there any human beings in the images?
	Animal	Are there any living creatures other than humans in the images?

3.7.1.1 Composition

- **Diagonal dominance:** Leading lines in photos can guide a viewer's focal point, and the two diagonal lines are the longest leading lines in a photo (Figure 7). If this study positions the main elements of a photo along the two diagonal lines, it will lead the viewer's focal points through the whole photo and lead to a feeling of spaciousness (Zhang et al., 2017). To operationally define the diagonal dominance, this study first employs the saliency algorithm (Yang et al., 2013) to identify the main element of each project image. Then this study calculates the shortest distance between the main figure and the two diagonal lines, and the shorter distance represents better diagonal dominance since the main figure is placed close to any diagonal line. The original score measures diagonal dominance in a reversed order; that is, the lower the score, the better the diagonal dominance. For easy interpretation, this study subtracts each diagonal dominance score from zero to make the score positively measure diagonal dominance.

- **Symmetry:** If the main elements of a photo are distributed balanced across the vertical center line of the photo, symmetric distribution of the main elements can give rise to a feeling of order and provide aesthetic pleasure (Zhang et al., 2017). To operationally define "symmetry," this study first employs the saliency algorithm (Yang et al., 2013) to identify the main element for each project image. Then this study calculates the distribution of the main figure across the vertical center line and subtracts the bigger portion by the smaller portion, and since the main figure is symmetrically distributed across the central line, a smaller difference between the two portions represents better symmetry. The original score measures symmetry in a reversed order. For easy interpretation, this study subtracts each symmetry score from zero to make the score positively measure symmetry.
- **Color balance:** Like symmetry attributes, balance color across the vertical central line could also provide a feeling of order and aesthetic pleasure. To operationally define the color balance, this study divides each project image by the central vertical line, calculate the Euclidean color distance for each mirrored pixel on the left portion and the right portion, then calculate the mean for all Euclidean color distances for each pair of mirrored pixels. The original score measures color balance in a reverse order since a lower score represents less difference of color across the vertical central line. But for easy interpretation, this study subtracts each color balance score from zero to make the score positively measure color balance.
- **Rule of thirds:** As illustrated in Figure 7, if we want to divide any photo into a three-by-three grid with nine even portions, we will need two vertical lines and two horizontal lines, and there will be four intersections of those lines on the photo. Photographers believe that if we place main element on the intersection or on the lines, the photo will be aesthetically

pleasing by using an unbalanced composition to move the viewer's focal point (Zhang et al., 2017). Compared to the symmetry attribute, the rule of thirds uses an unbalanced composition to present the feeling of something unusual. To operationally define the rule of thirds, this study employs the saliency algorithm to find the main element for each project image (Yang et al., 2013). Then this study calculates distance between the main element and the four intersections, and by the rule of thirds, the lower the distance, the better the score. However, for easy interpretation, this study subtracts each score from zero to make the higher score now be the better one.

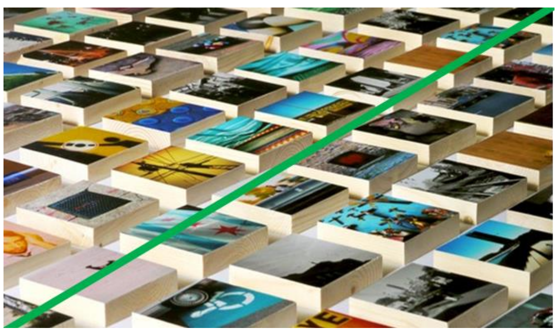
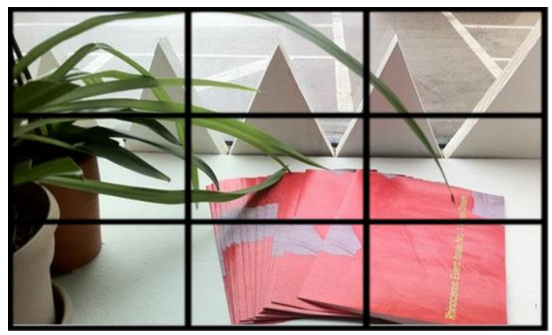
	
Diagonal dominance	Rule of thirds

Figure 7 Examples of image attributes: composition.

3.7.1.2 Color

Color can express different feelings to the viewers. For instance, red can be perceived as a hazardous color (Braun et al., 1994). Green may be taken as trustworthy in certain cultures (Aslam, 2006). Thus, based on the guidelines by Wang et al. (2013) and Zhang et al. (2017), this study includes some color-related attributes such as warm hue, saturation, brightness, contrast of brightness, and image clarity to measure image emotions in this study.

- **Warm hue:** This is defined as the proportion of pixels in the overall pixel count with warm color hues. To find this number, this study first converts to the HSV color space. By the definition of Wang et al. (2013), if a pixel's hue is smaller 30 or bigger than 110 in the HSV color space, this study could consider that pixel to have a warm color hue.

- **Saturation and brightness:** Past research provides evidence that different saturation and brightness settings of color could affect viewers' emotional feelings regarding color (Valdez & Mehrabian, 1994). To calculate the brightness and saturation, this study calculates the average brightness and saturation across every pixel in each project image using the Python-based PIL package.
- **Contrast of brightness:** Even distribution of brightness across all pixels in a photo could make the content clear to viewers. This study calculates the contrast of brightness by the standard deviation across the brightness of all pixels. The higher contrast of brightness indicates a more even distribution of brightness in project images.
- **Image clarity:** Based on the definition of Wang et al. (2013), this study measures image clarity for each pixel for which the brightness score falls between 0.7 and 1 on a 0-1 scale under the HSV color space. If more pixels fall into this range, it means higher intensity of hues, which makes the image clearer.

3.7.1.3 Main element-background relationship

The differences between the main element and the background, which are here measured by size, color, and texture, will make the main element stand out more in the image.

- **Size difference:** This study first employs the saliency algorithm (Yang et al., 2013) to find the main element of each project image. Then this study computes the proportion of the main element in each photo. A higher size difference score suggests that the main figure occupies a larger proportion of the project image.

- **Color difference:** Again, the main element of each project image is detected by the saliency algorithm (Yang et al., 2013). Euclidean distance is calculated by the average color of the main figure and the average color of the background. A higher score of color difference represents a greater color difference between the main figure and the background.
- **Texture difference:** To calculate texture difference, this study first employs the edge detection algorithm on the main figure and on the background of the project images, respectively. Then this study computes the edge density for the main figure and for the background to define the different texture, and this study further subtracts the edge density of the main figure from that of the background to give us the texture difference score. A higher texture difference score indicates a higher texture difference between the main figure and the background.

3.7.1.4 Image content

Previous research has shown that image content affects the outcome of e-commerce (Bland et al., 2007). Thus, this study introduces the presence of people or other living creatures in project images as an image attribute. Additionally, since evidence has shown that textual content affects crowdfunding success (Yuan et al., 2016), this study introduces the presence of text in project images as another image attribute. These attributes are manually extracted by two domain experts.

3.7.2 Data

This study uses Python and a widely used image processing package, Pillow, to measure the above attributes of all 840 project images in our data sample. The descriptive statistics are listed in Table 6.

Table 6 Descriptive statistics of image attributes.

Attributes	Min	Max	Mean	Std.
Diagonal dominance	-30.458	-0.148	-6.624	5.733
Symmetry	-3246	0	-468.36	405.894
Color balance	-296.143	-2.484	-90.672	36.807
Rule of thirds	-45.267	-0.781	-21.841	7.577
Warm hue	0.207	100	70.372	25.053
Saturation	0	0.9652	0.279	0.176
Brightness	1.401	254.465	140.495	49.509
Contrast of brightness	17.8546	117.946	60.118	15.545
Image clarity	0	99.963	34.814	26.785
Size difference	0.3107	69.333	5.197	5.534
Color difference	1.8642	408.3868	122.340	85.569
Texture difference	0	0.115	0.092	0.070
Text in image	0	1	0.488	0.500
Human in image	0	1	0.35	0.478
Animal in image	0	1	0.05	0.218
N	840			

3.7.3 Results

This study uses image attributes to explain the possible emotions of project images and find that overall, some attributes can explain emotions well, as shown in Table 7.

Table 7 Results of using image attributes to explain image emotions.

DV / IV	Amusement	Awe	Contentment	Disgust	Excitement	Fear	Sadness
Diagonal dominance	-0.010	0.151	**-0196*****	0.014	-0.009	0.052	0.046
Symmetry	0.000	0.001	0.000	0.000	0.001	**-0.003****	-0.002
Color balance	-0.016	**0.081*****	0.006	**-0.079*****	-0.004	**0.028****	-0.001
Rule of thirds	0.103	**-0.175***	0.093	0.061	-0.035	-0.092	0.000
Warm hue	0.003	-0.028	**-0.029***	-0.032	**0.076*****	**-0.029***	**-0.036****
Saturation	**5.783***	0.754	1.852	**7.216***	**7.279***	**-8.848*****	**-12.848*****
Brightness	-0.005	**-0.026****	0.012	**0.031****	-0.006	**-0.033*****	-0.012
Contrast of brightness	-0.007	**0.113****	-0.011	**-0.098***	0.036	0.028	**-0.068****
Size difference	-0.077	-0.180	0.125	0.081	**-0.052**	**-0.147***	0.068
Color difference	**-0.015****	**-0.020****	-0.003	0.001	-0.008	0.002	0.003

Table7 Results of using image attributes to explain image emotions. (Cont.)

DV / IV	Amusement	Awe	Contentment	Disgust	Excitement	Fear	Sadness
Texture difference	1.910	-11.995	6.101	-7.047	9.768	9.089	1.912
Text in image	**-4.252*** **	**-6.641*** **	**-2.802*** **	-2.017	-0.794	0.621	-1.022
Human in image	**-2.167***	**-5.551*** **	0.382	**-7.970*** **	**12.243*** **	**-2.012** **	-0.971
Animal in image	**-7.430 *** **	-4.395	**14.445*** **	-3.823	**-5.859** **	0.587	1.971
Constant	**17.409*** **	**20.783*** **	**8.400*** **	**12.139** **	**3.427**	**18.495*** **	**21.345*** **
R^2	5.6%	10.3%	11%	7.6%	14.6%	5%	4.5%

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

This study will explain these associations in more depth in the discussion section, but this study notes that, for instance, contentment is positively associated with having animals in project images, but diagonal dominance, warm hue, and text in project images are negatively associated with contentment. Color balance is positively associated with fear, but symmetry, warm hue, saturation, brightness, size difference, and humans in project images are all negatively associated with fear. Warm hue, saturation, and contrast of brightness are negatively associated with sadness. Note that image clarity is deleted due to multicollinearity. The results show that image attributes are related to emotions of project images. Thus, H1 is supported.

4. Study 2: Experiment of Designs attributes on Image Emotions

To further verify the empirical results from section 3. We conduct experiments with image attribute manipulations.

4.1 Experiment Design

This study creates a crowdfunding project that raises funds to trap, neuter, and return stray cats humanely. This made-up project contains a project image of the same stray cat and a short textual description of the project. The objective of this experiment is to test the effect of emotions in project images on the emotions of crowdfunding projects. The project image and description of the made-up crowdfunding project are attached in Appendix E. The experiment is run in the form of a survey form where this study uses Qualtrics to establish and collect our experiment data.

There are two groups of experiments. From the results of our empirical analysis, this study chooses warm hue and brightness to manipulate since they are easier to modify. Thus, this study manipulates warm hue (High vs. Low) in the first group and brightness (High and Low) in the second group. The reason this study chooses to manipulate only one attribute in each group is that manipulating any color attribute in images will affect other color attributes. Although this study tries our best to manipulate only the warm hue and brightness, the slightly changing in other attributes is inevitable. Thus, to minimize the effects of this issue, this study only manipulates one color attribute in one group. Each subject is assigned to view one of the project images and the associated same textual description. After reading through the made-up crowdfunding project, subjects are asked to rate their emotional response in a 7-points Likert scale (Machajdik & Hanbury, 2010) and if they are willing to pledge for such a crowdfunding project. This experiment is created online and distributed through online Amazon Mechanical Turk. To ensure all participants are proficient in English and maintain the quality of the sample, this study required all participants on the Mturks need to be located in United States, have historical approval rate over 85%, and obtain mater qualification.

Group 1:

In group 1, this study manipulates warm hue in the stray cat's image. This study uses the pillow, a python imaging library, to change the warm hue in the image. Then this study uses our algorithm to verify the manipulation on warm hue is successful. The cat's image with high or low warm hue are listed in following Table:

Table 8 Manipulation on Warm Hue

Manipulation	Warm hue: High	Warm hue: Low
Image		
Warm hue value measured by our algorithm	99.0971	0.3169

Group 2:

In group 2, brightness is manipulated by pillow as well. Then this study uses our algorithm to measure the brightness and check the manipulation is successful. The cat's image with high or low brightness are listed in following Table:

Table 9 Manipulation on Brightness

Manipulation	Brightness: High	Brightness: Low
Image		
Brightness value measured by our algorithm	196.4323	77.5801

4.2 Data

This study collected data in October of 2018 on the Mturks. In total, 170 samples are were collected. After removing invalid samples such as participants who never browse any crowdfunding platform before, 131 samples remain. The descriptive statistics are shown in Table 10.

Table 10. Descriptive statistics.

Variable	Min	Max	Mean	SD
Pledge Intention	0	1	0.67	0.471
Amusement	1	7	2.17	1.458
Awe	1	7	2.74	1.577
Anger	1	7	2.77	1.671
Contentment	1	7	2.77	1.486
Disgust	1	7	2.37	1.525
Excitement	1	7	2.46	1.432
Fear	1	6	2.14	1.335
Sadness	1	7	4.50	1.614
Age	24	65	39.46	9.621
Education	1	8	4.22	1.279
Income	1	12	5.76	2.844
Gender	0	1	0.57	0.497
N	131			

The descriptive statistics are provided for each treatment in our experiment in following table.

Table 11 Descriptive statistics of the data from the lab experiment.

Group	Warm hue				Brightness			
Manipulation	High		Low		High		Low	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Pledge Intention	0.60	0.498	0.78	0.420	0.71	0.462	0.60	0.497
Amusement	2.47	1.502	2.25	1.368	1.94	1.347	2.06	1.608
Awe	2.93	1.530	2.88	1.661	2.79	1.572	2.40	1.557
Anger	2.47	1.332	3.06	1.625	2.50	1.692	3.03	1.917
Contentment	3.17	1.367	3.03	1.555	2.47	1.261	2.49	1.652
Disgust	2.17	1.117	2.59	1.563	2.24	1.558	2.46	1.771
Excitement	2.67	1.269	2.59	1.388	2.38	1.498	2.23	1.555
Fear	1.77	0.817	2.75	1.566	2.12	1.387	1.91	1.269
Sadness	4.13	1.592	4.78	1.408	4.53	1.600	4.54	1.821
Age	37.73	7.939	40.09	11.035	40.97	10.179	38.89	9.106
Education	4.10	1.348	4.13	1.338	4.47	1.398	4.17	1.043
Income	5.73	2.840	5.78	2.837	6.12	3.409	5.43	2.279
Gender	0.53	0.507	0.47	0.507	0.62	0.493	0.66	0.482
N	30		32		34		35	

4.3 Experiment Results

This study employs independent samples t-test to test if changes in image attribute, warm hue and brightness, can make differences to the emotional feeling of participants. The results (t value) are listed in the Table 12.

Table 12 Results of independent samples t-test

Manipulation	Amusement	Awe	Anger	Contentment	Disgust	Excitement	Fear	Sadness
Warm hue	0.594	0.144	-1.573	0.363	-1.231	0.215	**-3.127*** **	**-1.700***
Brightness	-0.324	1.046	-1.213	-0.043	-0.552	0.418	0.636	-0.033

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

From the results in Table 12, we can see the results from the warm hue manipulation group is similar to our empirical results, and lower warm hue can bring participants the higher sadness and the higher fear. However, this study does not observe the effect of brightness manipulation in this study. The results from the warm hue manipulation group can provide experimental supports for our empirical results from Kickstarter and further verify H1.

4.4 Emotions' effect on the intention

In this section, this study pools all data points from the two groups, which contains 131 samples and perform a logistic regression to predict participants' intention to pledge by their emotional responses. Although the results from this section may not provide efficient causality inference as experiments, the results should still provide insights based on the correlation between emotions and pledge intention. The results of logistic regression are provided in the following table.

Table 13 Effect of image emotion on the pledge intention

IV \ DV		Intention
Emotions	Amusement	-0.160
	Awe	**0.589****
	Anger	0.312
	Contentment	0.066
	Disgust	-0.217
	Excitement	0.057
	Fear	0.106
	Sadness	**0.464*****
Control Variable	High_WarmHue (Dummy)	-0.370
	Low_WarmHue (Dummy)	0.338
	High_Brightness (Dummy)	-0.519
	Income	-0.136
	Education	**0.599*****
	Age	-0.019
	Gender	-0.093
Constant		**-4.003****
N		131
Cox & Snell R^2		0.284
-2 log likelihood		122.101

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

From the results from the above tables, the logistic regression indicates that sadness is effective in increasing the odds of pledge intention, which is consistent with our empirical results and verifies the H2. However, other effective emotions such as contentment and fear are lost in the analysis. This study argues that this issue might due to the design of our experiment is dominated by sadness, so the effect of other emotions might be overpowered by the sadness. This study does not elaborate the effect of awe since this study does not find the effect of awe in our empirical results.

5. Discussions

In this research, this study employs the SOR model to explore the behavior of backers on a crowdfunding platform. Overall, our results are consistent with the proposed model. In examining our research question, this study contributes to the emotion-related literature and shed light on a new way to examine the impact of emotions in project images on the performance of crowdfunding projects. This study performs both empirical analyses and lab experiments, and obtain consistent results to show that emotions do play a critical role in fundraising performance. For example, our empirical evidence from Kickstarter.com supports that emotions in project images affect crowdfunding performance in a way that can be measured by three important metrics: the number of backers, amount raised, and percentage of fundraising goal achieved. Besides, the effect of image emotions is more significant than other control variables, such as text emotions.

To draw out the practical value and managerial implications of this research, this study investigates whether and what image attributes can explain emotions in project images. This study extracts image attributes defined in existing literature (Wang et al., 2013; Zhang et al., 2017) to analyze 840 project images from Kickstarter.com. The results show significant evidence that some image attributes can explain emotions well, which is consistent with the literature, arguing that there are emotional responses to any type of artwork including photographs and paintings (Barry, 2006; Carroll, 2003; Mendelson & Papacharissi, 2007; Silvia, 2005).

Text emotions might be effective for the success of crowdfunding projects as well. So, we include text emotions as control variables in our empirical model. The results show that only textual sadness emotions are related to the outcomes of crowdfunding. Furthermore, we also try the different text emotion detection tool from IBM. The results show that only textual sadness emotion is related to only one of the outcome metrics. Thus, we infer that image emotions are more likely to catch users' attention and have stronger influences on user decisions than textual emotions.

Overall, our empirical and experimental results support the hypotheses this study proposed based on the SOR model, the image attribute (S) can affect the emotion of project images (O), and the emotion of project images (O) can further affect the crowdfunding performance. To enrich the contribution of this study, this study also conducts an additional moderation in the context. Prior literature shows that goal is a significant factor for crowdfunding performance (Mollick, 2014), so this study includes goal as a control variable in our model. To further test the possible moderation of goal, this study uses an average goal as the threshold to separate all the crowdfunding projects in our empirical dataset into higher goal projects and lower goal projects. The results show that contentment is effective in the lower goal projects, and fear and sadness are effective in the higher goal projects. The results might indicate negative emotions such as fear and sadness can better trigger backers to pledge on the higher goal projects. This might due to higher goal projects imply a greater difficulty to be funded, and negative emotions might provide a larger motivation to make humans pledge. On the other hand, for the lower goal projects, contentment plays an important role for backers since lower goal projects might imply easier fundraising so that backers can be motivated with positive emotions. The results are tabulated in Appendix F.

This study provides practical and theoretical implications in the following sections.

5.1 Practical Implications

The results of this research provide practical and actionable insights for crowdfunding seekers and platforms to improve the performance/outcome of charity-like crowdfunding projects. First, although previous studies have shown that emotions are strong factors motivating financial donation, there is still the question of how to gauge emotions. This study provides practical instruction to those who are interested in improving the performance of crowdfunding projects, particularly from the perspective of image emotions.

Our work also builds on previous studies of particular emotional states; for instance, prior studies show that sadness is a possible and powerful emotion in fundraising, and here this study presents similar findings to show that sadness can prompt people to provide financial aid to those in need (Liang et al., 2016; Small & Verrochi, 2009). Nevertheless, this study is also able to specifically point out that warm hue, saturation, and contrast of brightness are among many image attributes that are negatively related to sadness with statistical significance. In other words, decreasing values of these attributes in the project images can potentially increase the viewers' sadness, which might benefit crowdfunding performance in practice. In contrast, contentment works differently. It significantly affects crowdfunding performance, which is consistent with existing literature finding that consumers who donate are more contented (Krishna, 2011). The literature also argues that contentment plays an important role in reciprocal altruism, meaning that when one helps others, that person expects others will help them in the future (Jackson, 2002; Trivers, 1971). Thus, this study believes that when possible backers are in content, they are more likely to help others and this is possibly due to the reciprocal altruism mechanism. According to our results, images including animals correlate positively with contentment, while diagonal dominance, warm hue, and images with text all have a negative correlation with contentment. Thus, practitioners who are looking to trigger sadness and not contentment with their project images can intentionally control image attributes such as warm hue, saturation, and contrast of brightness.

Our results also suggest that practitioners should manipulate image attributes carefully. For example, some image attributes such as low saturation and low warn hue are found to be associated not only with sadness but also with fear. While sadness improves crowdfunding performance, these unexpected image emotions such as fear have a negative effect.

