Algorithms and User Behaviors

Yifan Yu

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Yong Tan, Chair

Ming Fan

Apurva Jain

Program Authorized to Offer Degree:

Business Administration

University of Washington

Abstract

Algorithms and User Behaviors

Yifan Yu

Chair of the Supervisory Committee:
Yong Tan
Department of Information Systems and Operations Management

This dissertation delves into the burgeoning and critical field of artificial intelligence (AI), focusing on the increasingly intricate interaction between sophisticated algorithms and complex user behaviors. In an era of digitalization and AI, understanding the interaction between algorithms, particularly those grounded in AI, and user behavior, has become a necessity rather than a luxury. The vast and diverse applications of AI algorithms have now permeated various aspects of human life, predicting and affecting behaviors in ways previously unforeseen. Yet, the extent to which AI algorithms predict, shape, and in turn are shaped by, user behaviors remains a vastly unexplored frontier. As such, this dissertation seeks to elucidate these interactions, bridging the gap in knowledge and contributing to the optimal design and effective use of AI algorithms in various industries. The first essay studies how emotion-aware algorithms can be

developed to advance predictions of user behaviors. After developing an algorithm for multi-dimensional emotion detection in texts and conducting predictive analyses, the chapter presents a laboratory experiment study, underscoring the causal mechanisms through which the algorithm generates predictive power to user behaviors. The second essay combines video analytics algorithms and econometric models to understand user attention and clicking behaviors to video advertising. After empirically demonstrating that algorithms can be used to predict and understand user behaviors in various business contexts, the third essay focuses on the scenario where algorithms may shape user behaviors, particularly strategic behaviors. By formulating a game-theoretical model, it identifies the conditions conducive for adopting emotion-aware AI and determines the optimal allocation policies. The essay also delves into the welfare impacts of emotion AI on individuals, organizations, and society at large. The results suggest that a stronger AI is not always socially desirable and necessitates regulation on data-driven allocation. Moreover, it outlines scenarios in which AI adoption can be more profitable than employing human labor for emotion recognition and resource allocation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Professor Yong Tan. Prof. Tan has been a great mentor and role model for me. He has guided me all the way through my Ph.D. study and provided me with consistent support. His wisdom and enthusiasm for research are admirable and have greatly influenced me. He provides tremendous care to all his students. I feel fortunate to have Prof. Tan as my academic advisor. And I personally cannot imagine a better advisor.

My thanks also go to the other committee members, Professor Ming Fan, Professor Apurva Jain, Professor Quan Wen, and Professor Yingfei Wang. Their insightful comments and constructive suggestions have helped me to improve this dissertation. I have also learned a lot in research expertise, teaching professionality, and from the fantastic courses that they have provided.

I would also like to express my appreciation to staff at the Foster School of Business for their generous support during my study. My gratitude extends to all fellow doctoral students who have shared this journey and provided me with long-standing help.

I sincerely thank my parents (Xinmin Yu and Jingzhi Tu) and the whole family for their unconditional love and support. My wife Wendao Xue and our baby Dawson Yu are the permanent source of inspiration and courage for me. Wendao also greatly helped me in research with her economic thinking and expertise in econometric theories. I have always been very proud of my brother Qixuan Yu, who is a chemistry PhD graduating from Yale University. We have supported each other in the journey of life since childhood.

DEDICATION

To Dawson. Be the change that you wish to see in the world.

Chapter 1. INTRODUCTION

The burgeoning field of artificial intelligence (AI) has become integral to our digital era, with the increasingly intricate relationship between advanced algorithms and complex user behaviors emerging as a critical area of study. This dissertation aims to understand this interaction between algorithms and user behavior in various business contexts. I particularly focus on algorithms with the capacity to understand unstructured data, including text, images, and videos. These algorithms include deep neural networks, language models, and computer vision models. Notably, this focus is on par with the recent trend of developing and applying artificial general intelligence (AGI) in business because many algorithms used or discussed in this dissertation, including deep neural networks and language models, serve as building blocks of AGIs.

From a broader perspective, understanding algorithm-user interactions is no longer luxury, but rather necessity for business researchers and practitioners, because advanced and complex algorithms have deeply permeated numerous facets of our lives, predicting and affecting our behaviors in ways that we could not have previously imagined. However, despite the widespread influence of AI and unstructured-data-driven algorithms, there remains a considerable gap in our understanding of the extent to which these algorithms predict, influence, and in turn are influenced by, our behaviors. This is particularly significant as business researchers and practitioners still find themselves in the preliminary stages of discovering this intriguing landscape. Therefore, through this dissertation, the aim is to shine a light on these complex interactions. In doing so, the goal is not only to augment our understanding but also to contribute to the refinement of AI and unstructured-data-driven algorithm design and the effective implementation of these algorithms across various contexts, including online reputation management, digital advertising, customer services, managerial responses, and product sales.

Chapter 3 focuses on emotion artificial intelligence, the algorithm that recognizes and interprets various human emotions beyond valence (positive and negative polarity), which is still in its infancy but has attracted attention from industry and academia. Based on discrete emotion theory and statistical language modeling, this chapter proposes an algorithm to enable automatic domain-adaptive emotion lexicon construction and multi-dimensional emotion detection in texts. With a large-scale dataset of China’s movie market from 2012 to 2018, I construct and validate a domain-specific emotion lexicon and demonstrate the predictive power of eight discrete emotions (i.e., surprise, joy, anticipation, love, anxiety, sadness, anger, and disgust) in online reviews on box office sales. Results show that representing overall emotions through discrete emotions yields higher prediction accuracy than does valence or latent emotion variables generated by topic modeling. To understand the source of the predictive power from a theoretical perspective and to test the cross-culture generalizability of the prediction study, this work further conducts an experiment in the U.S. movie market based on theories on emotion, judgment, and decision making. Results show that discrete emotions, mediated by perceived processing fluency, significantly affect the perceived review helpfulness, which further influences purchase intention. This chapter shows the economic value of emotion AI in the context of online reviews and movie sales, generates insight into the mechanism of predictive and causal effects, and has managerial implications for online review platform design, movie marketing, and cinema operations.

Chapter 4 aims to understand user behaviors by combining econometric models and deep learning algorithms (including computer vision, video analytics, and natural language processing). This chapter focuses on the context of digital advertising because it provides both multi-modal data of advertisements (ads) and click-stream data of user behaviors. In particular, this chapter concentrates on outstream video advertising, a new form of video advertising that is becoming increasingly popular in the industry recently. Differing from the instream video ads embedded in video content, outstream video ads auto-play in non-video environments when a user navigates to them. On online shopping websites and apps, sponsored outstream video ads target shopping queries and occupy high-visibility placements. This chapter investigates the effectiveness of video ads in driving user attention and clicking behaviors, by conducting a large-scale, query-level observational study on click-stream data. Results show that video ads attract consumer attention and are more effective when products are less differentiated from each other in a market. Contingent on consumer attention, video content features that facilitate efficient consumer learning or signal product quality significantly increase consumers' likelihood of clicking a product. This chapter contributes to the literature on video advertisement and video analytics. It provides important implications for practitioners to understand consumer decisions and advertising effectiveness, create effective outstream video ads, and improve video ad recommendation systems.

After empirically demonstrating that algorithms can be leveraged to predict and understand user behaviors in Chapters 3 and 4, the third essay in Chapter 5 focuses on the scenario where algorithms may shape user behaviors, particularly strategic behaviors. I choose to continue with the context in Chapter 3, the adoption of emotion AI. When organizations adopt emotion AI to recognize individuals' negative emotions and accordingly match limited resources to them, strategic users are incentivized to game the system by escalating emotional intensities. The economic value of AI may be undermined by gaming behavior, algorithmic noise in emotion detection, and the spillover effect of negative emotions. This chapter develops a game-theoretical model to identify the conditions under which adopting emotion AI is valuable to the organization and design the associated optimal allocation policies. Analytical results show that adopting emotion AI is always valuable if the spillover effect of negative emotions is negligible compared to resource misallocation loss, regardless of algorithmic noise and gaming behavior. This chapter also quantifies the welfare impacts of emotion AI on individuals, the organization, and society. Notably, results show that a stronger AI is not always socially desirable and regulation on data-driven allocation is needed. Finally, this chapter characterizes a broad set of conditions under which adopting emotion AI is more profitable for the organization than hiring employees to recognize emotions and allocate resources. In sum, this work provides implications for designing, adopting, and regulating emotion AI.

Chapter 2. LITERATURE REVIEW

This chapter has two parts. The first part relates to the first and the third essays and reviews the literature on emotions and AI. In particular, I review the theories of emotions, which serve as theoretical foundation for algorithm design and business applications. Then, I review the literature on the economic significance of emotions in an important business context, online reviews and sales. After illustrating the theoretical and economic significance of emotions, I review the design, application, and regulation of emotion AI. Finally, I review the work on strategic interaction between AI and its users.

The second part relates to the second essay, aiming at reviewing the literature on consumer behavior toward digital advertisements. Specifically, I review the literature on consumer attention and clicking behaviors in digital advertising. Then, I review the literature on the elaboration likelihood model and video content analytics.

2.1 EMOTIONS AND AI

2.1.1 *Theories of Emotions*

There are three fundamental theories of emotion, i.e., dimensional emotion theory, discrete emotion theory, and cognitive appraisal theory (Watson & Spence, 2007). Dimensional emotion theory (Barrett & Russell, 1999) focuses on the effect of emotions from dimensions such as valence (positive and negative polarity) and arousal (the extent to which a person is energized by an experience) (Yin et al., 2017). Discrete emotion theory identifies specific basic emotions and delineates their effects (Plutchik, 1984; Tomkins, 1962). Discrete emotions are rooted in human evolution, with expression and recognition fundamentally the same across all individuals, regardless of ethnic or cultural differences (Plutchik, 2001). The theory originated from the theory of evolution, in the late 19th century, of Charles Darwin, who argued that certain basic emotions evolved from natural selection. Neuroimaging analyses have shown that these discrete emotions are linked to discrete neural signatures and certain structures of human brains (Saarimäki et al., 2016; Vytal & Hamann, 2010). Cognitive appraisal theory argues that emotional experiences result from cognitive appraisals of an event and its situational environment and highlights the effects of specific emotions, which supports discrete emotion theory (Desmet, 2010).

Discrete emotion theory and dimensional emotion theory are two competing theories in terms of measuring consumer emotions (Havlena & Holbrook, 1986; Lerner et al., 2004, 2015; Plutchik & Kellerman, 1982; Russell & Mehrabian, 1974). On the one hand, Havlena & Holbrook (1986) surveyed twenty participants about their emotions in consumption experiences and found that the emotional dimensions of Russell & Mehrabian (1974) yield good predictions for seven of the eight discrete emotions of Plutchik & Kellerman (1982). In this regard, emotional dimensions are more favorable than discrete emotions because, without losing much information, less data needs to be collected when using emotional dimensions. On the other hand, more recent lab experiments show that discrete emotions, such as anxiety and anger (Yin et al., 2014) and sadness and disgust (Lerner et al., 2004), are similar in emotional dimensions but have disparate effects on consumers' cognition and purchase intention, which suggests that the effects of discrete emotions cannot be fully predicted by emotional dimensions.