This study also shows the value of machine learning in processing unstructured data such as images. Practitioners such as data scientists can apply machine learning on unstructured data to extract useful features for further data analysis, thus avoiding possible response bias and noise that may arise in traditional Human Intelligence Tasks (HITs) such as surveys.

5.2 Theoretical Implications

Existing research has shown that emotion is effective in motivating charitable donation (Baberini et al., 2015; Liang et al., 2016; Small & Verrochi, 2009). Nevertheless, there is still a lack of studies that consider emotions in crowdfunding scenarios. To contribute to this line of research, this study investigates how emotions in project images affect the performance of crowdfunding projects. Our results indicate that emotions, such as sadness, contentment, and fear, are significant drivers that affect potential backers' donation to public benefit crowdfunding projects.

Emotions are considered to be important responses from viewers who express their aesthetic appreciation to the artwork (Barry, 2006; Carroll, 2003). However, due to individual heterogeneity, viewers tend to hold different subjective viewpoints toward artwork, which makes it difficult to capture and measure the characteristics and emotions of individual works of art. In this research, this study develops a deep neural network-based image emotion classifier, which not only serves our particular purpose but also shows the implication and value of deep learning techniques for emotion- and fundraising-related research. In addition, this study applies existing advanced technologies developed by Wang et al. (2013) and Zhang et al. (2017) to transform subjective measures to objective ones to measure the aesthetic features that may affect emotions.

6. Conclusions

In this research, this study performs both empirical analyses on data collected from Kickstarter.com and lab experiments on made-up data to study how emotions in project images affect the performance of crowdfunding projects. Results show that sadness is a possible and powerful emotion in fundraising, and along with contentment and fear, may significantly affect fundraising performance. With these findings, this study makes theoretical contributions to the existing stream of literature. In addition, this study demonstrates an application of advanced machine learning techniques in studying performance issues in crowdfunding, and it provides actionable and practical guidance for crowdfunding seekers and platforms. This study demonstrates how to implement new algorithms to classify intangible emotions conveyed by images and to further discover the role of emotions in crowdfunding-related research.

There are a few limitations to the present research. First, our deep learning algorithm is built upon the emotion set from a paper on emotion learning (Machajdik & Hanbury, 2010) that is also used by You et al. (2016). This study uses this emotion set because it is the most frequently used and well-developed emotion set for studies on image emotions to date. This also allows us to compare our algorithm to those used by other researchers. However, it is possible that other emotion sets may offer an even better way of categorizing emotions. Second, image emotion is detected by our trained deep neural network model, and text emotions are identified based on LIWC. Seeking better emotion classifiers is also one planned avenue for our future research. Third, the focus of this study is limited to crowdfunding projects in the public benefit category on Kickstarter.com. All models and variables this study uses in this research are based on the existing literature in the domain of charity fundraising. In our future work, this study plans to extend our research to other categories.

References

Aboubaker Ettis, S. (2017). Examining the relationships between online store atmospheric color, flow experience and consumer behavior. *Journal of Retailing and Consumer Services*, *37*, 43–55.

Ahlers, G. K. C., Cumming, D., Günther, C., & Schweizer, D. (2015). Signaling in Equity Crowdfunding. *Entrepreneurship Theory and Practice*, *39*(4), 955–980.

Ali, A. (2019, August 11). Study: Google's AI is the best when it comes to image recognition accuracy. Retrieved September 22, 2019, from https://www.digitalinformationworld.com/2019/08/amazon-google-or-ibm-microsoft-which-company-does-the-best-job-at-image-recognition.html

Allison, T. H., Davis, B. C., Short, J. C., & Webb, J. W. (2015). Crowdfunding in a Prosocial Microlending Environment: Examining the Role of Intrinsic Versus Extrinsic Cues. *Entrepreneurship Theory and Practice*, *39*(1), 53–73.

Amit, E., Algom, D., & Trope, Y. (2009). Distance-dependent processing of pictures and words. *Journal of Experimental Psychology: General*, *138*(3), 400–415.

Animesh, A., Pinsonneault, A., Yang, S.-B., & Oh, W. (2011). An Odyssey into Virtual Worlds: Exploring the Impacts of Technological and Spatial Environments on Intention to Purchase Virtual Products. *MIS Quarterly*, *35*(3), 789–810.

Aslam, M. M. (2006). Are You Selling the Right Colour? A Cross-cultural Review of Colour as a Marketing Cue. *Journal of Marketing Communications*, *12*(1), 15–30.

Baberini, M., Coleman, C.-L., Slovic, P., & Västfjäll, D. (2015). Examining the Effects of Photographic Attributes on Sympathy, Emotions, and Donation Behavior. *Visual Communication Quarterly*, *22*(2), 118–128.

Barry, A. M. (2006). Perceptual Aesthetics: Transcendent Emotion, Neurological Image. *Visual Communication Quarterly*, *13*(3), 134–151.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*, *29*(5), 585–609.

Bland, E., S Black, G., & Lawrimore, K. (2007). Risk-Reducing and Risk-Enhancing Factors Impacting Online Auction Outcomes: Empirical Evidence from eBay auctions. *Journal of Electronic Commerce Research, 8*.

Braun, C. C., Sansing, L., & Silver, N. C. (1994). The Interaction of Signal Word and Color on Warning Labels: Differences in Perceived Hazard. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *38*(14), 831–835.

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231.

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.

Carroll, N. (2003). Art and Mood: Preliminary Notes and Conjectures. *The Monist*, *86*(4), 521–555.

Chaudhuri, A., & Micu, C. C. (2014). The effect of hedonic and utilitarian verbal descriptions and pictures on willingness to try an innovation: Do emotions matter? *Journal of Marketing Communications*, *20*(5), 307–324.

Choi, J. N., Sung, S. Y., Lee, K., & Cho, D.-S. (2011). Balancing cognition and emotion: Innovation implementation as a function of cognitive appraisal and emotional reactions toward innovation. *Journal of Organizational Behavior*, *32*(1), 107–124.

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1800–1807.

Chung, S. H., Goswami, A., Lee, H., & Hu, J. (2012, August 12). *The Impact of Images on User Clicks in Product Search*. Presented at the International Workshop on Multimedia Data Mining.

Cui, X., & Lai, V. S. (2013, June 18). *E-Loyalty to Online Auction Websites: A Stimulus-Organism-Response Model*. 14. PACIS 2013 Proceedings.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *1*, 886–893 vol. 1.

Di, W., Sundaresan, N., Piramuthu, R., & Bhardwaj, A. (2014). *Is a Picture Really Worth a Thousand Words? - On the Role of Images in E-commerce*. Presented at the International Conference on Web Search and Data Mining.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., … Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*(4), 712–717.

Enge, E. (2019, July 30). Image Recognition Accuracy Study. Retrieved October 10, 2019, from https://www.perficientdigital.com/insights/our-research/image-recognition-accuracy-study

France, C. R., & France, J. L. (2018). Fear of donation-related stimuli is reported across different levels of donation experience. *Transfusion*, *58*(1), 113–120.

Gilkeson, J. H., & Reynolds, K. (2003). Determinants of internet auction success and closing price: An exploratory study. *Psychology and Marketing*, *20*(6), 537–566.

Goswami, A., Chittar, N., & Sung, C. H. (2011). *A Study on the Impact of Product Images on User Clicks for Online Shopping*. Presented at the International World Wide Web Conference 2011, Hyderabad, India.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Hobbs, J., Grigore, G., & Molesworth, M. (2016). Success in the management of crowdfunding projects in the creative industries. *Internet Research*, *26*(1), 146–166.

Hong, Y., Hu, Y., & Burtch, G. (2018). Embeddedness, Prosociality, and Social Influence: Evidence from Online Crowdfunding. *MIS Quarterly*, *42*(4), 1211–1224.

Hsiao, R.-L. (2003). Technology fears: Distrust and cultural persistence in electronic marketplace adoption. *The Journal of Strategic Information Systems*, *12*(3), 169–199.

Jackson, T. (2002). Evolutionary psychology in ecological economics: Consilience, consumption and contentment. *Ecological Economics*, *41*(2), 289–303.

Kang, L., Jiang, Q., & Tan, C.-H. (2016). Remarkable advocates: An investigation of geographic distance and social capital for crowdfunding. *Information & Management*.

Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and Predicting Image Memorability at a Large Scale. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2390–2398.

Kickstarter. (2019a). Telling your story—Kickstarter. Retrieved June 22, 2019, from https://www.kickstarter.com/help/handbook/your_story?ref=handbook_started

Kickstarter. (2019b, June 22). Kickstarter Stats—Kickstarter. Retrieved June 22, 2019, from https://www.kickstarter.com/help/stats

Kong, S., Shen, X., Lin, Z., Mech, R., & Fowlkes, C. (2016). Photo Aesthetics Ranking Network with Attributes and Content Adaptation. *ArXiv:1606.01621 [Cs]*. Retrieved from http://arxiv.org/abs/1606.01621

Krishna, A. (2011). Can supporting a cause decrease donations and happiness? The cause marketing paradox. *Journal of Consumer Psychology*, *21*(3), 338–345.

Liang, J., Chen, Z., & Lei, J. (2016). Inspire me to donate: The use of strength emotion in donation appeals. *Journal of Consumer Psychology, 26*(2), 283–288.

Lin, M., & Viswanathan, S. (2015). Home Bias in Online Investments: An Empirical Study of an Online Crowdfunding Market. *Management Science*, *62*(5), 1393–1414.

Liu, B. (2018, November 21). Which is Better—Kickstarter or Indiegogo? Retrieved June 22, 2019, from Zendure website: https://zendure.com/blogs/news/which-is-better-kickstarter-or-indiegogo

Liu, L., Dzyabura, D., & Mizik, N. (2018). *Visual Listening In: Extracting Brand Image Portrayed on Social Media* (SSRN Scholarly Paper No. ID 2978805). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2978805

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, *2*, 1150–1157 vol.2.

Luqman, A., Cao, X., Ali, A., Masood, A., & Yu, L. (2017). Empirical investigation of Facebook discontinues usage intentions based on SOR paradigm. *Computers in Human Behavior, 70*, 544–555.

Machajdik, J., & Hanbury, A. (2010). Affective Image Classification Using Features Inspired by Psychology and Art Theory. *Proceedings of the 18th ACM International Conference on Multimedia*, 83–92.

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

Mendelson, A. L., & Papacharissi, Z. (2007). Reality vs. Fiction How Defined Realness Affects Cognitive & Emotional Responses to Photographs. *Visual Communication Quarterly*, *14*(4), 231–243.

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the international affective picture system. *Behavior Research Methods*, *37*(4), 626–630.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, *29*(1), 1–16.

Ordanini, A., Miceli, L., Pizzetti, M., & A. Parasuraman. (2011). Crowd-funding: Transforming customers into investors through innovative service platforms. *Journal of Service Management*, *22*(4), 443–470.

Peck, J., & Childers, T. L. (2003). To Have and to Hold: The Influence of Haptic Information on Product Judgments. *Journal of Marketing*, *67*(2), 35–48.

Peng, C., Yeong, & Kim, Y. G. (2014). Application of the Stimuli-Organism-Response (S-O-R) Framework to Online Shopping Behavior. *Journal of Internet Commerce*, *13*.

Silvia, P. J. (2005). Emotional responses to art: From collation and arousal to cognition and emotion. *Review of General Psychology*, *9*(4), 342–357.

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv 1409.1556*.

Singh, P., Katiyar, N., & Verma, G. (2014). *Retail Shoppability: The Impact Of Store Atmospherics & Store Layout On Consumer Buying Patterns*. *3*.

Small, D. A., & Verrochi, N. M. (2009). The Face of Need: Facial Emotion Expression on Charity Advertisements. *Journal of Marketing Research*, *46*(6), 777–787.

Statista. (2019). Crowdfunding—Worldwide | Statista Market Forecast. Retrieved June 22, 2019, from Statista website: https://www.statista.com/outlook/335/100/crowdfunding/worldwide

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463.

Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, *123*(4), 394–409.

Vieira, V. A. (2013). Stimuli–organism-response framework: A meta-analytic review in the store environment. *Journal of Business Research, 66*(9), 1420–1426.

Wang, X., Jia, J., Yin, J., & Cai, L. (2013). Interpretable aesthetic features for affective image classification. *2013 IEEE International Conference on Image Processing*, 3230–3234.

Wood, S. L., & Moreau, C. P. (2006). From Fear to Loathing? How Emotion Influences the Evaluation and Early Use of Innovations. *Journal of Marketing*, *70*(3), 44–57.

Yang, C., Zhang, L., Lu, H., Ruan, X., & Yang, M. H. (2013). Saliency Detection via Graph-Based Manifold Ranking. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 3166–3173.

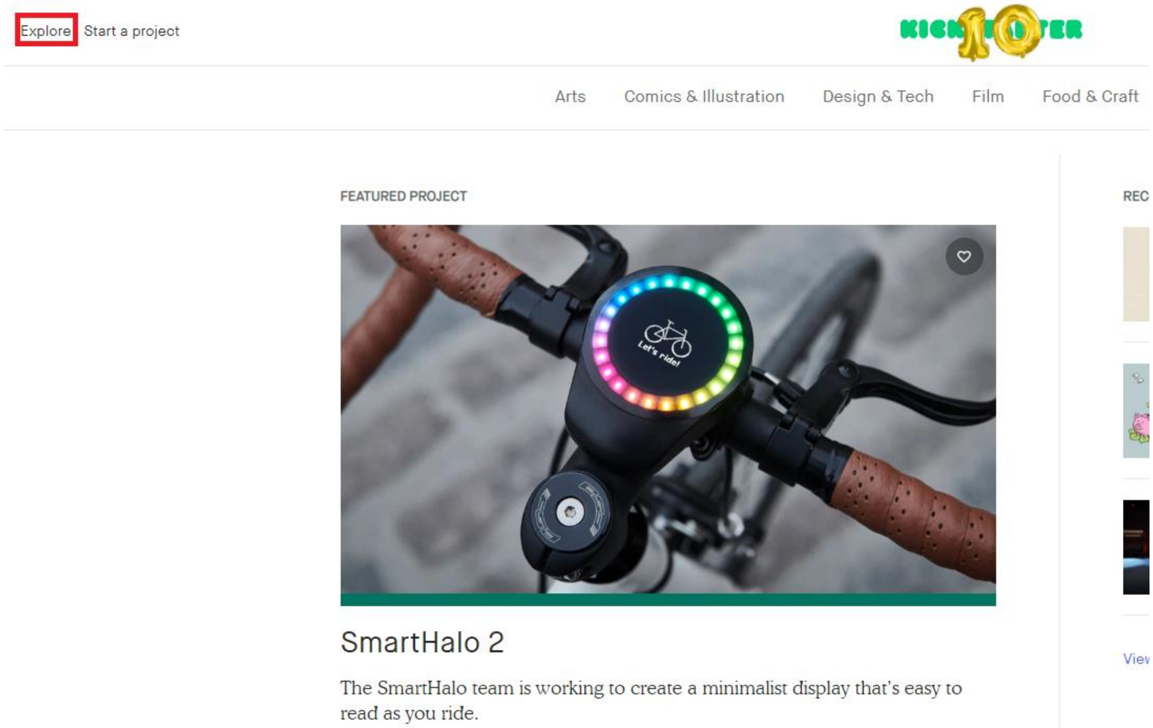
You, Q., Luo, J., Jin, H., & Yang, J. (2016). Building a Large Scale Dataset for Image Emotion Recognition: The Fine Print and the Benchmark. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 308–314. Retrieved from http://dl.acm.org/citation.cfm?id=3015812.3015857

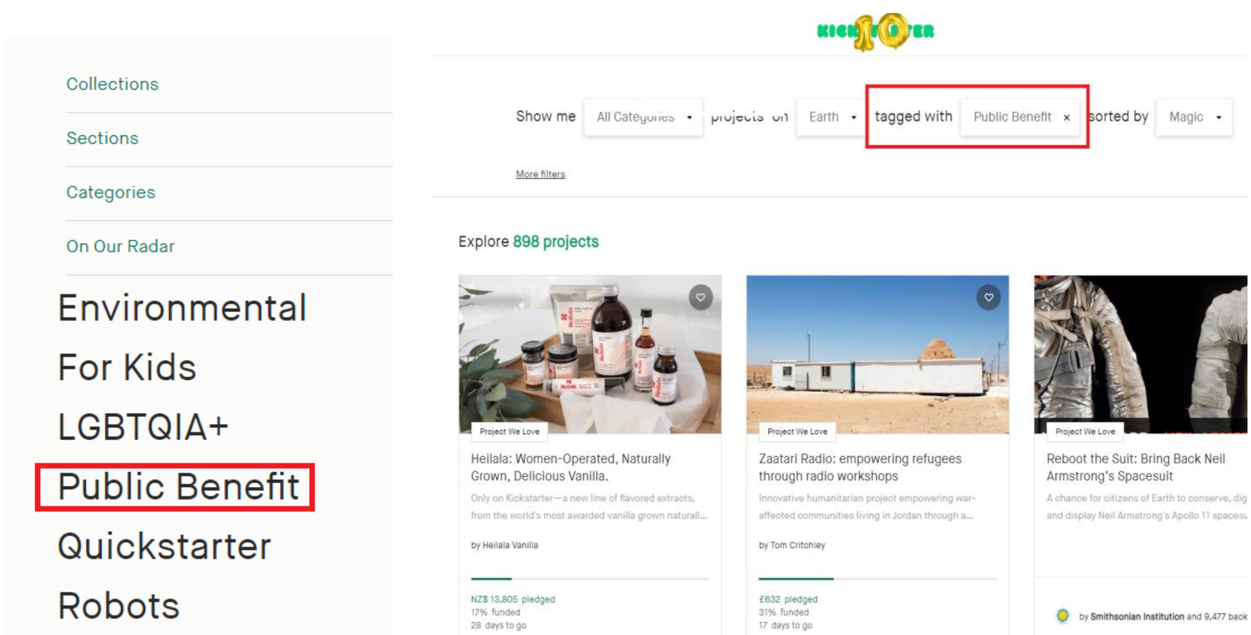
Yuan, H., Lau, R. Y. K., & Xu, W. (2016). The determinants of crowdfunding success: A semantic text analytics approach. *Decision Support Systems*, *91*, 67–76.

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2017). *How Much Is an Image Worth? Airbnb Property Demand Estimation Leveraging Large Scale Image Analytics* (SSRN Scholarly Paper No. ID 2976021). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2976021

Zheng, H., Li, D., Wu, J., & Xu, Y. (2014). The role of multidimensional social capital in crowdfunding: A comparative study in China and US. *Information & Management*, *51*(4), 488–496.

Appendix A. “Public Benefit” Category in Kickstarter.com





Appendix B. Emotion Detection via a Deep Learning Framework

1. Data

We first chose an artistic photo dataset of 807 pictures, available from the authors (Machajdik & Hanbury, 2010) on the imageemotion.org website, as our training dataset (see examples in Figure B1). Machajdik et al. downloaded the images from the deviantart.com website. In the previous literature, this dataset has proven to be as effective for the purpose of image emotion classification as the IAPS dataset, so we choose this dataset because it could be easily expanded into several thousands of pictures for model training. Next, we collect 4,959 pictures from the same deviantart.com website by searching for eight specific emotions. The results we obtained for 807 (796 after filtering) pictures were very close to the results with 4,959 pictures included, which gives us some confidence in the model stability without much dependence on the training sample size. The bigger dataset does provide much better and more stable accuracy on the 10-fold cross validation test. Finally, we use the dataset of 22,443 human-labeled pictures compiled by You et al. (2016)[3] for training.

Figure B1. Examples of training images with emotion labels.

2. Methods

- **Features**: We build a supervised learning model to predict the emotions of project images. The training data is a set of human-labeled pictures. We mainly focus on the following features: (1) instead of the low-level features that are commonly used in most image classification tasks, we use mid-level features. Specifically, we use the adjective noun pairs (ANP) because we believe that they are closer to the task of predicting emotions than are low-level features, such as color distributions, area, etc. (2) The objects shown in images

[3] http://www.cs.rochester.edu/u/qyou/deepemotion/index.html

are another important feature that tends to affect users' emotional responses. For example, a spider in a picture often leads to a fear emotion.

- **Model:** Deep neural networks have had great success in analyzing unstructured data and have been rapidly developed in many domains. We expect that they, together with medium-level features, will outperform commonly used classification approaches like random forests, support vector machines, and XGBoost. To achieve the best accuracy, we divide the ANP into adjectives and nouns, and supplement the features with objects/tags extracted for each image (up to five objects/tags) using Google Vision API. Google Vision API offers object detection service that automatically assigns objects/tags to images based on a model trained on millions of images. We believe that objects provided by Google Vision API service may supplement ANPs and enhance our predictive performance.

We first implement commonly used supervised machine learning classifiers, namely random forests, support vector machines (Breiman, 2001), and XGBoost. They are considered as the best baselines, especially among the traditional machine learning models, and they also offer easy implementation, less parameter tuning, and low computational cost. For the support vector machine, the best results were obtained by using a one-vs-rest LinearSVC classifier. To improve the performance, we designed a simple fully connected deep neural network with three hidden layers with 128 neuron units each, the activation function of ReLU, a dropout rate of 0.8, and a softmax layer as the output. It is implemented using the Python Pytorch package.