In view of dimensional emotion theory, existing management research on online reviews has focused mainly on "valence analysis" (e.g., Asur & Huberman, 2010; Hennig-Thurau et al., 2015; Rui et al., 2013; T. Song et al., 2019), which uses supervised machine learning models to classify a review as positive, negative, or neutral. More recent management research focuses on "discrete emotion analysis," i.e., detecting the intensities of discrete emotions from large-scale online content (Felbermayr & Nanopoulos, 2016; Malik & Hussain, 2017; Nguyen et al., 2020) by detecting emotion words in texts with existing emotion lexicons. An emotion lexicon contains emotion words and their associated emotional intensities of different emotion categories.

State-of-the-art practice, however, fails to consider domain differences and the evolutional nature of emotional expressions (Oliveira et al., 2016; Xue et al., 2014; Yin et al., 2014). Emotional expressions are domain-dependent and evolutional over time (Oliveira et al., 2016; Xue et al., 2014) and cannot be fully captured by a general and static emotion lexicon (e.g., RenCECps, LIWC). For example, Yin et al. (2014) used a domain-independent lexicon (LIWC) to capture anger and anxiety words in Yahoo! retailer reviews. Only 2.57% of reviews were detected to contain anxiety and anger words, which was "lower than expected." The authors noted that "these low values are not surprising given the use of a predefined dictionary that does not take context into consideration" (p. 550).

Notably, from an algorithmic and data-driven perspective, there is a third emotion analysis approach, referred to as "emotion topic analysis," i.e., use of a topic modeling approach to extract latent emotion topics (X. Yu et al., 2012).

2.1.2 *Economic Significance of Emotions: The Case of Online Reviews*

Emotions affect cognitive outcomes, including judgment and decision making (Lerner et al., 2015; Loewenstein, 2000). To be more specific and to motivate the context of the first essay, this section particularly focuses on the literature on emotions in online customer reviews. Emotions in reviews affect consumers' perceived review helpfulness and their attitudes toward products (Junyong Kim & Gupta, 2012; Yin et al., 2014). Perceived helpfulness is a cognitive result of reviews as well as a predictor of product sales (Mudambi & Schuff, 2010). Perceived helpfulness of reviews is defined as the extent to which the reviews are perceived by consumers to facilitate their purchase-decision process (Yin et al., 2014). When reviews are perceived as helpful, the recall of these reviews will affect consumers' attitudes and intentions (Filieri et al., 2018).

Table 2.1. Summary of Key FactorsTable 2.1 provides a summary of the key factors related to perceived review helpfulness, product evaluation, purchase probability, and purchase intention. Anger, anxiety, valence, and arousal (East et al., 2017; Junyong Kim & Gupta, 2012; Xiao et al., 2018; Yin et al., 2014, 2016, 2017) are of comparable or higher effect sizes in terms of other factors, e.g., review quality, credibility (Teng et al., 2017), and reviewer characteristics (A. H. Huang et al., 2015; Ngo-Ye & Sinha, 2014; Yin et al., 2016), showing the economic significance of these emotions in online reviews. Despite their potential significance, the effects of discrete emotions other than anger and anxiety are understudied, which is an important concern because one cannot expect the effects of anger and anxiety to be generalized to other discrete emotions.

Table 2.1. Summary of Key Factors

Dependent Variable	**Independent Variable**	**Normalized Effective Size**[a]	**Mechanisms Tested**	**Source**
Perceived Helpfulness	Valence (Rating)	Not significant	NA	Mudambi & Schuff (2010)
		0.07	NA	Schindler & Bickart (2012)
		0.07	NA	Huang et al. (2015)
		NA	Confirmation Bias	Yin et al. (2016)
	Anger	0.11	Perceived Cognitive Effort	Yin et al. (2014)
	Anxiety	0.36	Perceived Cognitive Effort	Yin et al. (2014)
	Reviewer Engagement	NA	NA	Ngo-Ye & Sinha (2014)

	Reviewer Impact	NA	NA	Huang et al. (2015)
	Review Ranking	0.12	NA	Yin et al. (2016)
	Review Length	0.15	NA	Yin et al. (2016)
	Arousal	0.18	Perceived Effort	Yin et al. (2017)
	Adjectives, State Verbs, Action Verbs, Readability, Subjectivity	NA	NA	Krishnamoorthy (2015)
Product Evaluation	Anger	-0.38	Review Information Value	Kim & Gupta (2012)
		-0.25	Perceived Problem Seriousness, Perceived Reviewer Rationality	Xiao et al. (2018)
	Valence	0.24	NA	Teng et al. (2017)
	Review Quality	0.17	Emotional Strength	Teng et al. (2017)
Purchase Probability	Positive/Negative Word of Mouth	NA	NA	East et al. (2017)
Purchase Intention	Review Credibility	0.20	NA	Teng et al. (2017)
	Two-sided Reviews	0.25	Information Helpfulness	Filieri et al. (2018)
	Food-related Disgust	-0.16	NA	Shimp & Stuart (2004)

Note. a. The effective sizes are normalized as the number of unit change in the dependent variable when there is a one-unit change in the independent variable. NA = not applicable.

According to the cognitive appraisal theory of emotion (Lerner et al., 2004, 2015) and neuroimaging research on emotion (Saarimäki et al., 2016; Vytal & Hamann, 2010), discrete emotions are independent of each other not only because they are evoked by different cognitive appraisals but also because they are linked to discrete neural signatures and certain structures of the human brain. Further, the existing mechanisms, i.e., perceived writer cognitive effort (Yin et al., 2014) and perceived writer rationality (Xiao et al., 2018), are built on anger and anxiety. The mechanism to explain the effects of other discrete emotions in online reviews, however, remains understudied.

2.1.3 *Emotion AI: Design, Application, and Regulation*

Researchers have identified emotion as an influential mediator in the interaction between organizations and individuals (T. Song et al., 2019; Y. Yu et al., 2023). Although emotion AI is still in its infancy, it attracts much attention from industry and academia (McStay, 2020). The functionality of emotion AI includes two parts, detecting emotions and responding to them. Current management research has mainly focused on the first part because it is arguably the foundation for the economic value of emotion AI. Researchers design and implement algorithms to detect emotional content in online reviews (Y. Yu et al., 2023), pictures (X. Wang et al., 2022), videos (Zhou et al., 2021), and live streaming (Lin et al., 2021).

Another emerging research stream focuses on identifying the impacts of using AI to display emotions to customers in service contexts. Chatbots' positive emotional expressions affect customers' service evaluation (Han et al., 2022). Relatedly, humanizing chatbots with emotional features, including humor and emojis, is beneficial for transaction outcomes (Schanke et al., 2021). Using empathetic technologies to display emotions to customers has almost zero marginal cost, and such a response can be considered unlimited resources that the firm can allocate to customers. This dissertation differs from this research by focusing on limited resource allocation. Although practitioners started using AI to allocate limited resources based on detected emotions, the design of such AI systems and associated policies is understudied. Notably, using AI to interpret and respond to human emotions raises concerns. First, inferring emotions using algorithms, such as algorithms that analyze facial expressions, always contains errors (Barrett et al., 2019). Therefore, one may be concerned that allocation based on detected emotions is impractical. Second, AI technologies sometimes collect private and sensitive data, including physiological and biometric data, to improve emotion detection accuracy. There is a call for regulating emotion AI's data collection (Crawford, 2021). Urgently needed yet currently scant is research on the roles of algorithmic errors in emotion AI applicability and the welfare impacts of regulating emotion AI.

A key question to customer service managers is whether to hire customer service employees or AI systems to detect negative emotions and allocate limited resources. Scalability is AI's obvious advantage over employees because the AI service is scalable at almost zero marginal cost. In many cases, AI has to be the first interface to customer emotional expressions due to scalability. For example, millions of online reviews may be generated for a newly-released movie in one day (Y. Yu et al., 2023), which human customer service systems do not have the capability to read and respond to. Therefore, comparing human and AI service systems is managerially meaningful only in contexts where scalability is not a bottleneck for human systems.

Human and AI service systems have distinct features regarding recognizing and responding to emotions, besides scalability. First, humans are the ground truth of recognizing others' emotional expressions, whereas algorithmic emotion detection is almost always imperfect (Y. Yu et al., 2023).

Second, humans have behavioral responses to others' emotional expressions, but algorithms essentially do not. An employee may be susceptible or feel empathetic about certain negative emotional expressions. In this case, they tend to allocate more resources than requested to agents who express a higher level of negative emotions. An employee can feel emotionally abused by agents' negative emotional expressions (Causon, 2021; Poddar & Madupalli, 2012). Besides, they may have negative emotional states induced by agents' negative emotional expressions because emotions are contagious (Campagna et al., 2016; Y. Yu et al., 2020). Therefore, sometimes employees may retaliate against the agent by allocating fewer resources than expected (Kumar Madupalli & Poddar, 2014). In addition, when employees feel abused, they may request their supervisor to allow them to end the service, resulting in lower-than-average allocation to the agent.

Third, employees who deal with negative emotions suffer from severe psychological stress (Causon, 2021). As a result, agents' negative emotions increase employees' turnover intentions (Poddar & Madupalli, 2012). Employee turnover incurs costs to the firm because the firm needs to recruit and train new staff. In addition, the firm must take legal and economic obligations even if an employee retains after suffering emotional abuse. Due to the potential heterogeneity in employees' sensitivity to negative emotional expressions, such spillover costs are heterogeneous across different agent-employee interactions.

2.1.4 *Strategic Interactions between AI and Users*

The third essay of this dissertation relates to the literature on signaling games (Spence, 1973) and particularly builds on the framework of "muddled information" (Frankel & Kartik, 2019, 2022). The traditional literature on signaling games studies the situation where agents strategically send signals to inform a principle about their private types. Because the benefit and cost of signaling can be heterogeneous among agents of different types, the principle may design an allocation mechanism to fully separate agents with different types based on their rational signaling behavior. The literature argues that if either dimension of an agent's type is homogeneous, a full separating equilibrium is possible (Kartik et al., 2007; Spence, 1973). With heterogeneity in two dimensions, the source of screening (heterogeneity in natural intensities) is muddled by the irrelevant information induced by heterogeneity in gaming ability.

This work differs from the current literature on muddled information (Frankel & Kartik, 2019, 2022) by considering the noise in signaling, the spillover effect of signals, and the variance in allocation. Signaling noise is intuitively considered to decrease signaling efficiency and welfare. However, in strategic interaction, the noise may reduce the gaming incentive and thus have intricate impacts on market equilibrium. The spillover effect significantly changes the equilibrium. Without spillover, the firm prefers a separating equilibrium (adopting AI) over a pooling equilibrium. However, the spillover is the key factor that may undermine the existence of a separating equilibrium. Finally, the allocation rules are considered fixed in the literature. However, this assumption is questionable when human employees have behavioral responses to negative emotions. Indeed, allocation variance caused by employees' behavioral responses can make AI outperform human systems.