3.Results

To evaluate performance, we use 10-fold cross validation. The LinearSVC model demonstrated an overall accuracy of 18% for an eight-emotion classification task. The results suggest that the support vector machine model only works only slightly better than random guess (12.5%). Note that even when we add low-level features (like color distribution, area, perimeter, aspect ratio, extent, solidity, angle), they do not add any predictive power and therefore we do not use them in any models. The only features that are predictive are ANPs (separated into adjectives and nouns) and objects included in images, which were also separated when an object label includes two words – for example, "human action" (separately coded as "human" and "action"). The experiments show that such separation into single words increases the performance. The random forest and XGBoost methods show similar results even after trying all possible combinations of parameters (so-called "tuning"). In spite of low accuracy in eight-category classification, we do see some hope for using our features to classify emotions from the decision function (shown in Figure B2). Each decision function reports performance on a classification task of one-vs-rest for each emotion (a.k.a. binary classification). The ROC curve shows the possible trade-off between the true positive and false positive rates.

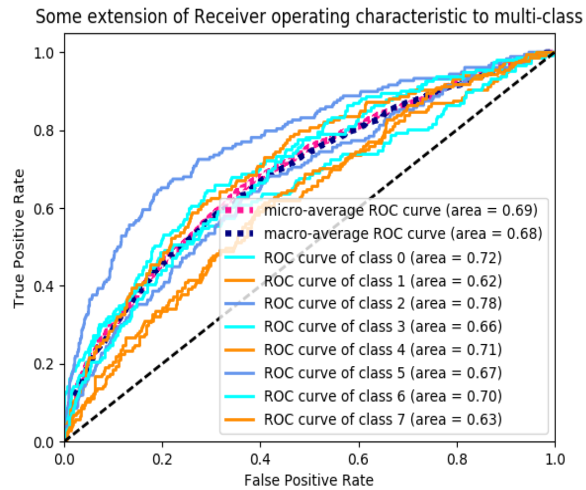


Figure B2. ROC curves for 8 categories of emotions (0 – Amusement, 1 – Anger, 2 – Awe, 3 – Contentment, 4 – Disgust, 5 – Excitement, 6 – Fear, 7 – Sadness).

For a good classification performance, we expect the area under the curve (ROC) to be close to 90% for each category. Figure B2 shows that the average ROC is 0.69, and the best performance is achieved for the class 2, which is the emotion of awe. The worst performance of 62% (as compared to 50% random guess) is seen for the first class – anger, followed by sadness with 63% area. Our results are still good in comparison with published papers using low-level features, but the 10-fold cross-validation score is relatively low, meaning that the model is not stable.

To improve accuracy and stability, we decide to employ a deep neural network classifier as mentioned above. The results obtained below are very promising. First, we train our model using a small dataset of 807 (796 with objects using Google Vision API) images, and after tuning parameters, such as number of neurons, the batch size, different activations and optimizers, and different number of epochs, the best accuracy we can achieve is 41.25% for our eight-category classification task. The high standard deviation reflects the fact that the size of training set is small and, when split into ten folds, each might by chance include more pictures of one class than of another. Instead of generating a decision function for neural networks, we show a confusion matrix (Figure B3) that shows correctly predicted classes on the main diagonal, and incorrectly predicted classes off the main diagonal. As we expected, the test set seems unbalanced.

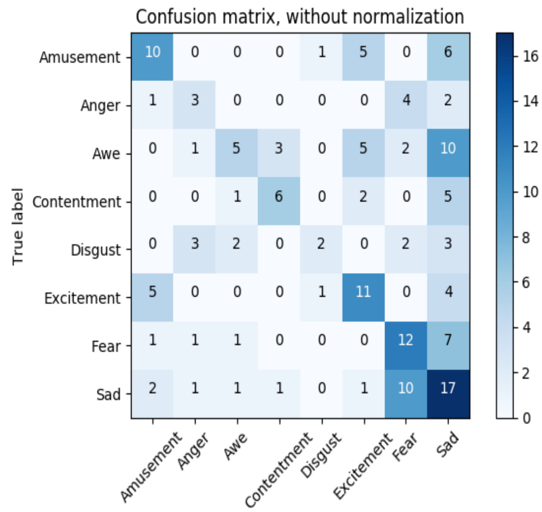


Figure B3. Confusion matrix for 796 images and 8 emotions.

For that reason we decide to use a large dataset. After removing duplicates and some pictures that were clearly noisy, 4,959 images remained in the final dataset with 530 to 694 pictures per category (average is 620). The results are also good: 36.59% overall accuracy. High cross-validation score with low standard deviation is a good sign of stability for the model performance on different sizes of training sets. Although the dataset seems a bit unbalanced, we tested models on the balanced dataset of 530 images per category for a total of 4,240 images and got a 4% lower accuracy score. The confusion matrix for the balanced dataset of 4,240 images is reported in Figure B4.

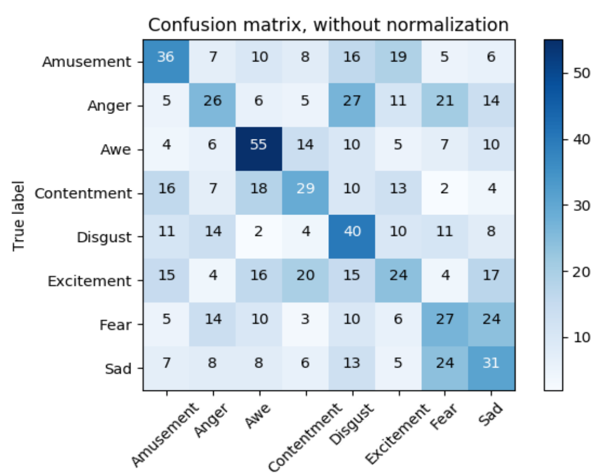


Figure B4. Confusion matrix for 4,240 images and 8 emotions.

We can clearly see that the algorithm works the best for predicting the “Awe” category (55 correct predictions out of 111) and the “Disgust” category (40 correct predictions out of 100). The confusion points, which we see by looking at the high numbers off the main diagonal, are between fear and sadness, anger and disgust, anger and fear, and amusement and excitement. Overall, the main diagonal elements are higher that the off-diagonal ones, which indicates good performance.

Appendix C. Alternative Model

Table C1. Coefficient estimations of various variables using the alternative model for different dependent variables.

IV \ DV		Backers (#)	Amount (1000 USD)	Percentage (%)
Image Attributes	Diagonal dominance	3.364	-0.264	-0.103
	Symmetry	0.017	0.005	-0.036
	Color balance	0.833	**0.155****	0.086
	Rule of thirds	**7.710****	**0.767*****	0.732
	Warm hue	-1.087	0.049	-0.084
	Saturation	2.043	3.694	-0.863
	Brightness	-0.333	-0.033	-0.073
	Contrast of brightness	0.092	0.001	**-1.379****
	Size difference	0.942	0.180	-0.222
	Color difference	0.405	0.041	0.115
	Texture difference	-199.102	-3.512	-33.326
	Text in image	26.744	-3.641	-13.123
	Human in image	-17.260	-1.846	-27.423
	Animal in image	65.410	**15.086***	10.052
Control Variables	Anxiety in text description	3.816	0.531	3.843
	Anger in text description	15.125	1.245	-4.431
	Sadness in text description	**34.548***	0.488	7.897
	Preset goal	**0.002*****	**0.000*****	**0.000****
	Project popularity	0.834	0.036	0.074
	Length of full description	**0.174*****	**0.011*****	**0.006*****
	Number of images	**7.652*****	**1.121*****	**2.501*****
	Number of videos	**185.273*****	**18.292*****	44.370
	Duration	-0.069	0.076	-1.23
Constant		190.182	9.151	**236.195*****
R^2		22.0%	29.4%	5.9%
N		840	840	840

Appendix D. Results of Randomly Split Sample

Table D1. Effect of image emotion and text emotion on crowdfunding performance
(Randomly split samples 1, N=630)

IV \ DV		Backers (#)	Amount (1000 USD)	Percentage of goal achieved
Emotion in Project Images	Amusement	-0.798	-0.180	0.590
	Awe	-0.708	-0.033	-0.543
	Contentment	**6.016*** **	**0.305***	**2.646*** **
	Disgust	-0.858	-0.075	-0.004
	Excitement	0.435	0.096	-0.325
	Fear	**-5.293** **	-0.175	-0.904
	Sadness	**9.252*** **	**0.548*** **	1.116
Control Variables	Anxiety in text description	6.627	0.387	1.947
	Anger in text description	15.021	1.114	-6.497
	Sadness in text description	**48.396** **	1.469	4.641
	Preset goal	**0.002*** **	**0.000*** **	**-0.000** **
	Project popularity	**1.451** **	0.078	0.150
	Length of full description	**0.212*** **	**0.011*** **	0.010
	Number of images	**7.798*** **	**1.061*** **	**2.952** **
	Number of videos	**205.563*** **	**18.369*** **	**31.799***
	Duration	0.022	-0.011	**-1.781***
Constant		**-278.632** **	**-20.568** **	**134.555** **
R^2		25.3%	31.89%	6.2%

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

Table D2. Effect of image emotion and text emotion on crowdfunding performance
(Randomly split samples 2, N=630)

IV \ DV		Backers (#)	Amount (1000 USD)	Percentage of goal achieved
Emotion in Project Images	Amusement	-1.707	-0.245	-0.018
	Awe	-1.453	0.105	0.006
	Contentment	**5.331****	**0.370***	**3.692*****
	Disgust	-1.521	-0.129	0.089
	Excitement	-0.930	-0.093	-0.181
	Fear	**-6.230****	-0.362	-0.645
	Sadness	**7.867*****	**0.404***	0.100
Control Variables	Anxiety in text description	-6.446	0.082	0.639
	Anger in text description	19.489	1.183	-5.979
	Sadness in text description	**59.936****	2.023	3.173
	Preset goal	**0.002*****	**0.000*****	**-0.000***
	Project popularity	**1.194***	0.040	-0.233
	Length of full description	**0.179*****	**0.010****	0.010
	Number of images	**8.316*****	**1.312*****	**3.374*****
	Number of videos	**184.114*****	**18.711*****	**43.495*****
	Duration	-0.190	0.050	**-1.412***
Constant		-148.133	-15.371	**116.852****
R^2		24.5%	29.5%	8.7%

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

Table D3. Effect of image emotion and text emotion on crowdfunding performance
(Randomly split samples 3, N=630)

IV \ DV		Backers (#)	Amount (1000 USD)	Percentage of goal achieved
Emotion in Project Images	Amusement	0.412	-0.10	**0.148*** **
	Awe	-1.128	-0.051	-0.624
	Contentment	**4.496***	0.189	**1.970** **
	Disgust	-1.504	-0.132	-0.201
	Excitement	-0.806	-0.082	-0.489
	Fear	**-5.794** **	**-0.331***	-1.162
	Sadness	**7.924*** **	**0.406** **	0.865
Control Variables	Anxiety in text description	15.376	1.243	4.109
	Anger in text description	0.200	0.471	-5.035
	Sadness in text description	-1.161	-1.510	8.389
	Preset goal	**0.005*** **	**0.001*** **	**-0.001*** **
	Project popularity	0.747	0.040	0.364
	Length of full description	**0.138*** **	0.003	0.009
	Number of images	**9.397*** **	**1.142*** **	**2.723** **
	Number of videos	**173.491*** **	**17.556*** **	23.468
	Duration	-0.435	-0.140	-1.605
Constant		-174.440	-10.256	**157.318*** **
R^2		31.2%	49.5%	5.2%

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

Table D4. Effects of image emotions on crowdfunding performance (N= 840)
With additional positive and negative emotion from LIWC

IV \ DV		Backers (#)	Amount (1000 USD)	Percentage of goal achieved	Backers (#)	Amount (1000 USD)	Percentage of goal achieved
Emotions in Title Images	Amusement	-1.089	-0.179	0.111			
	Awe	-1.238	0.088	-0.240			
	Contentment	**4.323****	**0.322***	**1.990****			
	Disgust	-1.599	-0.110	-0.124			
	Excitement	-0.319	-0.004	-0.337			
	Fear	**-4.585****	-0.273	-0.989			
	Sadness	**6.722*****	**0.383****	0.742			
Control Variables	Anxiety in text description	25.209	3.130	9.509	21.716	3.003	8.829
	Anger in text description	34.839	3.116	0.072	34.409	3.467	1.593
	Sadness in text description	**58.701****	3.133	14.480	**55.634***	2.879	14.382
	Positive emotion in text description	-5.392	-0.619	-1.241	-4.633	-0.578	-1.038
	Negative emotion in text description	-21.261	-2.522	-6.841	-21.071	-2.561	-7.279
	Preset Goal	**0.002*****	**0.000*****	**0.000****	**0.002*****	**0.000*****	**0.000****
	Project popularity	**0.888***	0.028	0.102	0.751	0.019	0.062
	Length of full description	**0.178*****	**0.011*****	0.006	**0.174*****	**0.011*****	0.005
	Number of Images	**7.544*****	**1.119*****	**2.561*****	**7.868*****	**1.138*****	**2.543*****
	Number of Videos	**181.477*****	**18.310*****	**40.070*****	**189.061*****	**18.859*****	**43.540*****
	Duration	0.202	0.077	-1.224	0.115	0.082	-1.218
Constant		-140.9	**-15.707***	**131.156*****	**-141.198***	**-16.136****	**131.610*****
R^2		24%	29.3%	5.9%	20.8%	27.6%	4.5%

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

Table D5. Effects of image emotions on crowdfunding performance (N= 840)
Text emotions based on IBM Tone Analyzer

IV \ DV		Backers (#)	Amount (1000 USD)	Percentage of goal achieved	Backers (#)	Amount (1000 USD)	Percentage of goal achieved
Emotions in Title Images	Amusement	-0.898	-0.172	0.227			
	Awe	-1.065	0.103	-0.218			
	Contentment	**4.377****	**0.326***	**1.969****			
	Disgust	-1.599	-0.104	-0.116			
	Excitement	-0.467	-0.011	-0.331			
	Fear	**-4.357****	-0.269	-0.955			
	Sadness	**6.693*****	**0.375****	0.826			
Control Variables	Anger in text description	29.696	25.034	92.823	6.619	22.047	72.517
	Disgust in text description	50.232	-11.580	-37.394	-44.028	-18.083	-57.041
	Fear in text description	252.824	15.377	114.571	246.222	15.003	122.947
	Joy in text description	-91.963	-8.365	54.627	-122.089	-9.983	47.002
	Sadness in text description	111.733	-15.214	**139.440***	84.632	-16.443	132.032
	Preset Goal	**0.002*****	**0.000*****	**0.000**	**0.002*****	**0.000*****	**0.000****
	Project popularity	**1.059****	0.037	0.113	**0.936***	0.028	0.076
	Length of full description	0.182***	0.012***	0.005	0.180***	0.012***	0.005
	Number of Images	**7.359*****	**1.093*****	**2.597*****	**7.595*****	**1.105*****	**2.550*****
	Number of Videos	**185.650*****	**18.352*****	**41.295*****	**193.729*****	**18.943*****	**44.880*****
	Duration	0.129	0.070	**-1.239***	0.048	0.074	**-1.236***
Constant		-151.973	-14.578	62.087	-124.684	-13.003	73.411
R^2		23.7%	29.1%	6.1%	20.6%	27.5%	4.7%

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

Appendix E. Made-up Crowdfunding Project for the Experiment

TNR- Trap, Neuter, and Return

by CatVille

On the street of town Donhou, numerous cats live here. They bring energy and vitality to the town. However, they are also a huge burden for Donhou. Although people love cats, the impact of cats still needs to be addressed.

Traditionally, government officers would catch stray cats and take them to shelter. If nobody adopts these cats, after 14 days in shelter, mercy killing will be the destination of these cats. Due to various environmental factors that might be dangerous to cats, stray cats usually have a short life span between 2 and 3 years. Since they only have such short life span, it would be cruel to just take their life away. Thus, we launch this campaign to help do birth control for these stray cats.

The TNR (Trap, Neuter, and Return) program catches stray cats and releases them back to the street they live after sterilizing. Furthermore, sterilizing could reduce the aggressive action and courting noise of cats, which makes sterilized stray cats' life span longer than usual and improves the life quality of residents.

Although you might never meet any cats in Donhou, you can still help them with your donation. Your help means a lot for both us and the stray cats in Donhou.

Appendix F. Moderation of Different Goals

The mean goal of all 840 projects in our empirical dataset is USD 27600.91. We use the mean to separate 840 projects into high and low goal groups to test the possible moderation of different goals.

Table F1. Effects of image emotions on crowdfunding performance
(Higher goal project, mean = 82568.27) (N= 207)

IV \ DV		Backers (#)	Amount (1000 USD)	Percentage of goal achieved	Backers (#)	Amount (1000 USD)	Percentage of goal achieved
Emotions in Title Images	Amusement	-2.278	-0.628	-0.357			
	Awe	-0.869	0.552	**1.008***			
	Contentment	-3.445	-0.050	-0.267			
	Disgust	-5.312	-0.261	-0.302			
	Excitement	0.684	-0.217	-0.152			
	Fear	**-13.555****	-0.967	-1.082			
	Sadness	**21.548*****	**1.084***	0.102			
Control Variables	Anxiety in text description	-34.234	-3.633	-9.969	-43.655	-4.855	-9.748
	Anger in text description	43.384	0.948	-0.976	41.868	3.670	1.817
	Sadness in text description	58.657	1.005	1.563	41.321	-1.096	0.195
	Preset Goal	**0.002*****	**0.000*****	0.000	**0.002*****	**0.000*****	0.000
	Project popularity	**4.499*****	0.201	0.003	4.835***	0.192	-0.018
	Length of full description	**0.267*****	0.014	0.011	**0.272*****	**0.016***	0.015
	Number of Images	3.215	**1.287****	1.005	2.830	**1.245****	0.935
	Number of Videos	**278.981*****	**33.412****	**48.424*****	**279.596*****	**32.588*****	**46.468*****
	Duration	0.782	0.219	0.283	2.660	0.355	0.333
Constant		-243.554	-15.060	51.458	-349.154	-28.917	32.975
R^2		31.0%	27.3%	17.9%	21.6%	22.7%	15%

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

Table F2. Effects of image emotions on crowdfunding performance
(Lower goal project, mean = 9625.81) (N= 633)

IV	DV	Backers (#)	Amount (1000 USD)	Percentage of goal achieved	Backers (#)	Amount (1000 USD)	Percentage of goal achieved
Emotions in Title Images	Amusement	-0.060	0.016	0.154			
	Awe	-1.288	-0.061	-0.539			
	Contentment	**6.839*** **	**0.323*** **	**3.261*** **			
	Disgust	-0.535	-0.042	-0.179			
	Excitement	-1.623	-0.033	-0.166			
	Fear	-1.588	-0.094	-0.934			
	Sadness	0.139	0.046	0.982			
Control Variables	Anxiety in text description	16.466	1.638	2.162	17.148	1.590	1.504
	Anger in text description	2.519	-0.343	-2.009	-2.046	-0.620	-5.009
	Sadness in text description	9.158	0.511	11.569	9.126	0.423	10.875
	Preset Goal	**0.010*** **	**0.001*** **	**-0.005*** **	**0.009*** **	**0.001*** **	**-0.005*** **
	Project popularity	-0.453	-0.017	0.196	-0.559	-0.023	0.133
	Length of full description	**0.071** **	0.002	0.019	**0.069** **	**0.002**	0.018
	Number of Images	**7.687*** **	**0.703*** **	**4.655*** **	**7.606*** **	**0.695*** **	**4.643*** **
	Number of Videos	**112.978*** **	**6.529*** **	**46.323** **	**120.186*** **	**6.821*** **	**49.019** **
	Duration	-2.035	-0.135	-1.205	-1.742	-0.123	-1.096
Constant		2.186	-3.108	**132.359** **	-20.452	-3.231	**134.718*** **
R^2		17%	23.3%	7.9%	13.3%	21.2%	5.5%

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

CHAPTER THREE

WHICH COLOR PERSUADES YOU?

THE EFFECT OF COLOR DESIGN

ON WARNING MESSAGES AND MALVERTISING

Jian-Ren Hou

Jingguo Wang

Jie Jennifer Zhang

Ying-Feng Kuo

Abstract

Warning message is an important feature for web browsers to stop users from browsing malicious websites, and malvertising is a serious concern for cybersecurity. How to design effective warning messages to stop users and what design of malvertising can draw users to click are important issues for cybersecurity. There are still rare researches have tested the background color effect on warning message and malvertising in the information system area. Prior researches of color effects usually indicate the effect of each specific color. However, the results might be too broad to follow. We employed standard color definition, RGB and HSL models, to conduct the research. The results indicate that red, green, blue, brightness, and lightness are significantly effective in changing users' likelihood to be persuaded by the warning messages and malvertising, and the effect size of each background color factor is provided. The results of this study provide insights for researchers to probe into warning messages and malvertising and contribute applicable and objective standards for practitioners to design warning messages and cope with malvertising.

Keywords: Cybersecurity, Warning message, Malvertising, color, Conjoint analysis

1. Introduction

When we explore the Internet, sometimes, a warning message pops up by browsers such as Google Chrome. The message alerts that the website requested is probably malicious and suggests that stopping browsing. The possible consequences for users to visit a malicious website could be malware (such as ransomware, virus, trojan horse) infection, or even taking advantage of users' computing capacity to mine cryptocurrency such as Monero (Pressman, 2017). While there could be false positive, the safest strategy for users is leaving the possible malicious webpage for sure. However, users may or may not take the advice. For cybersecurity propose, how to design effective warning messages to stop browsing possible malicious websites is critical.

Malvertising is a combination of malware and advertising (Huang et al., 2018). Often, malvertising pop-ups to draw one to follow its instructions or click the advertising. Compare to warning messages, malvertising tries to draw users to click it. It then possibly redirects users to some malicious websites or infecting users' computer systems with malware (Network Security, 2015). The mainstream of malvertising attack is based on redirection, users click on web advertising they perceived as legitimate which should lead them to websites they requested, but the advertising actually redirect users' requested web page to malicious websites and users' computer system would possibly be infected by malware such as trojan and ransomware (Li et al., 2012; Pathak, 2016; Sood & Enbody, 2011). One main strategy of malvertising is to threaten users to click the malvertising. For instance, malvertising displays the message, "Virus detected! Scan your device now!". By using this kind of fake and exaggerated message, malvertising might be able to threaten users to click. On an annual basis, reports show that recovery and repairing costs could be over 1 billion USD in damages caused by malvertising attacks (Invincea, 2015; Network

Security, 2015). Gaining in-depth insights into how users are lured to click malvertising may help scholars and practitioners to develop the strategies for dealing with malvertising.

While there may be different situation factors and design elements influencing users' choices and decisions, the purpose of this study is modest. This study investigates how the background colors may change the effectiveness in stopping users from browsing the warned website and drawing users to click malvertising. The default design of Google Chrome's warning messages is red in the background. Observing from our daily lives, we may discover that signs which want us to stop doing something are usually in red (ex: stop sign, traffic signal, or "do not enter" sign). Previous research also provides evidence that red is perceived as the most hazardous color by human subjects (Braun et al., 1994). We could infer that employing red might be effective in designing persuasive warning messages. For malvertising, previous research has rarely been examining how the background color of malvertising draws users to click, though some studies in other research areas indicate that blue signals trustworthy (Aslam, 2006; Jacobs et al., 1991).

However, to the best of our knowledge, there is still a lack of systematic research examining the effects of the background color of cybersecurity-related messaging. This study aims to show the effects of different color settings (the combination of red, blue, and green with the RGB model, as well as that of saturation and lightness with the HSL model). The RGB color model is generally applicable to mainstream operation systems such as Windows and Mac, and objective. Our study is to show the different effects of the three primary colors. The HSL model defines colors by degrees of hue, lightness, and saturation. This study shows how different levels of lightness and saturation can affect the effectiveness of warning messages and malvertising.

To find the effect of different background color settings for warning messages and malvertising, this study conducts a web choice-based conjoint analysis via online surveys. This study uses different color settings to generate several colors for our study and ask participants to choose between colors they perceive as effective for warning messages and malvertising. To ensure the generalizability of this study, this study collects data from two different cultures, American culture and Chinese culture. The data was collected via Amazon Mturks in United States and via social networks in Taiwan for the respective culture. After data collection, this study analyzes the effect of each color setting, and the findings indicate that most of the color settings are related to the likelihood whether the users stop browsing by warning messages or click the malvertising. For instance, increasing the red color in the background of warning messages might have a higher likelihood that users stop browsing, but increasing the red color in the background of malvertising might have a lower likelihood that users click malvertising. Our findings discover not only the effect but also the effect size of each color factor.

This study contributes in the following ways. First, this study shows the effects of the primary color elements, such as red, green, blue, saturation, and lightness, for researchers and practitioners to refer to design warning messages and to counter malvertising. Second, prior cybersecurity studies for warning messages mostly focus on how warning messages act to better alarm users or how users react to warning messages. The color background design is merely investigated. Third, most malvertising studies are on the stage of exploratory, to the best of our knowledge, this study is the first to investigate the color background setting of malvertising. Fourth, this study provides methodological contribution that traditional way to conduct color-related research through electronic display (e.g. LCD, OLED, or any type of screens) need to address the color shift issue between different displays, but our methodology provides a possible solution for this issue that the effects of color are examined through comparison instead of judgment on one color a time.

2. Literature Review

2.1 Warning Messages and Malvertising

Warning messages of the browsers are frequently discussed by literature. As cybercrime grows, warning messages protect internet users from possible threats such as phishing, fake websites, and malware infection. Reviewing the literature, this study can roughly see two different main streams of investigating warning message of browsers. The first stream focuses on the properties of users. Akhawe & Felt (2013) conclude that effectiveness of warning messages depends on users' demographic groups and suggests that demographics should be considered while designing warning messages. Using fMRI technology, Neupane et al. (2016) investigate what region in participants' brain is activated while seeing a warning message. Based on brain activity, they conclude that language comprehension, decision-making, attention, and problem-solving ability is critical while users trying to make decision about heeding the warning messages or not. Another stream focuses on the design of the warning messages. Before Google Chrome was released in 2008, Kuo et al. (2007) design a browser interface that will disable links and show warnings while detecting a possible malicious website. Egelman et al. (2008) conclude that if a browser only alarms users by pop-up messages, it will not be as effective as a warning message which interrupts users' current task. Sunshine et al. (2009) embedding an additional question on the first page of a multi-paged warning message design, the result indicates that the design is more effective than some conventional warning designs of Google Chrome and Firefox. Harbach et al. (2013) indicate that while reading warning messages, users prefer the linguistic contents, which are short, simple, and with less technical properties. The results from Harbach et al. (2013) echo the results from Neupane et al. (2016) that language plays an important role while users are reading warning messages. Weinberger & Felt (2016) design different length of periods that browser removes the suspected websites from exception list of warning messages, they conclude that a week is a balanced setting between a too long period (users are unprotected) and a too short period (users ignore the warning due to inertia). Silic et al. (2017) conclude that if color used in designing warning messages can fit the culture of the users, it can better express perceive risk and makes users to comply. Overall, various works regarding the design of browser warning messages were conducted. However, only Silic et al. (2017) investigate the effect of color in the research, and the results mostly infer the cultural factor of colors. The background color of warning messages has rarely been investigated.

Overall, academic works regarding malvertising are still on the stage of exploratory, and most prior literature states their review and opinions regarding how malvertising spread and suggests how to counter malvertising (Dwyer & Kanguri, 2017; Kumar et al., 2017; Pathak, 2016; Sood & Enbody, 2011). Only a few studies have conduct research on malvertising issues. Investigating the channel which can be utilized by malvertising, Xing et al. (2015) indicate that several ad-injection extensions in the Google Chrome store can bring malvertising to users' devices. Zabyelina (2017) explores if malvertising can be a chance for criminals to distribute substandard, spurious, falsely labeled, falsified and counterfeit (SSFFC) medical products. Huang et al. (2017) and Huang et al. (2018) use game theory to simulate how malvertising inspection should be deployed numerically depends on various factors such as detection rate, ratio of malicious advertisers, and ratio of advanced malicious advertisers. In short, there is still none of academic work considering the background color setting of malvertising and its effect on malvertising. This study is the pioneer to explore the background color setting in malvertising.

Reviewing related works regarding browser warning messages and malvertising, this study discovers there is a research gap needs to be filled with background color setting to investigate how warning messages stop user and how malvertising draws users to click. Although colors and their effects are frequently investigated in different research areas (Braun et al., 1994; Luximon et al., 1998), the academic works regarding warning messages are still rare. For malvertising, there are lack of academic works. On top of that, traditional academic color works tend to provide a broader guideline about how to employ color (Braun et al., 1995; Braun et al., 1994; Luximon et al., 1998), our results can better provide a flexible and widely adapted guideline for warning message designers and malvertising investigators. All papers reviewed in this section are summarized in Appendix A.

3. Theoretical Background

3.1 The RGB and HSL Model

RGB and HSL are two color models this study employs in this research to define a background color setting. The RGB model is a classic addictive color space, and each color setting is resulted from the addition of different proportions of red, green, and blue (Kalist et al., 2015). To alter a background color setting, the designer may change the proportion of red, green, or blue. On the other hand, HSL provides a different way to define colors. In the HSL model, the base of a color is defined by the hue (H). Hue defines a similar and approximated color the designers desired. Saturation (H) defines how saturated the color is. Higher saturation creates a color that is closer to the original hue, and lower saturation creates a color that is closer to gray. Lightness (L) defines how light is the color. Higher lightness creates a more whitish color, and lower lightness creates a more blackish color. After defining the base hue, designers can change the saturation and lightness to alter the color (Kalist et al., 2015; Saravanan et al., 2016). The HSL model has been widely adopted by web designers and researchers (Kalist et al., 2015; Valdez & Mehrabian, 1994). Different and interchangeable terms are also used, such as chroma and value replacing saturation and lightness respectively. More technical details regarding the RGB and HSL model are provided in Appendix B.

The visual presentation would affect human subjects' judgment (Shedler & Manis, 1986), and colors are critical for message receivers' judgment (Braun et al., 1995). Warning messages and malvertising usually pop up when we are trying to do our task. For example, when we try to download software from a website, warning messages and malvertising suddenly pop up and interrupt our task. Users usually need to make the decision, accepting the message or not, in a very short time. Dual process theory elaborates human's thoughts and decision making in two different systems, one system is fast, intuitive, and automatic, and another one is slow, calculated, and thought-out (Chaiken, 1980; Kahneman, 2013; Samson & Voyer, 2012). The first process is called system 1 or heuristic process, and the second process is called system 2 or systematic process, depending on different scholars in different research areas (Samson & Voyer, 2012). This study argues that users would like to make a system 1 decision (Kahneman, 2013) in such a situation since they may try to make a decision rapidly and intuitively and back to their task as soon as possible. In the case of system 1 processing, users would like to focus on cues rather than the content itself (Chaiken, 1980). Thus, instead of the message itself (warning message and malvertising), background color plays an important role in users' decision making.

Since RBG and HSL models are two independent color models, this study proposes two respective research models for RGB and HSL. In the first subsection, this study uses the RGB model as the base and reviews the effects of adding or subtracting three primary colors, red, green, and blue. In the second subsection, this study uses HSL model and prior literature to discuss how saturation and lightness of color affect the users. On top of the effects of color factors, this study then adapts our study context, warning messages and malvertising, to develop our hypotheses that how background color of warning messages and malvertising affect users' decision-making.

3.2 The effects of red, green, and blue

To achieve effective design for warning messages, the color should be included since the color is an important and widely accepted factor regarding warning messages and warning signals design (Braun et al., 1995; Fagan et al., 2015). Red is perceived as more hazardous for the detergent product (Braun et al., 1995). Red is perceived to be significantly more hazardous than blue, black, and green (Braun et al., 1994), and red is perceived to be significantly more hazardous than other all other colors (ex: yellow, orange) which are frequently used in the warning (Luximon et al., 1998). Red is already studied by multiple research to show its effect and meaning of hazard, and red shows a strong effect of stopping users (Braun et al., 1994; Luximon et al., 1998).

Green is the complementary colors (opposite color) of red, which means red and green can create the strongest contrast. Scholars employ green as a comparison to red in their research (Elliot et al., 2007; Gnambs et al., 2010). Green mostly refers to the positive meaning, such as happiness, pure, and love. (Aslam, 2006; Gil & Bigot, 2014; L. Jacobs et al., 1991). Compare to red, which expresses the sense of disagreeing, disturb, pain, hazard, and other aggressive meanings, the effect of green seems to express the feeling of quiet (Goldstein, 1942; Jacobs & Suess, 1975). Compare to the red, green expresses the significantly a lower level of hazard (Luximon et al., 1998). Prior literature even shows that if the same sign, such as a DEADLY sign, is printed in green, individuals would feel a lower level of hazard compare an identical sign, which is printed in red (Braun et al., 1994).

Compare to red, blue also expresses a less aggressive message, and the level of hazard expressed by blue is similar to green but not as high as red (Braun et al., 1994; Goldstein, 1942).

Besides, blue means trustworthy and high quality in multiple cultures (Aslam, 2006; L. Jacobs et al., 1991). The previous study also argues that the blue can enhance the strength of the argument and can further affect persuasion (Soldat et al., 1997).

3.3 The effects of saturation and lightness

Color related literature suggested that saturation could affect viewers' emotion feeling regarding colors (Valdez & Mehrabian, 1994). Among all the similar colors, which are based on the same hue, the color with the highest saturation is more preferred (Camgöz et al., 2002). High saturation brought the emotion of pleasure and even high arousal (Valdez & Mehrabian, 1994). Saturation provides a positive effect on increasing the click-rates of UGC (user-generated content) (Jalali & Papatla, 2016). If the advertisement employs the color with higher saturation, which may increase the likability of the advertisement received (Gorn et al., 1997).

Based on the color definition of the HSL color model, high lightness leads a color close to white, and colors with maximum lightness will be pure white. On the other hand, low lightness leads a color close to black, and colors with minimum lightness will be purely black (Kalist et al., 2015; Saravanan et al., 2016). Research finds out that white can bring a similar lower level of hazard as green (Luximon et al., 1998; Smith-Jackson & Wogalter, 2000). Compare the black, white can show a significantly lower level of hazard (Smith-Jackson & Wogalter, 2000), which might also show the different effects between the two polarities of lightness.

This study summarizes the effects and meaning of all color factors form prior literature in the above sections. In the next section, this study then develops our hypotheses on the top of the literature review.

4. Hypotheses Development

Based on the literature, this study argues that while users receive a warning message with more red color in the background, the user will feel the possible hazard and threats behind the warning messages, and they might want to comply with the warning messages and stop browsing to avoid the possible danger. Thus, red will be the best color to warn and stop users from browsing malicious websites. Similarly, in the case of malvertising, this study argues that red might remind users how hazardous the malvertising is, and users may refuse to click on the malvertising. Red may possess a strong effect on stopping users from clicking on malvertising. This study hypothesizes that:

Hypothesis 1a: ***Red background is positively related to users' likelihood of stopping browsing the website.***

Hypothesis 1b: ***Red background is negatively related to users' likelihood of clicking on malvertising.***

Green will reduce users' perception of the hazard level regarding a message. For warning messages, if there is less green color in the background color, the level of perceived hazard is decreasing as well. Users may overlook the possible threats behind the warning message, and they will be less likely to comply and stop browsing. As our malvertising setting is based on threatening users to click on the link, the perception of a decreasing level of hazard may cause users to overlook the intimidation of malvertising and be less likely to be threatened and click. This study proposes hypothesis 2a and 2b:

Hypothesis 2a***: Green background is negatively related to users' likelihood of stopping browsing the website.***

Hypothesis 2b***: Green background is negatively related to users' likelihood of clicking on malvertising***.

Based on the literature, the color of blue may enhance the trustworthy and persuasion of messages. For warning messages, blue in the background may increase the persuasion of the warning messages, and users would tend to accept the warning messages and stop browsing. For malvertising, the malvertising will threaten users to follow its instructions, so increasing blue in the background makes the malvertising more trustworthy and persuades users to click on it. In short, blue may have a positive effect on both warning messages and malvertising to let users follow its instructions. Thus, this study hypothesizes that:

Hypothesis 3a: ***Blue background is positively related to users' likelihood of stopping browsing the website.***

Hypothesis 3b: ***Blue background is positively related to users' likelihood of clicking on malvertising.***

Since findings suggest that saturation provides positive effects in color, and the effects can even be seen in advertisement-related literature, the high saturation might benefit both warning messages and malvertising. Both warning messages and malvertising are trying to persuade users to accept its message just like the advertisements and UGC, and literature shows the positive effect of high saturation in the context of advertisement and UGC. This infers that users might be more willing to accept message with a high saturation background color since they feel positive and may prefer the messages delivered by warning messages and malvertising. Thus, this study hypothesizes as followings.

Hypothesis 4a: ***High saturation background is positively related to users' likelihood of stopping browsing the website.***

Hypothesis 4b: ***High saturation background is positively related to users' likelihood of clicking on malvertising***.

High lightness brings a more whitish color, and low lightness brings a more blackish color. The literature concludes that white can bring a similar level of hazard of green color (Luximon et al., 1998). Thus, this study suggests that lightness will bring the same effect of green, which will lower the level of hazard. For the warning message, lower hazard level makes users ignore the possible harm and be less likely to comply the warning messages. While malvertising tries to threaten users to click on it, lower hazard level makes users underestimate the intimidation of the malvertising and be less likely to click on malvertising. This study hypothesizes as followings:

Hypothesis 5a*: High lightness background is negatively related to users' likelihood of stopping browsing the website.*

Hypothesis 5b*: High lightness background is negatively related to users' likelihood of clicking on malvertising.*

5. Methodology

5.1 Design

The research method of this study is a web choice-based conjoint analysis, and the participants receive the online survey for the experiment. The experiment contains several questions, and each question contains three choices, color 1, color 2, and none of the above. In each question, the experiment asks participants to choose the background color which they think better persuade them to follow the instruction of warning messages or malvertising depends on different experimental scenarios. First, this study needs to design the colors for the experiment. Since color can be converted between the RGB model and HSL model. This study chooses the HSL model to generate the colors in the experiment since it is more intuitive than RGB (Kalist et al., 2015), and this study can convert those colors to RGB for further analysis. In the HSL model, the hue is defined by a 360-degree circle, in which this study chooses six different hues (60, 120, 180, 240, 300, 360).

Second, this study chooses three levels of saturation (25, 50, 75) and two levels of lightness (25, 75). Hence, 6 (hues) x 3 (saturation) x 2 (lightness) creates 36 different colors. However, if this study uses 36 colors to create questions might be too many for the participants to make their choice, and which may create unnecessary boredom and may cause possible threats to validity. To reduce the possible threats of boredom, this study designs the choice-based conjoint analysis by AlgDesign package of R programming language, which allows us to reduce the colors to 18 colors without missing the variance among colors (Aizaki & Nishimura, 2008). Since we can convert the same color between the HSL model and RGB model, even the design is based on the HSL, this study can still analyze the data for the RGB model after converting. This study lists all the colors we use in the experiment in Table 1. Color ID is created and used to identify each color in this research.

Table 1 All colors in experiment

	HSL model			RGB model		
Color ID	Hues	Saturation	Lightness	Red	Green	Blue
1	360	25	25	80	48	48
2	60	25	25	80	80	48
3	240	25	25	48	48	80
4	60	50	25	96	96	32
5	120	50	25	32	96	32
6	240	50	25	32	32	96
7	360	75	25	112	16	16
8	180	75	25	16	112	112
9	300	75	25	112	16	112
10	120	25	75	175	207	175
11	180	25	75	175	207	207
12	300	25	75	207	175	207
13	360	50	75	223	159	159
14	180	50	75	159	223	223
15	300	50	75	223	159	223
16	60	75	75	239	239	143
17	120	75	75	143	239	143
18	240	75	75	143	143	239

The result yields eighteen different colors; this study then randomizes the order of these colors and create two lists with the same colors but randomized the order of 18 colors. This study pair the two lists to create the color comparison in eighteen questions. Each question contains three choices, one is with the color from the first list, one is with the color from the second list, and the third one is "none of above." The design for all eighteen questions is listed in Table 2. It is worth mentioning that this study also randomizes the order of choice 1 and choice 2 for each question to avoid possible validity threats from the order of choice.

Table 2 Question design

Question ID	Choice 1 (Color ID)	Choice 2 (Color ID)	Choice 3
1	13	14	None of the above
2	16	3	None of the above
3	2	16	None of the above
4	18	17	None of the above
5	4	8	None of the above
6	12	15	None of the above
7	3	11	None of the above
8	17	18	None of the above
9	1	13	None of the above
10	9	10	None of the above
11	6	12	None of the above
12	10	4	None of the above
13	15	2	None of the above
14	7	6	None of the above
15	11	7	None of the above
16	14	1	None of the above
17	5	9	None of the above
18	8	5	None of the above

For each question in the warning message scenario, this study reproduces the design of Google Chrome warning messages. On the other hand, since this study focuses on the malvertising which tries to threaten users to click, this study mimics one malvertising which tries to threaten users that there is a critical update for our Internet browser that users must install now. Furthermore, this study embeds a fake URL on the browser interface of the experiment, which can also be used to distinguish malvertising. Based on the same warning message and malvertising, this study changes the background color according to our design. Each question asks the subject of which choice is more persuasive than others. For example, this study asks, “Which warning better stop you from keeping browsing website?” for warning message scenario, and we ask, “Which advertisement better persuade you?” for malvertising scenario. By doing this, this study can see the effect between paired colors in persuading the subjects. This study shows our design for different scenarios in Figure 1. The URL of malvertising is designed to be random and suspicious, which might give people who are aware of malvertising to recognize the malvertising.

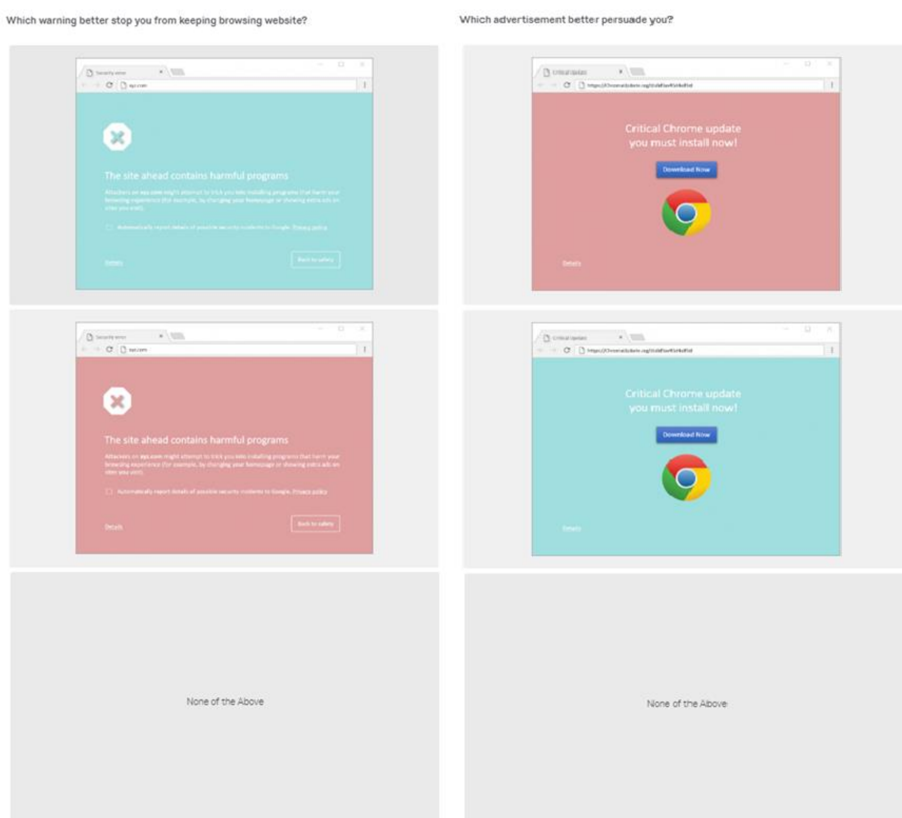


Figure 1 Design of warning message (left-hand side) and malvertising (right-hand side)

5.2 Procedure

The colors are the most important factors for this research. Thus, this study should make sure that every subject is able to distinguish different colors. After informed consent, this study employs the Ishihara test (Ishihara, 1972) to check if the subject is color-blind. Ishihara test is one of the most famous tests for testing color-blindness, and the previous study showed that the Ishihara test has a similar sensitivity in detecting color-blindness even through the electronic display (Semary et al., 2014). Following the suggested design, this study selects six questions in our Ishihara test, and this study displays one sample question from our Ishihara test in Figure 2. If subjects cannot pass the Ishihara test before the experiment, they are disqualified for our experiment, and the experiment stops immediately. The Ishihara test not only screens out color-blindness but also screens out those who do not pay attention to the experiment, which can assure the quality of our data. Before the actual experiment, survey page informs subjects that the task will be judging, among three available choices, which choice would better persuade the subjects, or none of these choices are persuasive. Moreover, each subject needs to answer all eighteen questions of eighteen paired colors. After collecting the data, this study collects demographic data of subjects.