The literature on strategic classification focuses on designing a machine-learning classifier interacting with strategic users who may manipulate the attributes used by the classifier to obtain more desirable outcomes (Hardt et al., 2016). This work differs from this literature by considering agents' types and allocation as continuous instead of discrete. Further, the classifier is also the allocation policy in this literature, whereas the algorithms used to detect emotion and allocation policies are separated in the current setting. Therefore, strategic classification focuses more on algorithmic transparency (Q. Wang et al., 2022), whereas this work focuses on the noise in detection and its impacts on allocation policy design. Notably, if the firm's allocation is binary and the agent has to invest a fixed cost before entering the game, the noise may benefit the firm by reducing gaming incentives (Braverman & Garg, 2020). When an allocation is allowed to be continuous and the agent does not have a fixed cost, this work shows that the variance of noise always makes the firm worse off. At the same time, it can unexpectedly increase social welfare.

A common assumption in the literature on muddled information and strategic classification is that signaling is costly. If agents do not have any costs in escalating their emotional expressions, rational agents will express emotions at infinite intensity. This is different from the observation in reality because agents have multiple costs of escalating emotional intensities. Although emotion escalation is not entirely equivalent to dishonesty, they may share similarities in occurring psychological costs and emotional exhaustion (Thielmann & Hilbig, 2019). For example, because emotional misrepresentation is commonly considered ethically inappropriate, it can undermine self-esteem and create negative feelings (Fulmer et al., 2009). Further, cognitive scientists argue that emotional expressions and subjective feelings cannot be fully separated, and even simulated negative expressions can partially activate corresponding negative emotional states, thereby carrying psychological costs to agents (Wood et al., 2016). In addition, more physical time and effort must be invested to express a higher level of emotion, e.g., writing more negative words to the system and speaking in a more intensive tone. For emotion AI embedded in smart wearable devices that detect physiological reactions (Hickey et al., 2021), agents need to manipulate data, including heart rates, sweating, and body temperature, which is even more costly.

2.2 CONSUMER BEHAVIOR TOWARD DIGITAL ADVERTISEMENTS

2.2.1 *Instream and Outstream Video Advertisements*

Marketing researchers have long known that online consumers tend to be “banner-blind” - consciously or unconsciously, they ignore banner-like ads. To enhance digital advertisement effectiveness, marketers are turning to video ads, which are more interactive, informative, and entertaining to consumers than traditional banner ads. Over the past couple of years, instream video advertising, delivered pre-roll, mid-roll, or post-roll within video content on video platforms (e.g., YouTube), has become an accepted norm.

More recently, a newer format of video ads, outstream, has become increasingly popular. Outstream video advertising turns up in non-video environments (e.g., online shopping websites including Amazon, news websites such as CNN.com), where a video ad begins playing as a user hovers over it. According to eMarketer, 77% of ad agencies and 70% of advertisers worldwide believe that outstream video ads are becoming the most important type of video ads; nearly 60% of advertisers think that outstream video ads provide better user experiences and higher return-on-investment than other formats of ads.

Several unique features distinguish outstream video ads in online shopping websites from instream video ads on video platforms. First, outstream video ads are specific to customers' purchase intention, as they directly target customers' search queries, whereas in the case of instream video ads, viewers' interests are ambiguous and are at most inferred based on the video content consumed. This feature can lead to different design principles for outstream and instream video ads. Second, outstream video ads exist in a competitive environment-many similar and competitive products are shown to customers together with the advertised product. Thus, the effectiveness of the outstream video ads may be shaped by the competitive dynamics. Third, outstream video ads cannot be skipped. Instream ads, in contrast, are frequently skipped and may be completely avoided by users with premium accounts, which reduces the impressions of the video ads and creates biased targeting due to video viewers' self-selection (Joa et al., 2018).

Despite the uniqueness of outstream video ads, the existing literature of marketing research on video ads has focused mainly on instream video ad viewership (Campbell et al., 2017; Goodrich et al., 2015; Joa et al., 2018) or their diffusion among consumers (Akpinar & Berger, 2017; Tellis et al., 2019). Marketing researchers have found principles of designing effective instream video ads. To begin, attention-getting tactics for instream video ads are redundant and may backfire (Campbell et al., 2017). Further, effective instream video ads are long (Goodrich et al., 2015), not necessarily relevant to the advertised products, and focus on entertainment values (Joa et al., 2018). Instream and outstream videos differ likely because online shopping consumers are goal-driven, aiming to efficiently learn the pros and cons of a product, whereas instream video ads viewers are likely to be "prosumers," who produce and consume video content, aiming to learn novel video-making ideas or merely consume the video ads for entertainment purposes (Joa et al., 2018). Comprehensive empirical investigation on the effectiveness of outstream video ads regarding driving click-through rates (CTRs) is lacking, possibly because it is challenging to collect large-scale data both on outstream videos ads and customers' click-streams.