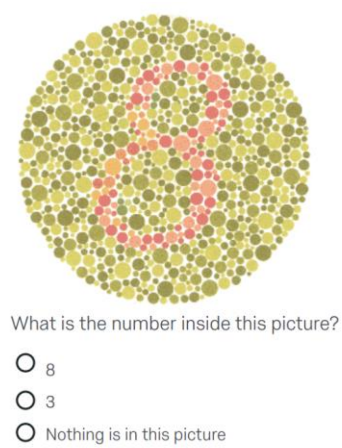


Figure 2 A sample Ishihara test question

This study tests the malvertising design by asking 112 subjects that are they aware of the advertisement is malicious. Twenty-five percent (25%) of subjects think that the advertisement is a malicious advertisement, and 75% think that the advertisement is normal. Although this study does not find any information regarding the proportion of Internet users who are aware of malvertising, this study believes our research can benefit from the malvertising design and the mixed sample (aware of and not aware of malvertising), which might be able to show the characteristic of malvertising.

5.3 Data collection

To generalize the results of this study, this study collects data from the US and Taiwan. In the US, this study posts the link of our experiment on the Mechanical Turk platform of Amazon to recruit subjects, and this study required our subjects must have a master qualification[4] to ensure the quality of our data. Besides, this study also requires the location of subjects must be in the US. In Taiwan, this study posts the link of our experiment on the authors' social network and asked for reposting those links to recruit subjects. It is worth mentioning that this study only released one link at a time, and this study also limits one subject can only join the experiment once. This study do this to avoid the effect of previous experience of who joins both experiments of warning messages and malvertising. After data collection and screening, the valid samples and demographic data from the US and Taiwan are listed in Table 3 and Table 4. Due to the education system and currency difference between the US and Taiwan, this study lists the demographic data in two separated tables.

[4] Amazon give master qualification to workers who showed good performance across different categories of task.

Table 3 Demographic data (US)

US			
Dimension	Category	Warning (n=89)	Malvertising (n=106)
Age	<20	0%	0.00%
	20-29	10.11%	11.32%
	30-39	51.69%	38.68%
	40-49	16.85%	23.58%
	>50	21.35%	26.42%
Education	Less than high school degree	0%	0.94%
	High school graduate (high school diploma or equivalent including GED)	11.24%	12.26%
	Some college but no degree	20.22%	13.21%
	Associate degree in college (2-year)	13.48%	16.98%
	Bachelor's degree in college (4-year)	38.20%	41.51%
	Master's degree	16.85%	9.43%
	Doctoral degree	0%	0.94%
	Professional degree (JD, MD)	0%	4.72%
Gender	Male	47.19%	45.28%
	Female	52.81%	54.72%
Income (Annually, USD)	Less than $20,000	17.98%	15.09%
	$20,000 to $39,999	22.47%	24.53%
	$40,000 to $59,999	20.22%	16.98%
	$60,000 to $79,999	22.47%	19.81%
	$80,000 to $99,999	7.87%	12.26%
	$100,000 or more	8.99%	11.32%

Table 4 Demographic data (Taiwan)

Taiwan			
Dimension	Category	Warning (n=102)	Malvertising (n=112)
Age	<20	1.96%	14.71%
	20-29	49.02%	57.84%
	30-39	34.31%	26.47%
	40-49	7.84%	7.84%
	>50	6.86%	2.94%
Education	Less than high school degree	0.00%	0.89%
	High school	0.00%	8.93%
	Bachelor's degree in college	59.80%	48.21%
	Master's degree	38.24%	41.07%
	Doctoral degree	1.96%	0.89%
Gender	Male	56.86%	60.71%
	Female	43.14%	39.29%
Income (Monthly, NTD)	Less than 20,000	21.57%	54.46%
	20,001 to 50,000	44.12%	27.68%
	50,001 to 80,000	20.59%	9.82%
	80,001 to 100,000	4.90%	3.57%
	100,001 to 120,000	2.94%	0.00%
	120000 or more	5.88%	4.46%

6. Analysis and Results

To analyze data collected, this study uses the survival package of the R programming language and apply the conditional logit model to analyze, and the analysis and data formation are based on choice-based experiment literature (Aizaki & Nishimura, 2008). Since literature also assumes the linear effect of the colors factors, this study also assumes the linear effect of factors of colors (Valdez & Mehrabian, 1994).

6.1 Study 1: RGB model

This study first converts the HSL model to the RGB model, and this study considers each choice of each question for each subject as one data point. If a subject selects the choice, this study labels that choice as 1. This study considers each question as one fixed effect, and this study uses the choice label as our dependent variable of the conditional logit model. A part of the data formation of our analysis is shown in Figure 3. In Figure 3, we can see that STR is used as a fixed effect in conditional logit model, 101 represents it is the first question for subject 1, and 102 represents it is the second question for subject 1, so on and so forth. Under each question, one choice can be selected by the subject among three options, and if the subject selects that choice, the RES will be labeled as 1. For example, subject 1 selected the first choice in the first question.

This study standardizes the value of RGB by dividing each value by 255, the maximum possible value of RGB, which can prevent extreme small coefficients in our analysis results.

STR	RES	RED	GREEN	BLUE
101	1	0.87451	0.623529	0.623529
101	0	0.623529	0.87451	0.87451
101	0	0	0	0
102	0	0.937255	0.937255	0.560784
102	1	0.188235	0.188235	0.313725
102	0	0	0	0
103	1	0.313725	0.313725	0.188235
103	0	0.937255	0.937255	0.560784
103	0	0	0	0
104	0	0.560784	0.560784	0.937255
104	0	0.560784	0.937255	0.560784
104	1	0	0	0

Figure 3 Data formation for RGB model

This study uses following model to analyze the conditional logit model:

$$logit\ (RES) = \beta 1 \cdot Red + \beta 2 \cdot Green + \beta 3 \cdot Blue + \sum \cdot STR_i + \varepsilon$$

Which i is the fixed effect of each question of each subject.

The results are listed in Table 5. The results are overall consistent between the US and Taiwan. Red can better persuade users to stop browsing the malicious website in warning messages, but red decreases the odds that users will click on the malvertising. Thus, H1a and H1b are supported.

According to our hypotheses, green should not stop users from browsing the possible malicious website in warning messages, and green should not make users click on the malvertising. The results support the H2a, which indicate that green will decrease the level of hazard, and users might not follow the instruction of warning messages. In addition, the results show that green will decrease the odds that users will click the malvertising. Hence, H2b is supported.

The results from the experiment support both H3a and H3b, showing that applying blue in warning messages design, the warning messages can better persuade users to stop browsing the possible malicious website. Besides, applying blue in malvertising design, the malvertising can draw users to click on malvertising.

Table 5 Analysis results of the RGB model

Scenario / RGB	Warning (US) R^2=0.087	Malvertising (US) R^2=0.032	Warning (TW) R^2=0.096	Malvertising (TW) R^2=0.045
Red	**3.5151*** (exp(b)=33.619, p=0.000)**	**-0.386** (exp(b)=0.680, p=0.032)**	**3.440 *** (exp(b)=31.187, p=0.000)**	**-1.302*** (exp(b)=0.272, p=0.000)**
Green	**-3.8022*** (exp(b)=0.022, p=0.000)**	**-1.496*** (exp(b)=0.224, p=0.000)**	**-4.383 *** (exp(b)=0.012, p=0.000)**	**-1.569*** (exp(b)=0.208, p=0.000)**
Blue	**0.4432** (exp(b)=1.558, p=0.022)**	**2.158*** (exp(b)=8.654, p=0.000)**	**0.612 *** (exp(b)=1.844, p=0.001)**	**1.9924*** (exp(b)=7.330, p=0.000)**

6.2 Study 2: HSL model

Testing H5 and H6, this study uses the HSL model instead of the RGB model in this analysis, and the data formation is in Figure 4. The data formation is the same as the RGB model. This study also standardizes the saturation and lightness by dividing the value by 100, the maximum value of saturation and lightness.

STR	RES	Hue	SATURATION	LIGHT
101	1	360	0.5	0.75
101	0	180	0.5	0.75
101	0	0	0	0
102	0	60	0.75	0.75
102	1	240	0.25	0.25
102	0	0	0	0
103	1	60	0.25	0.25
103	0	60	0.75	0.75
103	0	0	0	0
104	1	240	0.75	0.75
104	0	120	0.75	0.75
104	0	0	0	0

Figure 4 Data formation for HSL model

Since the hues is a complete circle, which 0 equals to 360, that makes the increasing of hue is not meaningful for our analysis. Thus, in addition to the fixed effect of each question, this study also considers hues as a fixed effect in our conditional logit model. Hence, the model this study uses to analyze is the following:

$$logit\ (RES) = \beta 1 \cdot Lightness + \beta 2 \cdot Saturation + \sum \cdot STR_i + \sum \cdot Hue_j + \varepsilon$$

Which i is the fixed effect of each question of each subject, and j is the fixed effect of each hue.

The results are listed in Table 6. Base on the results, the high saturation increases the odds that users will follow the warning messages and stop browsing the possible malicious website. H4a is supported. However, saturation does not have a significant effect on malvertising. H4b is not supported.

On the other hand, lightness brings a significantly negative effect on both warning messages and malvertising. That is, high lightness will not stop users from browsing the possible malicious website, and high lightness will not draw users to click on the malvertising. Thus, H5a and H5b are supported.

Table 6 Analysis results of the HSL model

SL \ Scenario	Warning (US) R^2=0.012	Malvertising (US) R^2=0.008	Warning (TW) R^2=0.015	Malvertising (TW) R^2=0.004
Saturation	**1.863** (exp(b)=6.443, p=0.027)**	1.258 (exp(b)=3.518, p=0.124)	**2.958** (exp(b)=19.529, p=0.000)**	0.746 (exp(b)=2.109, p=0.409)
Lightness	**-3.834*** (exp(b)=0.022, p=0.000)**	**-3.159*** (exp(b)=0.042, p=0.000)**	**-5.088*** (exp(b)=0.006, p=0.000)**	**-2.315*** (exp(b)=0.099, p=0.003)**

This study lists all our hypothesis testing and the results in Table 7.

Table 7 Summary of hypothesis testing

Hypothesis	**Results**
H1a: <u>Red background</u> is positively related to users' likelihood of stopping browsing the website.	Supported
H1b: <u>Red background</u> is negatively related to users' likelihood of clicking on malvertising.	Supported
H2a: <u>Green background</u> is negatively related to users' likelihood of stopping browsing the website.	Supported
H2b: <u>Green background</u> is negatively related to users' likelihood of clicking on malvertising.	Supported
H3a: <u>Blue background</u> is positively related to users' likelihood of stopping browsing the website.	Supported
H3b: <u>Blue background</u> is positively related to users' likelihood of clicking on malvertising.	Supported
H4a: <u>High saturation background</u> is positively related to users' likelihood of stopping browsing the website.	Supported
H4b: <u>High saturation background</u> is positively related to users' likelihood of clicking on malvertising.	Not Supported
H5a: <u>High lightness background</u> is negatively related to users' likelihood of stopping browsing the website.	Supported
H5b: <u>High lightness background</u> is negatively related to users' likelihood of clicking on malvertising.	Supported

6.3 Robustness check: Culture difference

The meaning of color might vary from culture to culture. Previous studies show that the same color might means totally different in various cultures (Aslam, 2006; Jacobs et al., 1991). For example, the meaning of red might be negative in US culture (Aslam, 2006). However, red means happiness and luck, and red is also the color of celebration in Chinese culture (Aslam, 2006; Jacobs et al., 1991). To check the robustness of our results, this study collects data from both US and Taiwan, which represent American culture and Chinese culture. This study uses the comparison method from literature to check the difference of the coefficients of conditional logit model (Paternoster et al., 1998). This study would like to see if our results hold in different cultures. The comparison results are listed in following tables. Table 8 lists the comparison results of warning messages between the US and Taiwan, and Table 9 lists the comparison results of malvertising between the US and Taiwan.

Table 8 Warning message comparison

Colors	**Taiwan Coefficient (SE)**	**US Coefficient (SE)**	**Coefficient Difference z-value (p)**
Red	3.440 (0.203)	3.515 (0.211)	-0.255 (p=0.399)
Green	-4.383 (0.223)	-3.802 (0.223)	**-1.845** (p=0.033)**
Blue	0.612 (0.187)	0.443 (0.194)	0.626 (p=0.734)
Saturation	2.958 (0.853)	1.863 (0.847)	0.911 (p=0.181)
Light	-5.088 (0.774)	-3.834 (0.75)	-1.164 (p=0.122)

Table 9 Malvertising comparison

Colors	**Taiwan Coefficient (SE)**	**US Coefficient (SE)**	**Coefficient Difference z-value (p)**
Red	-1.302 (0.186)	-0.386 (0.18)	**-3.525*** (p=0.000)**
Green	-1.569 (0.186)	-1.496 (0.178)	-0.283 (p=0.389)
Blue	1.992 (0.175)	2.158 (0.169)	-0.680 (p=0.248)
Saturation	0.746 (0.903)	1.258 (0.818)	-0.420 (p=0.337)
Light	-2.315 (0.769)	-3.159 (0.718)	0.802 (p=0.211)

From the results of the above tables, readers can see that most of the effects of colors are not significantly different between different cultures. Only green in warning message scenario and red in malvertising scenario show significant differences between different cultures. Although few coefficients are significantly different, the sign of each coefficient is still the same for all the effects of color factors. Thus, the culture difference should not be a threat to the external validity of this study.

6.4 Robustness check: Gender difference

Literature shows that there are gender differences that make males and females see the same color differently (Jaint et al., 2010). Thus, in this section, this study examines the effect of background color of warning messages and malvertising for males and females, respectively. For US and Taiwan, respectively, this study separates the sample into two biological genders and estimates the model again. The results of estimated model for two biological genders are listed in Appendix C. Then this study compares the coefficient between two biological genders (Paternoster et al., 1998). The comparison results are listed in following Tables.

Table 10 Gender comparison (Warning, US)

Colors	**Female Coefficient (SE)**	**Male Coefficient (SE)**	**Coefficient Difference z-value (p)**
Red	3.505 (0.292)	3.534 (0.306)	-0.069 (p=0.473)
Green	-4.118 (0.316)	-3.473 (0.315)	**-1.446* (p=0.074)**
Blue	0.737 (0.270)	0.126 (0.280)	**1.572* (p=0.057)**
Saturation	2.351 (1.257)	1.476 (1.161)	0.511 (p=0.305)
Light	-5.047 (1.145)	-2.765 (1.008)	**-1.496* (p=0.067)**

Table 11. Gender comparison (Malvertising, US)

Colors	**Female Coefficient (SE)**	**Male Coefficient (SE)**	**Coefficient Difference z-value (p)**
Red	-0.049 (0.242)	-0.795 (0.270)	**2.057** (p=0.020)**
Green	-1.961 (0.247)	-0.956 (0.259)	**-2.809*** (p=0.002)**
Blue	2.335 (0.231)	1.966 (0.248)	1.089 (p=0.138)
Saturation	1.594 (1.111)	0.851 (1.211)	0.452 (p=0.326)
Light	-3.166 (0.978)	-3.142 (1.060)	-0.017 (p=0.493)

Table 12 Gender comparison (Warning, Taiwan)

Colors	**Female Coefficient (SE)**	**Male Coefficient (SE)**	**Coefficient Difference z-value (p)**
Red	2.878 (0.298)	3.906 (0.279)	**-2.522*** (p=0.006)**
Green	-3.811 (0.322)	-4.867 (0.310)	**2.361*** (p=0.009)**
Blue	0.598 (0.279)	0.627 (0.253)	-0.076 (p=0.470)
Saturation	2.791 (1.320)	3.081 (1.119)	-0.168 (p=0.433)
Light	-5.239 (1.204)	-4.983 (1.010)	-0.163 (p=0.435)

Table 13 Gender comparison (Malvertising, Taiwan)

Colors	**Female Coefficient (SE)**	**Male Coefficient (SE)**	**Coefficient Difference z-value (p)**
Red	-1.37 (0.308)	-1.256 (0.238)	-0.293 (p=0.385)
Green	-1.519 (0.308)	-1.612 (0.235)	0.240 (p=0.405)
Blue	1.470 (0.290)	2.324 (0.222)	**-2.338*** (p=0.009)**
Saturation	0.628 (1.542)	0.804 (1.114)	-0.093 (p=0.463)
Light	-2.350 (1.301)	-2.292 (0.953)	-0.036 (p=0.486)

From the above tables, we can see that although the effects of some color factors are different between two biological genders, this study does not find any sign difference. That is, the effect size of color factors might vary between two biological genders, but the signs of color factors' effect are still the same between two biological genders.

7. Discussion

Red is perceived as the color with the highest hazard level among all colors (Braun et al., 1994; Luximon et al., 1998). The results of this study are consistent with the literature, in which red can warn users to stop browsing more effectively. Due to its effects of warning and arousing (Braun et al., 1994; Wilson, 1966), the red color can remind users of the possible dangerous aspect of malvertising and stop users from accepting the exaggerated message of malvertising. The green can decrease the level of hazard, and it is much less arousal than red (Braun et al., 1994; Wilson, 1966). Thus, applying the green in the background color might make users relax and neglect the warning, and green may also lower the users' vigilance and makes them ignore the exaggerated statements of malvertising and refuse to click the malvertising. Blue is considered a trustworthy and persuasive color (Aslam, 2006; L. Jacobs et al., 1991; Soldat et al., 1997). Our results support these findings by showing that blue can better persuade users to stop browsing after watching a warning message, and blue can better draw users to click on the malvertising.

Previous studies show that users prefer colors with higher saturation (Camgöz et al., 2002; Gorn et al., 1997). The effect of saturation is significantly positive in the warning message scenario, which is consistent with the literature. High lightness brings more whitish colors, and white is perceived as low level of hazard, which is similar to green (Braun et al., 1994; Luximon et al., 1998). Our results also show that high lightness will bring negative effects to both warning messages and malvertising, where green background in warning messages may make users ignore the danger and cannot stop users from browsing. Furthermore, malvertising cannot use its exaggerated intimidation to draw users to click on since green makes the malvertising feel less dangerous.

In the H4b, this study hypothesizes that high saturation will allure users to click the malvertising. However, the effect of saturation is not significant for malvertising. Prior literature mentioned that high saturation might yield higher pleasure, but the literature also shows that high saturation will yield higher arousal (Valdez & Mehrabian, 1994). This study argues that the higher arousal might stimulate users to see the abnormality of malvertising and further mitigate the pleasure brought by high saturation. This might cause the effect of saturation is not significant for malvertising.

7.1 Theoretical implication

This study discusses the contributions and theoretical implications in the following sections:

First, this research will be one of the first articles which concentrate on background color effect in the cybersecurity scenario, namely warning messages and malvertising. Numerous articles from other disciplines (ex: marketing) have provided evidence that color could affect humans in different ways, but can this study apply those findings in warning messages and malvertising is still questionable. Drawing on literature, this study applies those findings in this area and serve as a pioneer for color-related research in cybersecurity discipline.

Second, previous research only indicates which color should be used in a specific scenario. However, countless similar colors will be perceived as the same color by humans since humans could not tell the subtle differences among similar colors. What exact color should be used remain unanswered. This research addresses the above issue by controlling color objectively with widely accepted color models for the computer system to define colors. On top of that, the results of this study also provide a more elastic way too employ the colors. Our results show that 1) How to generate new effective colors by adding different proportions of the three primary colors 2) Among similar colors, how to adjust saturation and lightness to generate effective colors. This study not only shows the effect of color factors but also the effect size of them.

Third, there are a few studies that focus on malvertising design. This research will be the pioneer for exploring the design of malvertising. In the future, the researcher could reference the finding of this research and conduct future research to in-depth investigate how to counter the malvertising issues.

Fourth, researchers who employ online survey for researching color-related issues need to deal with the possible threat of hardware limitation, that is, different monitors might have a slight color shift in showing the exact same color. This might limit the external validity of the research's results since the same color might not be the same on different monitors. This research uses the choice-based method to collect the data. Although the color still shifts from monitors to monitors, the results of our research are based on the comparison between colors on the same monitor not based on the absolute values of color on the different monitors, which provides a different way to mitigate the possible effect of hardware limitation on color-related research. In addition to that, changing the background color might affect the readability of the message text itself, this may also be an issue for color researchers, the comparison based methodology should overcome this issue since participants can focus on the context and comparison between colors instead of the readability of the message text.

7.2 Managerial implication

To the best of our knowledge, this article is the first research which employs RGB and HSL model to control colors in the warning messages and malvertising design. Since the RGB and HSL model can be implemented by every modern computer system and are totally objective, this study could provide objective and applicable standards for practitioners to implement.

For warning message design, designers can adjust their color design by using different colors, saturation, or lightness to gain the desired effect they need without losing the aesthetic feature of their design. The results of this study provide a flexible way for designers to employ colors, saturation or lightness in warning message design. For malvertising, the current web browser can detect possible malicious website which may contain malvertising. In this situation, web browser developers could adjust the colors, saturation or lightness on the possible malicious websites or advertising according to our results. This might mitigate the damage from malvertising by making the malvertising less attractive and persuasive. This approach can also be applied to pop out advertisements, which might be malvertising as well.

7.3 Limitations & Future Research

There are several limitations inherent in this research. First, the reliability of treatments could be a possible threat to validity. The treatment of this research is graphical. However, different devices generate different output on the screen. For instance, some subjects might finish the survey on mobile devices, and the way of display such as layout on these devices might be different from personal computers. Although this study uses the same survey system, subtle differences in screen size still could be a threat to validity. Future research could address this issue by limiting the participants to use the same device to finish the survey.

Second, this study manipulates the color in the warning message of Google Chrome as our treatment. Since Google Chrome is the most popular web browser on earth, there should be many participants who have already known the standard color setting of Google Chrome warning message. They might notice the color difference between the standard warning message and manipulated ones. Thus, they might also do hypothesis-guessing in the experiment, which is a possible threat to validity here. To address this issue, future researchers could recruit subjects who are not using Google Chrome as their web browser in their daily life or design their own warning messages for research.

Third, in this study, although culture and gender do not affect the sign of the effect of color factors, there are some significant effect size differences. Thus, future researchers can test the effect of culture and gender in their context to provide in-depth insights for culture and gender differences in color effect.

Last but not least, there is a different strategy which malvertising employs to persuade users to click, in which the malvertising tries to use attractive incentives to allure users to click. For instance, “Congratulations! You just won an iPhone. Click here to redeem your new iPhone!” This study focuses on malvertising, which tries to threaten users to click the malvertising. Future researchers might want to test the second strategy to extend the insights for malvertising studies.

References

Aizaki, H., & Nishimura, K. (2008). Design and Analysis of Choice Experiments Using R: A Brief Introduction. *Agricultural Information Research*, *17*. https://doi.org/10.3173/air.17.86

Akhawe, D., & Felt, A. (2013, August 14). *Alice in warningland: A large-scale field study of browser security warning effectiveness*. 257–272.

Aslam, M. M. (2006). Are You Selling the Right Colour? A Cross-cultural Review of Colour as a Marketing Cue. *Journal of Marketing Communications*, *12*(1), 15–30. https://doi.org/10.1080/13527260500247827

Ba, S., & Pavlou, P. A. (2002). Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Q.*, *26*(3), 243–268. https://doi.org/10.2307/4132332

Braun, C. C., Mine, P. B., & Clayton Silver, N. (1995). The influence of color on warning label perceptions. *International Journal of Industrial Ergonomics*, *15*(3), 179–187. https://doi.org/10.1016/0169-8141(94)00036-3

Braun, C. C., Sansing, L., & Silver, N. C. (1994). The Interaction of Signal Word and Color on Warning Labels: Differences in Perceived Hazard. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *38*(14), 831–835. https://doi.org/10.1177/154193129403801407

Camgöz, N., Yener, C., & Güvenç, D. (2002). Effects of hue, saturation, and brightness on preference. *Color Research & Application*, *27*(3), 199–207. https://doi.org/10.1002/col.10051

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Dwyer, C., & Kanguri, A. A. (2017). Malvertising-A Rising Threat To The Online Ecosystem. *Journal of Information Systems Applied Research (JISAR)*, *10*, 29.

Egelman, S., Cranor, L. F., & Hong, J. (2008). You've been warned: An empirical study of the effectiveness of web browser phishing warnings. *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*, 1065. https://doi.org/10.1145/1357054.1357219

Fagan, M., Khan, M. M. H., & Nguyen, N. (2015). How does this message make you feel? A study of user perspectives on software update/warning message design. *Human-Centric Computing and Information Sciences*, *5*(1), 36. https://doi.org/10.1186/s13673-015-0053-y

Felt, A. P., Reeder, R. W., Almuhimedi, H., & Consolvo, S. (2014). Experimenting At Scale With Google Chrome's SSL Warning. *ACM CHI Conference on Human Factors in Computing Systems*.

Gil, S., & Bigot, L. L. (2014). Seeing Life through Positive-Tinted Glasses: Color–Meaning Associations. *PLoS One; San Francisco*, *9*(8), e104291. http://dx.doi.org.ezproxy.uta.edu/10.1371/journal.pone.0104291

Goldstein, K. (1942). Some Experimental Observations Concerning the Influence of Colors on The Function of the Organism. *American Journal of Physical Medicine & Rehabilitation*, *21*(3), 147.

Gorn, G. J., Chattopadhyay, A., Yi, T., & Dahl, D. W. (1997). Effects of Color As an Executional Cue in Advertising: They're in the Shade. *Management Science*, *43*(10), 1387–1400. Retrieved from JSTOR.

Harbach, M., Fahl, S., Yakovleva, P., & Smith, M. (2013). Sorry, I Don't Get It: An Analysis of Warning Message Texts. In A. A. Adams, M. Brenner, & M. Smith (Eds.), *Financial Cryptography and Data Security* (pp. 94–111). Springer Berlin Heidelberg.

HSL and HSV. (2017). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/HSL_and_HSV

Huang, C., Sakib, M. N., Kamhoua, C., Kwiat, K. A., & Njilla, L. (2018). A Bayesian Game Theoretic Approach for Inspecting Web-based Malvertising. *IEEE Transactions on Dependable and Secure Computing*, 1–1. https://doi.org/10.1109/TDSC.2018.2866821

Huang, C., Sakib, M. N., Kamhoua, C., Kwiat, K., & Njilla, L. (2017). A game theoretic approach for inspecting web-based malvertising. *2017 IEEE International Conference on Communications (ICC)*, 1–6. https://doi.org/10.1109/ICC.2017.7996807

Invincea. (2015, August 12). *Invincea 1H 2015 Advanced Endpoint Threat Report*. Retrieved from https://www.invincea.com/2015/08/invincea-research-details-latest-endpoint-threats-including-the-billion-dollar-malvertising-problem/

Ishihara, S. (1972). *Tests for Color-Blindness*. Kanehara Shuppan Co., Ltd.

Jacobs, K. W., & Suess, J. F. (1975). Effects of Four Psychological Primary Colors on Anxiety State. *Perceptual and Motor Skills*, *41*(1), 207–210. https://doi.org/10.2466/pms.1975.41.1.207

Jacobs, L., Keown, C., Worthley, R., & Ghymn, K. (1991). Cross-cultural Colour Comparisons: Global Marketers Beware! *International Marketing Review*, *8*(3). https://doi.org/10.1108/02651339110137279

Jaint, N., Verma, P., Mittal, S., Singh, A., & Munjal, S. (2010). Gender based alteration in color perception. *Indian J Physiol Pharmacol*, *54*(4), 366–70.

Jalali, N. Y., & Papatla, P. (2016). The palette that stands out: Color compositions of online curated visual UGC that attracts higher consumer interaction. *Quantitative Marketing and Economics*, *14*(4), 353–384. https://doi.org/10.1007/s11129-016-9178-1

Kahneman, D. (2013). *Thinking, Fast and Slow* (1st edition). New York: Farrar, Straus and Giroux.

Kalist, V., Ganesan, P., Sathish, B. S., Jenitha, J. M. M., & Basha.shaik, K. (2015). Possiblistic-Fuzzy C-Means Clustering Approach for the Segmentation of Satellite Images in HSL Color Space. *Procedia Computer Science*, *57*, 49–56. https://doi.org/10.1016/j.procs.2015.07.364

Kumar, S., Rautaray, S. S., & Pandey, M. (2017). Malvertising: A case study based on analysis of possible solutions. *2017 International Conference on Inventive Computing and Informatics (ICICI)*, 288–291. https://doi.org/10.1109/ICICI.2017.8365356

Kuo, C., Schneider, F., & Jackson, C. (2007). *Browser system and method for warning users of potentially fraudulent websites*. Retrieved from https://patents.google.com/patent/US20070130327A1/en

Li, Z., Zhang, K., Xie, Y., Yu, F., & Wang, X. (2012). Knowing Your Enemy: Understanding and Detecting Malicious Web Advertising. *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, 674–686. https://doi.org/10.1145/2382196.2382267

Luximon, A., W. Chung, L., & Goonetilleke, R. (1998). *Safety Signal Words and Color Codes: The Perception of Implied Hazard by Chinese People*.

Network Security. (2015). Malvertising hits dating websites. *Network Security*, *2015*(9), 2. https://doi.org/10.1016/S1353-4858(15)30075-1

Neupane, A., Saxena, N., Maximo, J. O., & Kana, R. (2016). Neural Markers of Cybersecurity: An fMRI Study of Phishing and Malware Warnings. *IEEE Transactions on Information Forensics and Security*, *11*(9), 1970–1983. https://doi.org/10.1109/TIFS.2016.2566265

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the Correct Statistical Test for the Equality of Regression Coefficients. *Criminology*, *36*(4), 859–866. https://doi.org/10.1111/j.1745-9125.1998.tb01268.x

Pathak, P. B. (2016). Malware a Growing Cybercrime Threat: Understanding and Combating Malvertising Attacks. *International Journal of Advanced Research in Computer Science*, *7*(2), 9–11.

Pressman, A. (2017, November 15). *The Growing Threat of Cryptocurrency Mining Malware*. Retrieved from http://fortune.com/2017/11/15/the-growing-threat-of-cryptocurrency-mining-malware/

Samson, A., & Voyer, B. G. (2012). Two minds, three ways: Dual system and dual process models in consumer psychology. *AMS Review*, *2*(2), 48–71. https://doi.org/10.1007/s13162-012-0030-9

Saravanan, G., Yamuna, G., & Nandhini, S. (2016). Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models. *2016 International Conference on Communication and Signal Processing (ICCSP)*, 0462–0466. https://doi.org/10.1109/ICCSP.2016.7754179

Semary, N., Mandour, S., & Marey, H. (2014). Ishihara Electronic Color Blindness Test: An Evaluation Study. *Ophthalmology Research*, *3*, 67–75. https://doi.org/10.9734/OR/2015/13618

Shah, R., & Patil, K. (2016). Evaluating Effectiveness of Mobile Browser Security Warnings. *Ictact Journal on Communication Technology*, *7*, 1373–1379.

Shedler, J., & Manis, M. (1986). Can the availability heuristic explain vividness effects? *Journal of Personality and Social Psychology*, *51*(1), 26–36. https://doi.org/10.1037/0022-3514.51.1.26

Silic, M., Njavro, M., & Oblakovic, G. (2017). Understanding Color Risk Appropriateness: Influence of Color on a User's Decision to Comply with the IT Security Policy—Evidence from the U.S. and India. In F. F.-H. Nah & C.-H. Tan (Eds.), *HCI in Business, Government and Organizations. Supporting Business* (pp. 412–423). Springer International Publishing.

Smith-Jackson, T. L., & Wogalter, M. S. (2000). Users' Hazard Perceptions of Warning Components: An Examination of Colors and Symbols. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(32), 6–55. https://doi.org/10.1177/154193120004403215

Soldat, A. S., Sinclair, R. C., & Mark, M. M. (1997). Color as an Environmental Processing Cue: External Affective Cues Can Directly Affect Processing Strategy Without Affecting Mood. *Social Cognition; New York*, *15*(1), 55–71. http://dx.doi.org.ezproxy.uta.edu/10.1521/soco.1997.15.1.55

Sood, A. K., & Enbody, R. J. (2011). Malvertising – exploiting web advertising. *Computer Fraud & Security*, *2011*(4), 11–16. https://doi.org/10.1016/S1361-3723(11)70041-0

Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., & Cranor, L. F. (2009). Crying Wolf: An Empirical Study of SSL Warning Effectiveness. *Proceedings of the 18th Conference on USENIX Security Symposium*, 399–416. Retrieved from http://dl.acm.org/citation.cfm?id=1855768.1855793

Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, *123*(4), 394–409. https://doi.org/10.1037/0096-3445.123.4.394

Weinberger, J., & Felt, A. (2016). *A Week to Remember: The Impact of Browser Warning Storage Policies*. 15–25. Denver, CO, USA.

Wilson, G. (1966). Arousal properties of red versus green. *Perceptual and Motor Skills*, *23*, 947–949. https://doi.org/10.2466/pms.1966.23.3.947

Xing, X., Meng, W., Lee, B., Weinsberg, U., Sheth, A., Perdisci, R., & Lee, W. (2015). Understanding Malvertising Through Ad-Injecting Browser Extensions. *Proceedings of the 24th International Conference on World Wide Web*, 1286–1295. https://doi.org/10.1145/2736277.2741630

Zabyelina, Y. G. (2017). Can criminals create opportunities for crime? Malvertising and illegal online medicine trade. *Global Crime*, *18*(1), 31–48. https://doi.org/10.1080/17440572.2016.1197124

Appendix A: Literature of Warning Messages and Malvertising

Table A1. Literature of Warning Messages

Reference	Main Findings	Method	Publication Type
(Kuo et al., 2007)	If the web browser detects a possible fraudulent webpage, the browser will grayed-out the webpage, disable the links, and alarm users by displaying warning messages in the browser.		Patent
(Egelman et al., 2008)	To cope with the phishing attack, web browsers alarm users by showing warning messages. This study examines the effectiveness of warning design. The results show that the warning should interrupt users' task to effectively alarm users. If the warning only passively alarms users by showing pop-up message, it is not as effective as the warning with interruption.	Laboratory experiment	Conference
(Sunshine et al., 2009)	This study designs a new SSL connection warning message with multi-page and question. In the first page, the question is used to assess how risky the connection is. Then the second page will show different warnings according to the answer of the first page. The result show that the design is better in persuading users than some conventional designs, however, it is not better than contemporary designs.	Laboratory experiment	Conference
(Harbach et al., 2013)	This study investigate that how linguistic properties of warnings affect user's preference. The results show that users prefer the warnings with simple headline, shorter sentence, less technical term, and no complicated grammatical construction.	Survey; Interview	Conference
(Akhawe & Felt, 2013)	Under users' consent, this study observes over 25 million real warning impressions and conclude that warning message is effective to stop user with Chrome and Firefox. Another result indicates that the effectiveness of warnings might be different between demographic groups, and the authors suggest that warnings should be designed while considering the demographics.	Field study	Conference

Table A1. Literature of Warning Messages (Cont.)

Reference	Main Findings	Method	Publication Type
(Felt et al., 2014)	The results of Akhawe & Felt (2013) indicate that warning message is effective. However, the result also shows that the warning message of Firefox is more effective than Chrome. Therefore, this study investigates the design aspect of warnings of Chrome and Firefox. This study replaces the default design of Chrome with six different designs. Observing over 130 thousand real warning impression, this study concludes that additional human images do not have or have small effect on the effectiveness of warning message. Other results including stylizing the warnings with google style also shows a small effect.	Field study	Conference
(Weinberger & Felt, 2016)	Browser HTTPS warning alarms users that the connection is unsafe. However, due to the possible false alarms, users might choose to except the warnings. How long should the browser memorize the exceptions? If it is too long, users might be unprotected from the attack. If it is too short, users might get used to the warnings and choose to ignore. From the results of the field experiment, the authors suggest that a one-week period will be a balance setting for exception policy for users to pay attention to the warnings. The setting was adopted by Google Chrome.	Field experiment	Conference
(Shah & Patil, 2016)	Focusing on mobile browser warnings, this study investigates twenty-three mobile browsers, the results show that most of mobile browser are failed to provide comprehensive security warnings. Among all the mobile browsers investigated, only four can provide comprehensive security warnings during possible security threats.	Field study	Academic Journal
(Neupane et al., 2016)	Employing fMRI technology, this study discovers that while facing a decision to follow a security warning or not, there are significant activations in the brain regions associated with language comprehension, decision-making, attention, and problem-solving ability in participants' brain.	fMRI; Laboratory experiment	Academic Journal
(Silic et al., 2017)	This study tests the color risk perception in different culture, namely US and India. It shows that if warning messages can employ colors that fit the culture of users, the warnings can bring higher perceived risk and lead users to comply.	Survey; PLS-SEM	Conference

Table A2. Literature of Malvertising

Reference	Main Findings	Method	Publication Type
(Sood & Enbody, 2011)	Reviewing essentials of how malvertising spread and attack users, the article also discusses how to counter malvertising using appropriate verification, redirection control, service audition, and system update.	Reviews and opinions	Journal
(Xing et al., 2015)	The article develops a program to investigate the relationship between ad-injecting browser extension and malvertising. The results show that several ad-injection browser extensions on Chrome store could lead users to receive malvertising while using their browser.	Empirical study	Conference
(Pathak, 2016)	Reviewing both malware and malvertising, this article discusses several channels that malvertising might use to spread, and it suggest that individual and business should both aggressively counter malvertising.	Reviews and opinions	Academic Journal
(Dwyer & Kanguri, 2017)	Reviewing how malvertising spreads through real time bidding mechanism of online advertising and what risks might be caused by malvertising to online advertising and online ecosystem.	Reviews and opinions	Academic Journal
(Kumar et al., 2017)	Suggesting solutions for the malvertising. It suggests that the malvertising issues should be handled by the online advertisement providers. The providers should detect malvertising in the advertisement inventory.	Reviews and opinions	Conference
(Zabyelina, 2017)	From case study, this study argues that malvertising might be an opportunity for criminals to sell and promote substandard, spurious, falsely labelled, falsified and counterfeit (SSFFC) medical products.	Case Study	Academic Journal
(Huang et al., 2018, 2017)	Using game theory, this study simulates that what malvertising inspection strategies will defenders employ while dealing with malvertising issues. The results show that depending on different detection rate, ratio of malicious advertisers, and ratio of advanced malicious advertisers, how different malvertising inspection strategy will be carried out.	Game Theory; Numeric Simulation	Academic Journal; Conference

Appendix B: RGB and HSL color model

1. RGB model:

RGB define one color in 1 set of 3 numbers, and each number represents the degree of one of the three primary colors, red, green, and blue (Kalist et al., 2015). And each number can ranges from 0 to 255.

1st number: Red

2nd number: Green

3rd number: Blue

We use Google's color picker service to demonstrate the manipulation of color in the RGB model.

For example, (255, 0, 0) means that we maximize the degree of red and minimize the degree of green and blue in this color. The output color will be the following:

Figure B1. Color Picker Output of (255, 0, 0) (Source: Google color picker)

Or we maximize the degree of green and minimize the degree of red and blue in this color (0, 255, 0). The output color will be the following:

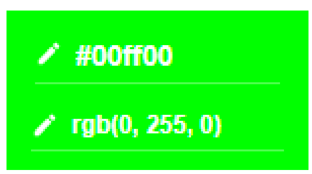


Figure B2. Color Picker Output of (0, 255, 0) (Source: Google color picker)

Following the same logic, we can choose to maximize 2 colors and minimize the 1 color to generate different colors. Thus, we will have the most basic 6 colors in the RGB model. I listed the colors I will employ in my research in the following table:

Table B1. RGB model and 6 basic colors

(R, G, B)	Color	Description
(255, 0, 0)	Red	Max Red, Min Blue, and Green
(0, 0, 255)	Blue	Max Blue, Min Red and Green
(0, 255, 0)	Green	Max Green, Min Red and Blue
(255,255,0)	Yellow	Max Red and Blue, Min Green
(255,0,255)	Magenta	Max Red and Green, Min Blue
(0,255,255)	Cyan	Max Blue and Green, Min Red

2. HSL model:

HSL is another way for a computer to define colors, which is based on three parameters, hue, saturation, lightness (Kalist et al., 2015).

H: Hues, defining the basic color hue, ranges from 0 to 360. The 0 equals to 360, which makes the hue a complete color circle.

S: Saturation, ranges from 0 to 100, the higher it is, the color it defines will be more closed to the original color (hue), the lower it is the color it defines will be more closed to the gray.

L: Lightness, ranges from 0 to 100, the higher it is, the color it defines will be more closed to the white. And the lower it is, the color it defines will be more closed to black.

Hues will define the basic color and we can use saturation and lightness to adjust the color based on the same color hue.

Following is a cylinder that shows the HSL color model.