2.2.2 *Two-stage Model of Advertising*

Attention and clicking are the two important stages of consumer behavior toward digital advertisements. In cognitive psychology, attention is the process of selectively allocating limited cognitive resources to specific information and ignoring a large amount of competing information and stimuli in the same cognitive environment (James, 2004; Posner & Snyder, 1975). In the online shopping context, although the product information and ads are presented on the screen, consumers do not allocate cognitive capacity to all content equally. Clicking can happen only when a product can attract a consumer's attention. Attention and clicking are also linked to the literature of business-to-consumer marketing funnels (Li & Xie, 2020; Rawal, 2013). A marketing funnel describes a consumer's journey to purchase, which includes awareness, consideration, and purchase. The first functionality of ads is to attract consumers' attention, which corresponds to the "awareness" step in marketing funnels (Rawal, 2013). Only contingent on consumer attention can ad content affect the consumer's probability of clicking, which corresponds to the "consideration" step in marketing funnels (Li & Xie, 2020). Therefore, to understand advertising effectiveness regarding CTR, it is important to focus on both stages.

Outstream video ads may affect both stages. First, one of the best approaches to attracting consumer attention is to break existing patterns of web content through a highly creative message, known as “creative disruption” (Colon, 2016). Relatedly, marketers often use surprise or unconventional content to attract consumer attention, termed as “guerrilla marketing” (Levinson, 2011). On shopping websites, the conventional information flow is of product textual descriptions and images. Outstream video ad impression, acting as a guerrilla marketing tactic, induces creative disruption to consumers by breaking the flow unexpectedly, and may effectively attract consumer attention. Second, if consumers pay attention to video ads, video content, as a type of highly interactive and user-friendly marketer-generated content, can persuade consumers to click the sponsored product (T. Song et al., 2019).

Few studies investigate the effectiveness of outstream video ads on the consumer two-stage decision, despite the important managerial implications. Further, although clicking behavior can be easily measured by keeping track of click-stream data, consumer attention is essentially unobservable in observational data. It may only be measured by using eye-tracking devices in lab experiments (Vehlen et al., 2021). Therefore, unobservable consumer attention provides critical challenges for marketers, platforms, and researchers to understand consumer behavior and advertising effectiveness in real-world business environments.

2.2.3 *Advertisement, Competition, and Limited Attention*

A controversial question in the marketing literature is whether the effectiveness of advertisement is affected by market competition levels, and if so, to what extent (Comanor & Wilson, 1979). On the one hand, Telser (1964) argued that advertisement is less effective in markets in which products are less differentiated (i.e., more competitive markets) because little empirical evidence suggests that advertising can effectively differentiate products in highly competitive markets and increase sales. On the other hand, other scholars argue that ads can be more effective in competitive markets (Comanor & Wilson, 1979) because ads can be successful in differentiating similar products (Tirole, 1988). One possible mechanism is that ads may signal unobservable product quality information (Nelson, 1970).

The controversial question contextualizes in the digital era with a new issue, that is, limited attention (Cetin & Bingol, 2014; Kong et al., 2019). An individual consumer's cognitive resources are too limited to process all the product information listed on shopping websites. Therefore, consumers pay attention only to attractive shopping content, and attention scarcity becomes a dominant factor affecting advertising effectiveness on online shopping websites (O'Donnell & Cramer, 2015). Ads are successful only when they can cut through the clutter of product information on e-commerce platforms. An analytical study suggests that the advertised product is predicted to obtain a higher market share in a more competitive market, because in such a market it is even harder for sellers to attract and hold consumers' attention (Cetin & Bingol, 2014). According to this proposition, outstream video ads should be more effective when products in a market are less differentiated.

However, little empirical evidence supports such a theory. The empirical challenges happen in several aspects. First, as noted, consumer attention is unobservable to advertisers and researchers. Second, it is difficult to obtain large-scale, cross-market data in which variation of competitive levels is meaningful for statistical identification. Third, when the number of markets is large, it is challenging to quantitatively measure product differentiation in each of the markets.

2.2.4 *Elaboration Likelihood Model and Video Content Analytics*

After a consumer pays attention to the content of the video ad, the Elaboration Likelihood Model (ELM) framework can be used to explain how various features of video content could further shape video ad effectiveness. The ELM proposed by Petty and Cacioppo has been an influential and widely-accepted theory in psychology and marketing (Petty & Cacioppo, 1986). It has been leveraged recently to explain the effect of visual elements in images on consumer engagement toward social media posts (Shin et al., 2020).

Based on the ELM framework, video features are sorted into two categories: features in the central route and those in the peripheral route. According to the ELM framework (Petty & Cacioppo, 1986), some consumers decide to buy a product by evaluating its pros and cons, that is, through the central route. Other consumers make decisions based on the superficial aspects of the product or cues that do not relate to the product itself, that is, by the peripheral route. If a video ad feature can facilitate the process of evaluating the pros and cons of the advertised products and affect clicking behavior, it is referred to as a feature in the central route. If a video ad feature affects clicking behavior through aspects that do not relate to the pros and cons of a product, it is referred to as a feature in the peripheral route, for example, visual aesthetics (Zhang et al., 2022) and celebrity endorsement (Shin et al., 2020).

Prior to this work, few studies have empirically tested whether video features in either route would influence outstream video ad effectiveness. It is an interesting question because it is difficult to know ex-ante how consumers make clicking decisions when they are shopping online. It could also be possible that video features in neither route would have a significant effect on CTR because, as some researchers argue, the effect of ads is more about attracting attention rather than persuading purchase decisions (Cetin & Bingol, 2014). To investigate the question, one needs to detect video features in both routes from video content.