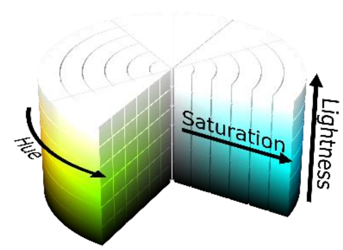


Figure 3. cylinder of the HSL model (“HSL and HSV,” 2017)

Appendix C: Estimated models under two biological genders

Table C1. Analysis results of the RGB model (Female)

RGB \ Scenario	Warning (US) R^2= 0.094	Malvertising (US) R^2= 0.025	Warning (TW) R^2= 0.076	Malvertising (TW) R^2=
Red	**3.505*** (exp(b)=33.278, p=0.000)**	-0.0491 (exp(b)=0.952, p=0.839)	**2.878 *** (exp(b)=17.777, p=0.000)**	**-1.370*** (exp(b)=0.254, p=0.000)**
Green	**-4.118*** (exp(b)=0.016, p=0.000)**	**-1.961*** (exp(b)=0.141, p=0.000)**	**-4.383 *** (exp(b)=0.022, p=0.000)**	**-1.519*** (exp(b)=0.219, p=0.000)**
Blue	**0.737*** (exp(b)=2.09, p=0.006)**	**2.335*** (exp(b)=10.329, p=0.000)**	**0.598 ** (exp(b)=1.818, p=0.032)**	**1.470*** (exp(b)= 4.348, p=0.000)**

Table C2. Analysis results of the the RGB model (Male)

RGB \ Scenario	Warning (US) R^2= 0.082	Malvertising (US) R^2= 0.041	Warning (TW) R^2= 0.113	Malvertising (TW) R^2=
Red	**3.534*** (exp(b)=34.2717, p=0.000)**	**-0.795*** (exp(b)=0.452, p=0.003)**	**3.906 *** (exp(b)=49.708, p=0.000)**	**-1.256*** (exp(b)=0.285, p=0.000)**
Green	**-3.473*** (exp(b)=0.031, p=0.000)**	**-0.956*** (exp(b)=0.384, p=0.000)**	**-4.866 *** (exp(b)=0.008, p=0.000)**	**-1.612*** (exp(b)=0.200, p=0.000)**
Blue	0.126 (exp(b)=1.134, p=0.653)	**1.966*** (exp(b)=7.142, p=0.000)**	**0.627 ** (exp(b)=1.872, p=0.013)**	**2.324*** (exp(b)=10.213, p=0.000)**

Table C3. Analysis results of the HSL model (Female)

SL \ Scenario	Warning (US) R^2= 0.019	Malvertising (US) R^2= 0.007	Warning (TW) R^2= 0.017	Malvertising (TW) R^2= 0.004
Saturation	**2.351* (exp(b)=10.492, p=0.061)**	1.594 (exp(b)=4.924, p=0.151)	**2.791** (exp(b)=16.300, p=0.034)**	0.628 (exp(b)=1.874, p=0.684)
Lightness	**-5.047*** (exp(b)=0.006, p=0.000)**	**-3.166*** (exp(b)=0.042, p=0.001)**	**-5.239*** (exp(b)=0.005, p=0.000)**	**-2.350* (exp(b)=0.095, p=0.071)**

Table C4. Analysis results of the HSL model (Male)

SL \ Scenario	Warning (US) R^2= 0.006	Malvertising (US) R^2= 0.01	Warning (TW) R^2= 0.014	Malvertising (TW) R^2= 0.004
Saturation	1.476 (exp(b)=4.375, p=0.203)	0.851 (exp(b)=2.342, p=0.482)	**3.081*** (exp(b)=21.789, p=0.006)**	0.804 (exp(b)=2.234, p=0.471)
Lightness	**-2.765*** (exp(b)=0.063, p=0.006)**	**-3.142*** (exp(b)=0.042, p=0.003)**	**-4.983*** (exp(b)=0.007, p=0.000)**	**-2.292** (exp(b)=0.101, p=0.016)**

CHAPTER FOUR

APPLYING DATA SCIENCE ON STRUCTURAL EQUATIONS MODELING (SEM): AN EXPLORATORY STUDY

Jian-Ren Hou

Sridhar Nerur

Jie Jennifer Zhang

Abstract

Structural equation modeling (SEM) is a commonly and widely adopted analysis to in-depth examine causality among latent variables, and SEM has been employed in numerous researches. SEM contains several statistical steps to perform rigorous estimations for the targeting research questions. This study tries to utilize the data science technique to support part of the procedure of SEM. Before distributing the survey, this study uses word embedding to convert the item into vectors which embedding semantic meanings. Based the word embedding vectors, this study tests the reliability (alpha) of the measurement model and detect possible common method bias in advance. After collecting the data, this study adopts cluster techniques in data science to test reliability. The results of this study can shed light on further SEM researchers.

Keywords: SEM, word embedding, doc2vec, GloVe, data science

1.Introduction

Structural equation modeling (SEM) is a popular research tool for researchers to probe into the human thinking process. This tool could provide an interpretable thinking sequence for a human being (Hair et al., 2009). However, the prerequisites of a reliable SEM are developing a reliable survey and meeting basic assumptions regarding statistic methods. Hence, this study aims to seek alternative and robust ways to verify and support the analysis from SEM.

Data science provides multiple methods to cluster similar objects together. Multidimensional scaling (MDS) and dendrogram can visualize the distances or hierarchy among objects by the similarity. By the visualization ability of these techniques, this study aims to answer the following research questions:

Before data collection,

1. Can the data science technique detect low-reliability (alpha) constructs in advance?
2. Can the data science technique detect the common method bias (CMB) in advance?

After data collection,

3. Can the data science technique support the results of confirmatory factor analysis (CFA) to show reliability?

For the first and second research questions, web embedding techniques such as Doc2Vec and GloVe can convert words or documents to vectors, and the vectors can embed semantic meaning and semantic similarity of words or documents. Thus, if this study can convert the text content of each item into vectors, then this study might be able to see if items are cohesive under each construct by data science cluster techniques. Based on the results, this study can detect possible low reliability (alpha) construct in advance. On the other hand, if this study considers all items under each construct as a document and converts the text content into vectors, and the results show the high similarity among different constructs, which might be able to detect possible common method bias in the survey design even before data collection.

For the third research question, after data collection, this study would like to apply the data science cluster techniques on collected data to see which item (questions, indicators) deviates more from others. The result could be additional support for CFA in SEM research.

This study contributes to the current research in the following ways. First, before data collection from pretest, pilot test, or actual test, SEM researchers have limited methods to examine the design of the survey. This study provides a more tangible and practical method to detect possible low-reliability construct or even a possible common method bias. Second, after data collection, factor analysis is employed by researchers to verify the reliability of the measurement. However, factor analysis requires statistical assumptions to be met to perform valid results. The data science techniques perform by this study can serve as a support method and a robustness check for the current factor analysis. Third, this study sheds light on using contemporary data techniques to enhance traditional statistical-based analysis. Future researches can further extend the idea of this study to other research methods.

This study has constructed this study as followings. In section 2, this study will briefly review the SEM analysis and the data science technique this study will employ in this study. In section 3, this study will employ data science techniques to answer our research questions. In section 4, this study will discuss and provide implications for this study, and limitations and further researches will be discussed.

2. Literature review

2.1 Structural Equation Modeling

Structural equation modeling (SEM) is an advanced multivariate technique that is used to probe possible sequential causality among latent variables by statistic/mathematical methods (Hair et al., 2009; Kaplan, 2009). Scholars from various research areas, such as psychology (Breckler, 1990; MacCallum & Austin, 2000), sociology (Golob, 2003), marketing (Fornell & Bookstein, 1982), information system (Urbach & Ahlemann, 2010), and environmental epidemiology (Sánchez et al., 2005), have adopted SEM in their studies. A typical SEM analysis usually consists of the measurement model and the structural model.

According to Hair et al., (2009), the measurement model represents the relationship between measurement items (questions, indicators) and latent variables, and measurement model can be used to detect the possible measurement error. Normally, in a survey-based research SEM analysis, there will be several questions used to measure one construct (latent variable). Measurement model would be used to describe the relationship among these items to the latent variable. Moreover, CFA (confirmatory factor analysis) technique would be used to categorize all these items to its construct, and factor-loadings and eigenvalue would be calculated to see if items are cohesive to other items which are belonging to the same latent variable. For example, it detects if the four questions are used to measure one latent variable by calculating the factor-loadings for each item. If one question's factor loading is far less than other questions, the low factor loading question might be removed due to the possibility that the question does not belong to the construct or an error has occurred.

2.2 Multidimensional Scaling (MDS) and Dendrogram

In multivariate analysis, multidimensional Scaling can visually create a map to represent the similarity among the object in the dataset (Kruskal & Wish, 2009). The 2-D map generated by MDS can show the similarity between objects by the distance on the map. A shorter distance implies a higher similarity between data points. MDS is widely adopted by different research areas such as marketing, political issues, or cultural issues (Hair et al., 2009).

On the other hand, a dendrogram is also a frequently used technique to represent the relationship between objects hierarchically (Hair et al., 2009). The tree diagram can categorize similar objects under the same or nearest branches. Dendrogram can hierarchically display the similarity between objects by a tree diagram, and literature mentions that dendrogram can be used to detect outliers in multivariate analysis (Hair et al., 2009).

2.3 Word embedding

To represent the similarity between the textual context in the computer science area, scholars try to use bag-of-word (BOW) and term frequency-inverse document frequency (TF-IDF). Although BOW and TF-IDF can be used to classify different documents, these frequency-based techniques are failed to capture the semantic similarity between words (Kusner et al., 2015). Mikolov et al., (2013) introduce the word2vec model which can use pre-trained neural network model to convert words to vectors, and the vectors can capture semantic meaning of word by learning the words which frequently co-occurs from a large size of corpora, for example, a dataset from Google news which contains about 100 billion words. Researchers later extend the word2vec model to capture not only the semantic meaning of words but also the semantic meaning of paragraphs or documents on textual content, which is the doc2vec model (Lau & Baldwin, 2016; Le & Mikolov, 2014). Similarly, Pennington et al. (2014) develop the GloVe model, which also uses a different static model to learn the co-occurrence of words and convert the textual content to vectors with semantic meaning.

3. Study

3.1 Environment

This study uses two different word-embedding models to convert textual content to vectors in this study, which are doc2vec (Le & Mikolov, 2014) and GloVe (Pennington et al., 2014). Practically, this study uses the pre-trained doc2vec model trained by Lau & Baldwin (2016), the pre-trained model was trained based on a news dataset which contains 0.9 billion tokens. For the GloVe model, this study uses the pre-trained GloVe model provided by Explosion, the developer of the spacy NLP library, and the model was trained based on news, blog, and comments. All the textual contents are being preprocessed before the actual word embedding convert process. The preprocess includes converting textual contents to lower cases, removing punctuations, and removing stopwords. The following tools are used in this study.

- Python 3.6
- genism for converting words or documents to vectors (doc2vec)
- spacy for converting documents to vectors (GloVe)
- nltk for removing stopwords
- sklearn for MDS and kmeans clustering
- scipy for dendrogram clustering
- matplotlib for visualization
- jupyter notebook

3.2 Study 1 – detecting low-reliability constructs

In study 1, this study aims to answer the first research question: Before data collection, can the data science technique detect low-reliability constructs in advance? This study uses doc2vec and GloVe to convert items under each construct into vectors and use data science cluster techniques to show the similarity among items. If the items used in a construct shows a long distance from other items under the same construct, this might imply possible low reliability (alpha).

The survey items are adopted from Kuo & Hou (2017). All items are listed in the following table:

Table 1 Survey Items

Construct	Indicator	Measurement	References
Brand community identification	BCI_1	I usually use "we" instead of "they" when I talk about Brand Community XYZ.	Algesheimer et al. (2005) Zhou et al. (2012)
	BCI_2	I see myself as a part of Brand Community XYZ.	
	BCI_3	I take the success of Brand Community XYZ as mine.	
	BCI_4	I feel complimented when I hear good remarks about Brand Community XYZ.	
	BCI_5	I feel insulted when I hear bad remarks about Brand Community XYZ.	
	BCI_6	I have a great interest in "how others think about Brand Community XYZ".	
Brand commitment	BC_1	I feel upset when I cannot get a product of Brand XYZ I want to buy.	Coulter et al. (2003) Raju et al. (2009)
	BC_2	I am a loyalty customer (user) of Brand XYZ.	
	BC_3	Brand XYZ is the best choice of mine.	
Self-brand connections	SBC_1	Brand XYZ reflects who I am.	Escalas and Bettman (2003)
	SBC_2	I perceive personal connections between Brand XYZ and myself.	
	SBC_3	I can use Brand XYZ to communicate who I am to others.	
	SBC_4	Brand XYZ is very suited for me.	

Table 1. Survey Items (Cont.)

Construct	Indicator	Measurement	References
Oppositional brand loyalty	OBL_1	I will not try any rival brand that offers similar products.	Kuo and Feng (2013)
	OBL_2	I have no interest in any rival brand even if it offers a diversity of products.	
	OBL_3	I will not consider buying products of any rival brand even if the products can better meet consumers' specific needs.	
	OBL_4	I will not consider buying products of any rival brand even if the products have better specifications.	
	OBL_5	I will not recommend products of any rival brand even if the products are generally considered better.	
	OBL_6	I will not try products of any rival brand even if the products are widely discussed by other people.	

This study uses doc2vec and GloVe to convert each textual item into a 300-D vector. Since the vector can represent the semantic meaning of items, this study then uses the data science cluster techniques, including MDS and dendrogram to visualize the similarity among items.

3.2.1 doc2vec model

In this section, this study uses the doc2vec model. The results are in the following Figures:

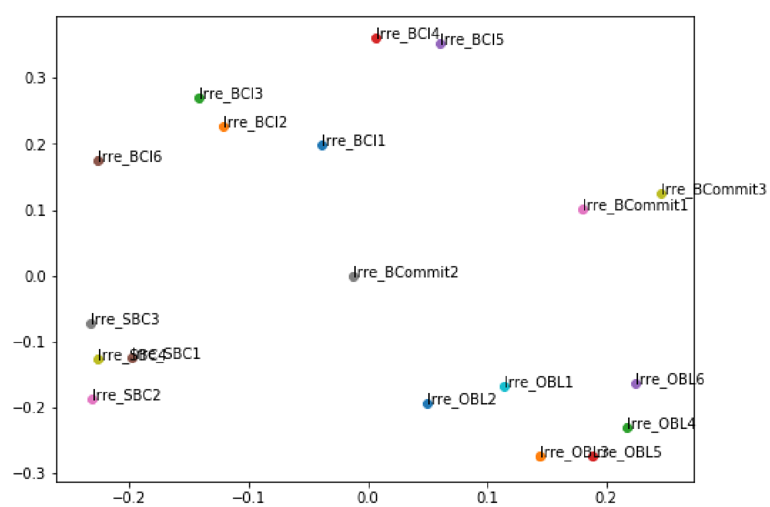


Figure 1 MDS on doc2vec model

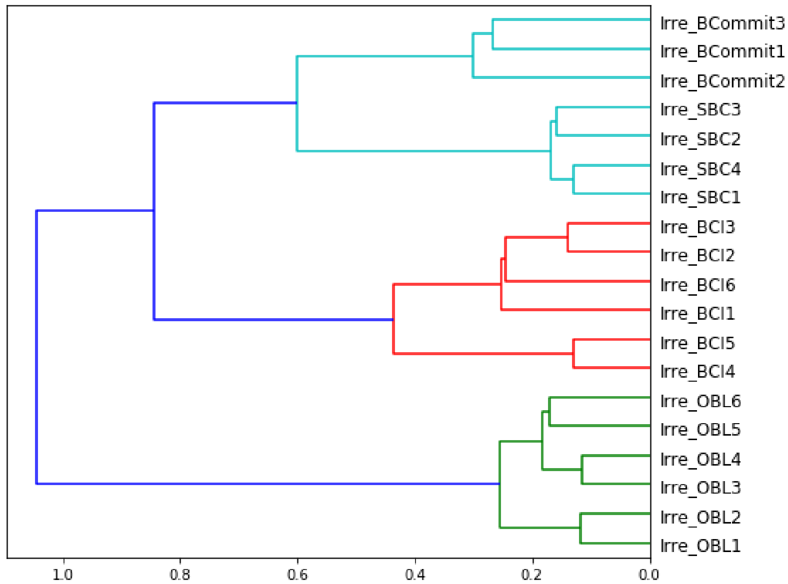


Figure 2 Dendrogram on doc2vec model

3.2.2 GloVe model

In this section, this study uses GloVe to convert items to vectors. The results are the following:

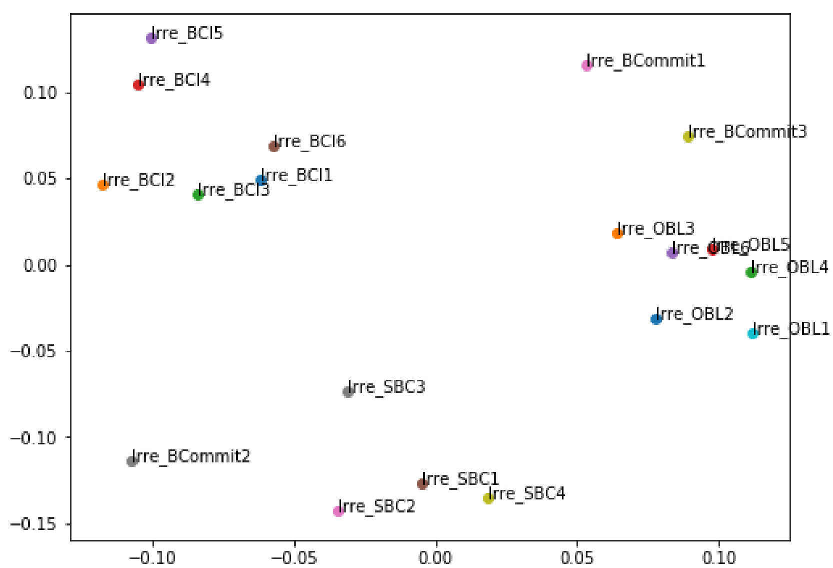


Figure 3. MDS on GloVe model

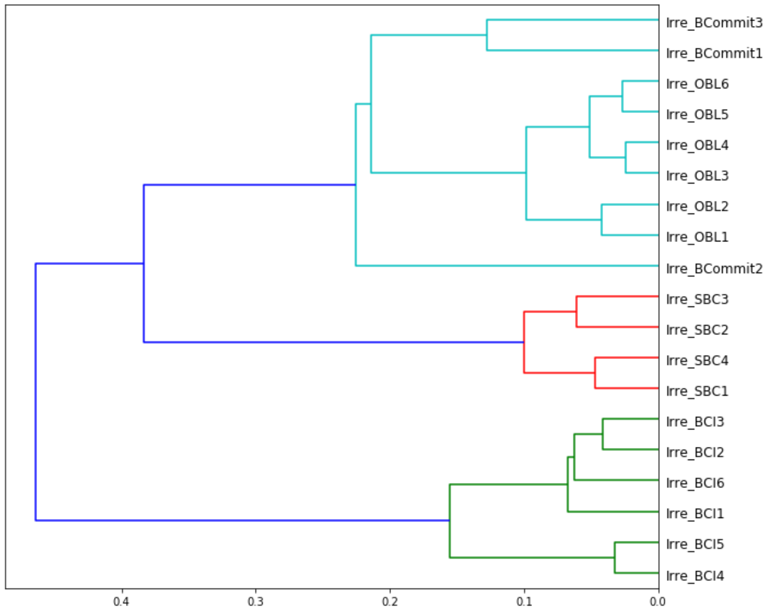


Figure 4. Dendrogram on GloVe model

3.2.3 Results from study 1

Based on the results and of MDS from doc2vec and GloVe model, readers can see that the distribution of the three items under the brand commitment construct (BCommit in the Figures) is more separated than other items under each construct. On top of that, the dendrogram from the GloVe model also categorizes one of the brand commitment items into a different branch, which indicates the deviation among the items of brand commitment construct. Therefore, the result shows that there is a possibility that the semantic meaning among the items which belong to the brand commitment construct is not consistent.

Table 2 Reliability of Kuo & Hou (2017)

Construct	Composite reliability	Cronbach's alpha
Brand community identification	**0.937**	**0.916**
Brand commitment	**0.898**	**0.828**
Self-brand connection	**0.941**	**0.917**
Oppositional brand loyalty	**0.955**	**0.944**

This study lists the reliability results from the paper of Kuo & Hou (2017). Although readers can see the composite reliability (CR) value and Cronbach's alpha both exceed the recommended threshold value of 0.7 (Bagozzi & Yi, 1988; Nunnally, 1978), the CR and alpha values of brand commitment is still slightly lower than other constructs. Moreover, our results from study 1 also indicate that the internal consistency of the brand commitment construct might be lower. Thus, the lower CR and alpha values of brand commitment might result from the inconsistently semantic meaning of brand commitment items. Moreover, the results also indicate that the semantic meaning of the second item of brand commitment (BCommit2) deviates from other items. If researchers can detect the deviation before data collection, they might be able to adjust the design of the item in advance to avoid possible lower CR or alpha values.

3.3 Study 2 – Detecting possible CMB in advance

The target of study 1 is to discover the similarity among items under the same construct. It is beneficial for SEM analysis to have high similarity among the items under the same construct, which might imply a higher reliability of the measurement. However, if discovering extremely high similarity among the different constructs might imply the possible common method bias (CMB). In this section, this study uses the word embedding techniques to convert all textual content under a construct into vectors, and the vectors can catch the semantic similarity among the constructs.

In study 2, this study would like to catch the constructs which are too similar in the same paper, which might imply a CMB issue in that paper. However, how to define an abnormally high similarity is critical for this study. Thus, this study tries to build a 2-D distribution map for numerous constructs from several papers to solve this issue. Practically, this study first gathers 30 articles from the top tier IS journals, this study collected 203 construct and its measurements from these research articles. This study then uses word embedding techniques to covert 203 constructs to corresponding 302 300-D vectors. By plotting and visualizing all 203 vectors for constructs on the MDS map, we can see the distribution of the semantic meaning of 203 constructs. Next, for each paper, this study calculates the Euclidean distance between every two constructs from all the constructs in the same paper, and this study calculates the average of all the Euclidean distances in the same paper. The average Euclidean distance for each paper can show how semantically cohesive are the constructs, and this can be an index for us to diagnose the possible CMB. This study uses bootstrapping 5000 times, and each time this study randomly draws 7 constructs from the MDS map since the average number of constructs for each paper is 6.77 (203 constructs / 30 papers). This study calculates the average Euclidean distance of the seven constructs and ends up with 5000 average Euclidean distances. These distances are used to build a distribution of the average Euclidean distances of the constructs. Finally, if we try to test the degree of similarity among all constructs for a specific paper, we can compare the average Euclidean distance of that paper to the 5000 average Euclidean distances. Assuming this study predefines the confidence level as 95%, if the average Euclidean distances of the paper are smaller than 95% of the 5000 bootstrapping average Euclidean distances, we might be able to say that the constructs for the paper are way too cohesive, which might imply a possible CMB issue.

In the next two sections, this study will use two different word-embedding models to perform the above test. The list of all 30 papers used in this section is provided in the supplementary file.

3.3.1 doc2vec

This study uses the doc2vec model in this section to perform the test proposed.

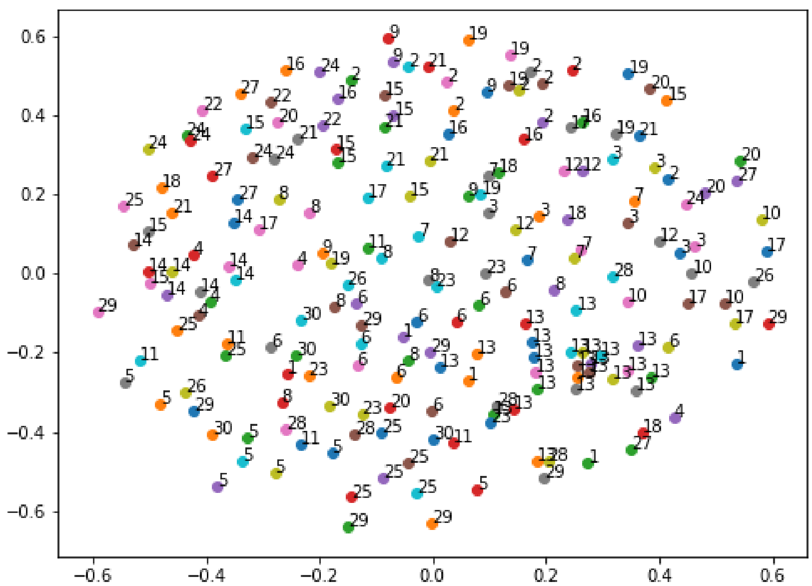


Figure 5. MDS on doc2vec model

In Figure 5, this study utilizes the MDS technique to plot all the word embedding vectors from the 203 constructs in the dataset. The numbers on the MDS map are the serial numbers used in this paper to encode and distinguish each paper, the constructs from the same article will have the number on the dot. The distance among the dots on the MDS map can catch the semantic similarity of constructs used in each paper.

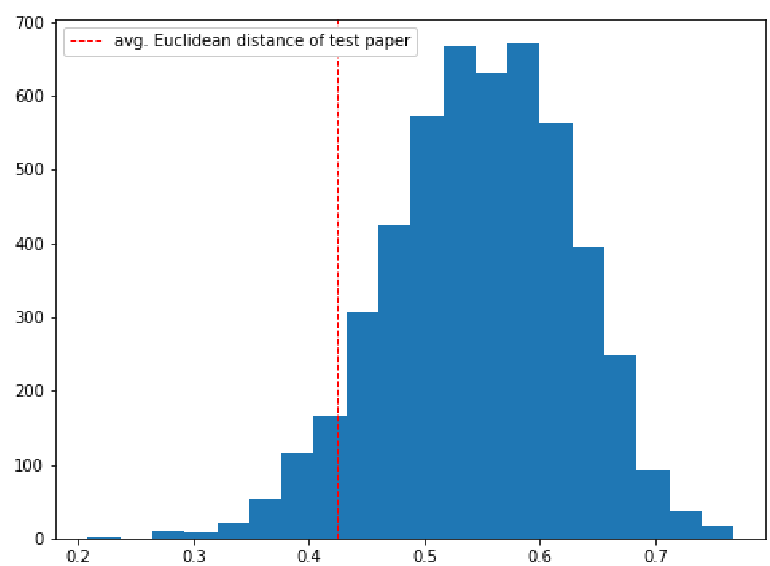


Figure 6. Distribution of average Euclidean distances (Bootstrapping)

In the bootstrapping distribution in Figure 6, this study plots the average Euclidean distance of the paper we would like to test as a red line. Furthermore, this study calculates how many times the average Euclidean distance of the test paper is smaller than the 5000 bootstrapping as the p-value of this test. The p-value for this test is 0.0656, which is not significant with the 95% confidence level.

3.3.2 GloVe

In this section, this study conducts an identical procedure with section 3.3.1 but replacing the doc2vec model with the GloVe model.

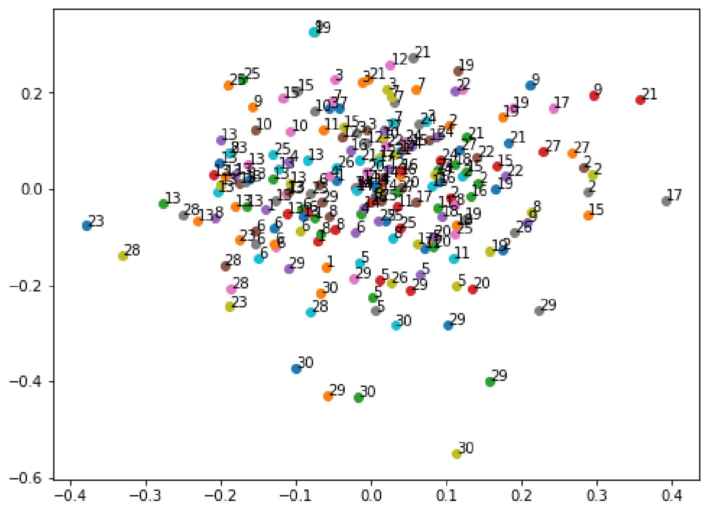


Figure 7. MDS on GloVe model

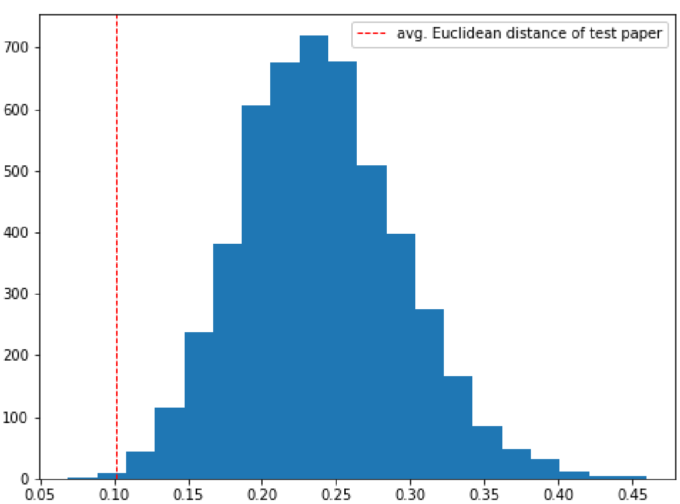


Figure 8. Distribution of average Euclidean distances (Bootstrapping)

The p-value calculated based on the above bootstrapping is 0.0012. It shows a significantly cohesive of the constructs from the test paper with a confidence level of 95%. The results are different from the result of section 3.3.1.

3.3.3 Results from study 2

In study 2, this study employs word embedding and MDS techniques to catch the significantly cohesive constructs. This study argues that this method might help SEM researchers to catch the possible CMB in advance. However, the result based on the GloVe model is different from the test result based on the doc2vec model in section 3.3.1. The results are different due to the two different word embedding techniques. However, if there is any possibility that CMB exists in the study, this study would like to suggest SEM researchers taking a cautious approach to address the issue of CMB in the survey design.

3.4 Study 3 – Supporting factor analysis in SEM

3.4.1 Data

The dataset and research model are both adopted from Kuo & Hou (2017). The dataset consists of a survey with 232 subjects. To input the data properly into data science techniques, this study rotates the data. Therefore, the data is now consisting 19 questions, which are categorized into four constructs, and each question contains 232 responses from 232 different participants, and all the 233 responses are considered as features of each item. By inputting this rotated dataset, this study should be able to see the relationship among items.

3.4.2 Cluster on the measurement model

For this section, this study would like to see if the results are similar to those of the traditional SEM approach of the measurement model. Since the results could indicate the similarity among questions, this study expects to compare the results to the factor-loadings from the research (Kuo & Hou, 2017). All factor loadings for each item under each construct are in the following table.

Table 3. Factor loadings table from Kuo & Hou (2017)

Construct	Indicator	Loading
Brand community identification	BCI_1	0.763
	BCI_2	0.858
	BCI_3	0.838
	BCI_4	0.913
	BCI_5	0.852
	BCI_6	0.839
Brand commitment	BC_1	0.832
	BC_2	0.876
	BC_3	0.882
Self-brand connection	SBC_1	0.901
	SBC_2	0.921
	SBC_3	0.925
	SBC_4	0.827
Oppositional brand loyalty	OBL_1	0.886
	OBL_2	0.927
	OBL_3	0.900
	OBL_4	0.888
	OBL_5	0.823
	OBL_6	0.875

Low factor loadings might indicate the item does not relate to other items and the construct itself correctly, which also indicates the possible deviation of the item. Thus, this study expects to see the results that show that the lower factor loading questions deviate from other questions on the MDS map and dendrogram.

3.4.3 Cluster on the real data

Different from Study 1 and Study 2 that use vectors generated by word embedding technique, Study 3 uses the responses data from real participants of the Kuo & Hou (2017). This study inputs the data into MDS and dendrogram to plot the following figures.

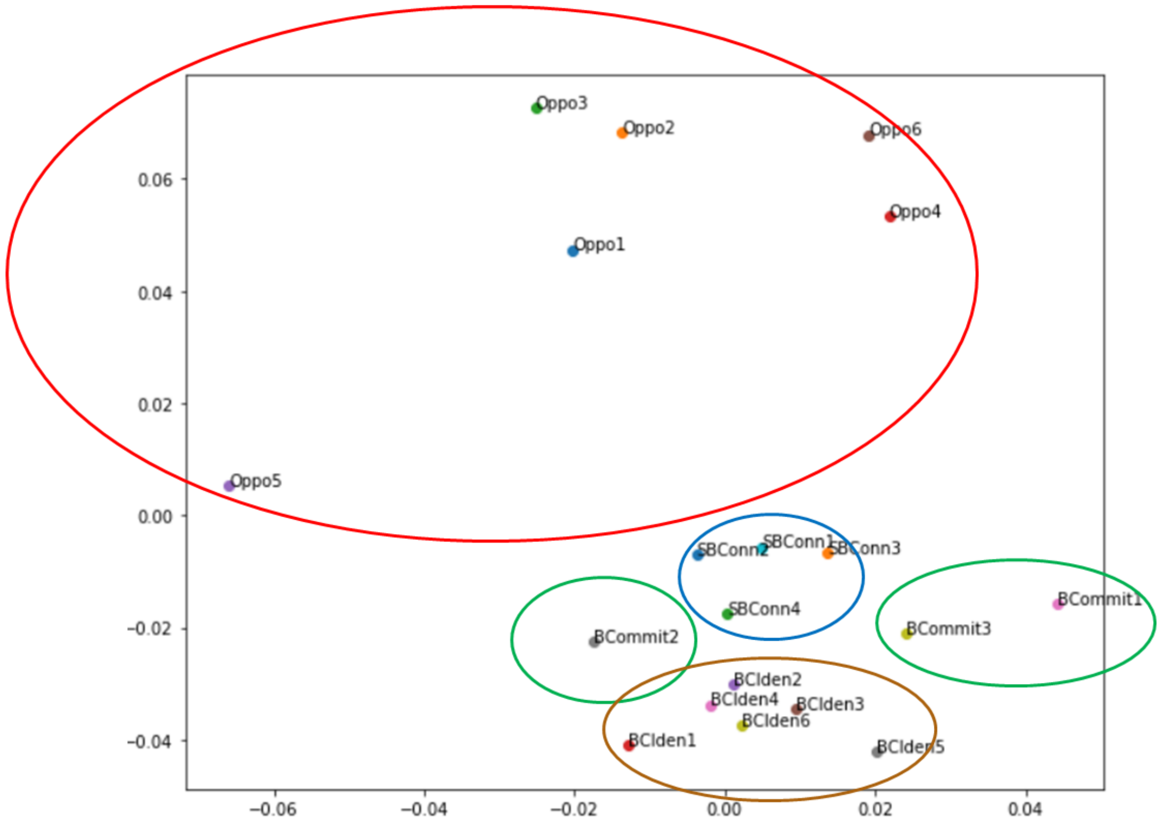


Figure 9. MDS on real data

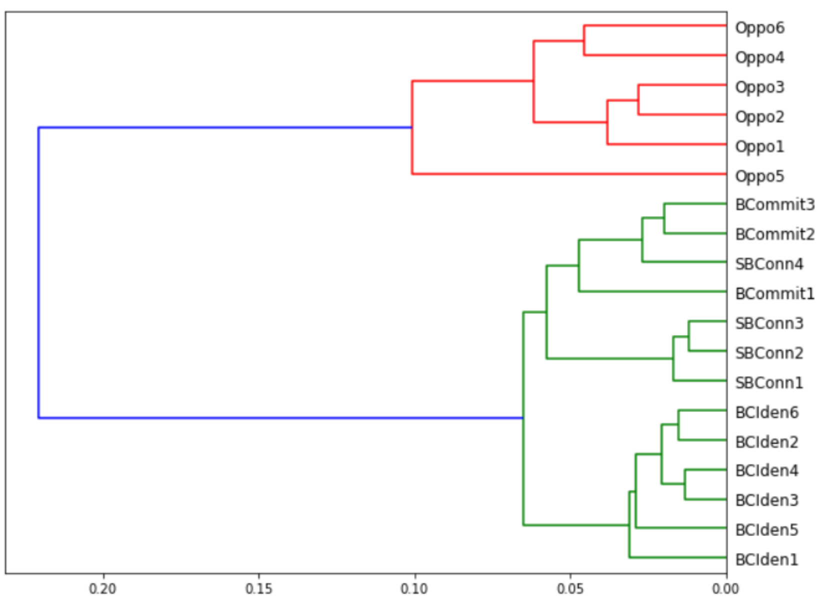


Figure 10. Dendrogram on real data

The results show that the SOM does show a similar result with the traditional statistic method (factor loading). The deviated points on each technique also show lower factor loading in the table in the research. For example, the 5th item is under the oppositional brand loyalty constructs. Thus, this result might show it supports the measurement model of SEM aside from other techniques. On top of that, since sometimes the statistical assumptions of factor analysis might not be met, the MDS and dendrogram could not only support the results of factor analysis but also be served as an alternative method to detect low factor loading items.

4. Discussion and conclusion

In Study 1, this study demonstrates how to employ word embedding techniques to detect possible low-reliability constructs by catching the semantically deviated items under each construct. The visualization techniques allow us to find these semantically deviated items before data collection. The results are consistent with those of Kuo & Hou (2017). The construct with semantically deviated items does show relatively lower reliability in the above paper. The techniques used in the study can help researchers catch possibly problematic items. However, since the word embedding technique can only catch the semantic meaning of textual contents, the techniques this study used in Study 1 are not able to catch other issues, which may cause low reliability such as layout design of the survey, displaying order of items in the survey, or scenario setting of the survey. Researchers should employ these techniques with cautious.

In Study 2, this study demonstrates a way to detect extremely and semantically cohesive constructs in SEM analysis. This can be a technique for SEM researchers to employ before distributing a survey. Using the technique in advance, the possibility that CMB caused by being extremely semantic close among the constructs can be lower if researchers catch these issues earlier and modify the design before data collection.

Study 3 provides an alternative way for SEM researchers to conduct measurement model analysis. This study expects that the technique can not only support the results of factor analysis but also possibly detect low factor loading items which factor analysis failed to catch due to unmet statistical assumptions.

In summary, these data science techniques shed light on supporting the widely adopted SEM analysis. The methodology suggested by Study 1 and Study 2 can be used for catching the possible issues in the survey design, and researchers can improve the design before data collection to avoid possible negative effects brought by survey design. The methodology suggested by Study 3 can be used to support the current measurement model analysis. It is worth mentioning that the methodology proposed by this study does not aim to replace the conventional statistical methodology, this study considers the methodology as extra improvements for the SEM analysis.

5. Limitation and future research

Despite our effort, the results of this research are constrained by the following limitation. This study only tests the methodology on the results of Kuo & Hou (2017). Future research can test the generalizability of the methodology proposed by using more SEM papers. The accuracy of word embedding highly relies on the quality of the pre-trained model. The doc2vec and GloVe models used in this study are trained by the news, Wikipedia, or blog post, which might limit its generalizability to text content in the academic area. Future research can further train the model with research article content, which might improve in accuracy compared to the model this study used in this study. In Study 2, limited by the labor power, this study uses the bootstrapping technique to resample from 30 papers to form the distribution and test the significance. Future research should gather more papers than 30 papers to form the distribution without bootstrapping. This should provide a more accurate distribution, which allows us to estimate better how cohesive the constructs are in the same paper to detect CMB.

References

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94. https://doi.org/10.1007/BF02723327

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, *107*(2), 260–273. https://doi.org/10.1037/0033-2909.107.2.260

Fornell, C., & Bookstein, F. L. (1982). Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. *Journal of Marketing Research*, *19*(4), 440–452. https://doi.org/10.1177/002224378201900406

Golob, T. F. (2003). Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological*, *37*(1), 1–25. https://doi.org/10.1016/S0191-2615(01)00046-7

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate Data Analysis* (7 edition). Upper Saddle River, NJ: Pearson.

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed). Los Angeles: SAGE.

Kruskal, J. B., & Wish, M. (2009). *Multidimensional scaling* (Nachdr.). Newbury Park, Calif.: Sage Publ.

Kuo, Y.-F., & Hou, J.-R. (2017). Oppositional Brand Loyalty in Online Brand Communities: Perspectives on Social Identity Theory and Consumer-Brand Relationship. *Journal of Electronic Commerce Research; Long Beach*, *18*(3), 254–268.

Kusner, M. J., Sun, Y., Kolkin, N. I., & Weinberger, K. Q. (2015). *From Word Embeddings To Document Distances*. 10.

Lau, J. H., & Baldwin, T. (2016). An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. *ArXiv:1607.05368 [Cs]*. Retrieved from http://arxiv.org/abs/1607.05368

Le, Q., & Mikolov, T. (2014). *Distributed Representations of Sentences and Documents*. 9.

MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modeling in Psychological Research. *Annual Review of Psychology*, *51*(1), 201–226. https://doi.org/10.1146/annurev.psych.51.1.201

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *ArXiv:1310.4546 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1310.4546

Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Sánchez, B. N., Budtz-Jørgensen, E., Ryan, L. M., & Hu, H. (2005). Structural Equation Models: A Review With Applications to Environmental Epidemiology. *Journal of the American Statistical Association*, *100*(472), 1443–1455. https://doi.org/10.1198/016214505000001005

Urbach, N., & Ahlemann, F. (2010). Structural Equation Modeling in Information Systems Research Using Partial Least Squares. *JITTA : Journal of Information Technology Theory and Application*, *11*(2), 5.

CHAPTER FIVE

Conclusion

This dissertation provides both theoretical and practical values for studying online human decision-making. In the first and second essays, this study investigates how visual cues, such as images and colors, are related to online human decision-making, such as participating in crowdfunding projects and dealing with cybersecurity issues. In the third essay, this study proposes a new methodology for SEM researchers to better employ SEM analysis to investigate complicated psychological issues of online human decision-making.

In the first essay, using deep learning techniques to extract emotions from the project images, this study further uses the emotions to predict the outcomes of crowdfunding projects. The results indicate that sadness and contentment positively affect the performance of crowdfunding projects, and fear negatively affects the performance of crowdfunding projects. On top of that, the first essay also provides practical guidelines for practitioners to design the project images for desired emotions. Moreover, the first essay employs experiments to verify our results from empirical studies. The experimental results show that image attribute modifications affect the emotional feelings of human subjects. From the data this study collected for the experiments, this study also observes the significant correlation between emotions and intention to pledge crowdfunding projects.

In the second essay, this study employs choice-based conjoint analysis to test the effect of color elements, such as red, green, blue, saturation, and lightness, on message acceptance of warning messages and malvertising. Using conditional logistic regression to analyze data, the second essay not only indicates whether the effect of each color element positively or negatively relates to the message acceptance but also how strong the effect is from each color factor. The results provide insights for both the designer for effective warning message design and the practitioners to deal with the malvertising issues, which is rarely investigated.

The third essay employs data science techniques to improve and support the current SEM procedure. Based on the word embedding techniques, the third essay provides practical guidance for SEM researchers to avoid possible common method bias and low reliability in the survey design before any actual data collection. After data collection, the cluster techniques in data science can also serve as an alternative way to perform factor analysis and to support the current factor analysis results.

In summary, this dissertation shows how to employ innovative techniques to perform analysis and to answer research questions. In the first essay, this study uses deep learning to extract emotion from images, while the traditional way to measure emotions is based on human subjects' self-reported response. The first essay sheds light on future research to employ deep learning techniques to enrich current empirical studies with more objective measurements. In the prior studies, the choice-based conjoint analysis is employed in testing the effect of different merchandises' attributes, and the prior color-related studies are focusing on the negative or positive effects of colors without measuring the effect size of color factors. In the second essay, this study uses choice-based conjoint analysis to test the effect of color factors, the results show not only the effect but also the effect size of color factors. The methodology and results of the second essay can provide an innovative way for future researchers to study color-related issues. In the third essay, data science techniques provide alternative ways to improve and support the current SEM procedure. This study hopes this dissertation can be an example of employing state-of-the-art techniques to support or improve current research methodology, and this study encourages future studies can adopt these innovative techniques to enhance their research further as well.