Despite the prevalence of online videos, management researchers have not explored the effects of video content features until recently, possibly due to the formidable methodological challenges regarding video content analytics. Most papers have been relying on manual coding to analyze videos (Goodrich et al., 2015; A. N. Smith et al., 2012; Teixeira et al., 2014), which is incompatible with today's large scale online video ad data that needs to be analyzed on e-commerce platforms. Further, although in the literature a few exceptions use automatic approaches to video analytics, they have been focused mainly on customer emotional reaction detection (Teixeira et al., 2012), customer facial expression analysis (Lu et al., 2016), instructor emotion analysis, and visual aesthetic analysis for online classroom videos (Zhou et al., 2021). In current literature, few studies investigate the content of video ads. The challenge of understanding video ad effectiveness does not rise only from the methodological side. Although computer vision techniques and deep learning models emerged as a must in industrial practice for video content understanding, their theoretical foundations in the management literature are not yet clear.

Chapter 3. PREDICTING SALES USING EMOTION-AWARE ALGORITHMS

With the prevalence of social media and online review platforms that make reviews ubiquitous, reading reviews has become an integral part of consumers' purchase decisions. The survey of the movie industry (detailed later in this chapter) shows that 75.2% of consumers will (and another 21.3% may) refer to online reviews before they decide to buy a movie ticket. The economic value of emotions in online reviews has attracted growing attention from the industry. A management perspective emphasizes the managerial salience of understanding and reacting to emotions (e.g., anger, sadness, surprise) in large-scale online reviews with the advanced technology of artificial intelligence (AI) and big-data analytics, known as "emotion AI" or "affective computing."[1] Indeed, 25% of Fortune Global 500 companies and over 1,400 brands have adopted emotion AI in their marketing research.[2]

Measuring emotions in online reviews and understanding their effects on consumer behavior and business performance, however, poses a theoretical and algorithmic challenge. I have reviewed the literature in the previous chapter and shown that there are three different algorithmic representations of customer emotions, i.e., valence, discrete emotions, and latent emotion topics. The motivation of this work originates from the algorithmic side. First, among the three emotion analysis approaches, which is the best to represent emotional information in reviews and of the highest predictive power in regard to sales? I expect valence analysis, the common practice of understanding emotion in online reviews, to be less effective than is discrete emotions analysis because, from a theoretical perspective, valence is merely one of the important dimensions of human emotions (Lerner et al., 2015), and lab experiments show that valence-based predictions of consumer cognition and behavioral intention yield contradictory results (Lerner et al., 2004; Yin et al., 2014). It is unclear, however, whether the data-driven latent emotion topics would outperform the theory-driven discrete emotions. Second, I consider whether combining different representations of emotions produces better predictive power. Investigating these two questions would improve the state-of-the-art algorithmic practice of representing emotions in online content and provide empirical evidence for relevant emotion

[1] https://mitsloan.mit.edu/ideas-made-to-matter/emotion-ai-explained
[2] https://www.affectiva.com/who/about-us/

theories. Third, I consider how to incorporate domain differences and the evolutional nature of emotional expressions and advance the state-of-the-art discrete emotion analysis approach.

To this end, I focus on the movie industry for the following reasons. First, the movie market is of high economic importance, with global box office revenue that reached a record $42.5 billion in 2019.[3] Second, social media and online review platform users have considerable interest in discussing movies, with substantial variance in their emotions. Third, the real-world business outcome (i.e., box office revenue) is publicly observable. More importantly, the movie industry was among the most disrupted industries during the COVID-19 pandemic, with billions of dollars lost in the global box office in 2020.[4] As more movie marketing and sales activities go digital, it is timely to leverage the economic value embedded in online user-generated content, which could potentially contribute to the digital resilience of the industry.

In Study 1, from China's movie market, the world's largest, I collect a large-scale dataset that contains 499 movies released between 2012 and 2018 and 3,257,871 microblogging messages related to these movies. I identify 1,214,310 movie reviews from the microblogging messages by utilizing a support vector machine (SVM) model and then conducting three emotion analyses on the reviews. First, I conduct valence analysis by following the state-of-the-art implementation in the literature (Hennig-Thurau et al., 2015; T. Song et al., 2019) and classify reviews as positive or negative. Second, to enable domain-adaptive discrete emotion analysis, I construct a domain-specific and up-to-date emotion lexicon by extending an existing general emotion lexicon built in 2008, Ren-CECps (Quan & Ren, 2010), and combining it with a popular neural network language model, the Word2Vec model.[5] After validating the newly generated lexicon with out-of-sample testing and human annotation, I utilize it to detect the intensities of eight discrete emotions in reviews, i.e., surprise, joy, anticipation, love, anxiety, sadness, anger, and disgust. I focus on these emotions not only because they are considered the most basic emotions that constitute other emotions, according to discrete emotion theory (Lerner et al., 2004, 2015; Plutchik & Kellerman, 1982; Tomkins, 1962; Yin et al., 2014), but also because

[3] https://www.boxofficepro.com/global-box-office-2019-record-42-5-billion/

[4] https://en.wikipedia.org/wiki/Impact_of_the_COVID-19_pandemic_on_cinema

[5] The Word2Vec model is a commonly used statistical language model proposed by Mikolov et al. (2013). It maps a word to a high-dimensional vector that indicates the meaning of the word in a context, using deep-learning techniques. It can exploit the semantic information of words and judge the semantic similarity between words by calculating the cosine distance between corresponding word vectors. Compared with other statistical language models, such as the neural network language model and recurrent neural network language model, Word2Vec provides better performance on measuring semantic word similarities, with a much lower computational cost (Mikolov et al., 2013).

they are the most commonly seen in online content (Quan & Ren, 2010). Third, I conduct emotion topic analysis by advancing the approach proposed by Yu et al. (2012). I interpret the latent emotion topics by leveraging the estimated topic-word distribution and the emotional intensities associated with the emotion words in the lexicon. Based on the results, I infer that there are two prevalent emotions in movie reviews. The first one is positive and activated, resembling "excited" or "delighted"; the second one is less positive and of low arousal, resembling "bored" or "disappointed."

After obtaining three representations for emotions, and combining them with other control predictors, I use an expanding window strategy (Geva et al., 2017; T. Song et al., 2019) to predict box office sales with four machine learning models, i.e., linear regression (LR), random forest (RF), support vector regression (SVR), and XGBoost (XGB), which are either popular or state-of-the-art for numeric prediction tasks.

I find that, first, compared to a basic lexicon, the newly constructed domain-specific and up-to-date lexicon for discrete emotion analysis achieves significantly higher prediction accuracy for box office sales across all prediction models, which demonstrates the practical value of addressing the domain difference and evolutional nature of emotional expressions. Second, discrete emotions achieve significantly higher predictive power than does valence. Under the SVR model, all discrete emotions can *individually* achieve a higher prediction accuracy than can valence. In addition, discrete emotions are of significantly higher predictive power than are latent emotion topics. Third, combining discrete emotions with valence or latent emotion topics does not improve prediction accuracy, which implies that the emotional information carried by valence and latent emotion topics has been absorbed in discrete emotions. These findings demonstrate the economic value of discrete emotions in online reviews and provide empirical evidence for discrete emotion theory.

The algorithmic study further motivates this work to investigate the source of the predictive power of discrete emotions in online reviews from a theoretical perspective. If the algorithmic predictive relationship is causal with an interpretable mechanism, the robustness of the predictive study can be further justified, and one can expect the predictive results to be generalizable to other contexts in which a similar mechanism exists. Moreover, although the existing work on emotion in online reviews has examined the effects of anxiety and anger on perceived helpfulness (Yin et al., 2014) and happiness and anger on product evaluation (Junyong Kim &

Gupta, 2012; Xiao et al., 2018), the effect sizes of other discrete emotions remain understudied. As emotional experiences consist of eight discrete emotions (Plutchik & Kellerman, 1982), the effects of emotions in reviews cannot be fully understood until we have knowledge of a comprehensive set of discrete emotions. More importantly, although the existing mechanisms, i.e., perceived writer cognitive effort (Yin et al., 2014) and perceived writer rationality (Xiao et al., 2018), help us to understand the effects of anxiety and anger, we lack a mechanism to understand the effects of other discrete emotions. Finally, this work is motivated by the potentially cross-cultural differences in the effects of emotions (Eid & Diener, 2001), which make one wonder whether the effects of emotions in online reviews remain significant for English users.

I thus conduct a randomized experiment in Study 2 to investigate the causal impacts of emotions in online reviews on consumer purchase intention and their underlying mechanism in the U.S. movie market. I find that all positive (negative) discrete emotions have a significantly positive (negative) effect on purchase intention. Discrete emotions have higher effect sizes than do the important factors investigated by the previous studies, which highlights their economic significance. Second, mediated by *perceived processing fluency* (the extent of feeling the ease of cognitive processing) (Reber et al., 2002), discrete emotions significantly affect the perceived review helpfulness. Such a mediation effect is significant when subjects are exposed to all discrete emotions, except anger. Further, the effect of anger is completely mediated by perceived writer rationality. Finally, I confirm that perceived helpfulness significantly influences purchase intention. The remainder of this chapter is organized as follows. First, I elaborate on the analyses and results of Studies 1 and 2. Second, I discuss the contribution and managerial implications of this work.

3.1 STUDY 1: A FIELD STUDY ON PREDICTION

Predicting future sales is a critical issue in various business domains. I focus on the Chinese movie market, the world's largest. In this industry, however, the movie theater's operation efficiency is surprisingly low. Among 9,504 theaters, only 15 achieved over 50 million RMB of revenue per year, and the average occupancy rate is only around 15%.[6] In what follows, I explain how weekly and daily box office predictions can enhance theater operation efficiency.

First, weekly box office predictions contribute to movie selection decisions. In this market, most theaters are small, with an average of only 5.3 screens.[7] Nevertheless, the total number of new movies in one month could be 30 to 40, which means that theater managers should choose one from at least three new movies for a single screen. These decisions are usually made on a weekly basis. An improper movie screening decision could lead to a 0.21 million RMB loss in revenue per week.[8] Thus, optimizing movie selection decisions according to box office sales predictions is of great importance for a majority of the theaters. Second, daily box office predictions help daily screening room arrangements. In a theater, the screening rooms are usually of different sizes (ranging from tens to hundreds of seats). If one movie is predicted to be popular the next day, the cinema managers can move it to a larger room or temporarily add more screening schedules without running into a ticket stockout problem.

Third, daily demand predictions contribute to cinemas' inventory management. Cinemas offer low-profit-margin tickets to attract an audience and sell high-profit-margin food and drinks to make a profit. The inventory management decisions regarding perishable food and drinks need to be made on a daily basis. Thus, each day, cinema managers face a newsvendor problem, whereby they need to order a certain amount of perishable inventory in advance to meet the next day's uncertain demand. Accurate predictions of daily demand are critical to reducing the cost of inventory. To address these managerial challenges, this work first focuses on daily box office predictions and includes additional analysis on weekly predictions.

3.1.1 *Data Collection*

Fourteen research assistants were invited to collect all publicly available movie samples from 2012 to 2018 from maoyan.com, a public movie database in China. For each movie during its screening period, I collected its daily box office revenue from maoyan.com and all of the microblogging reviews related to the movie from weibo.com, the largest microblogging platform in China. I deleted movies without any reviews and obtained a sample of 589 movies. In addition, I deleted the movie samples that contained missing data in their box office revenue and

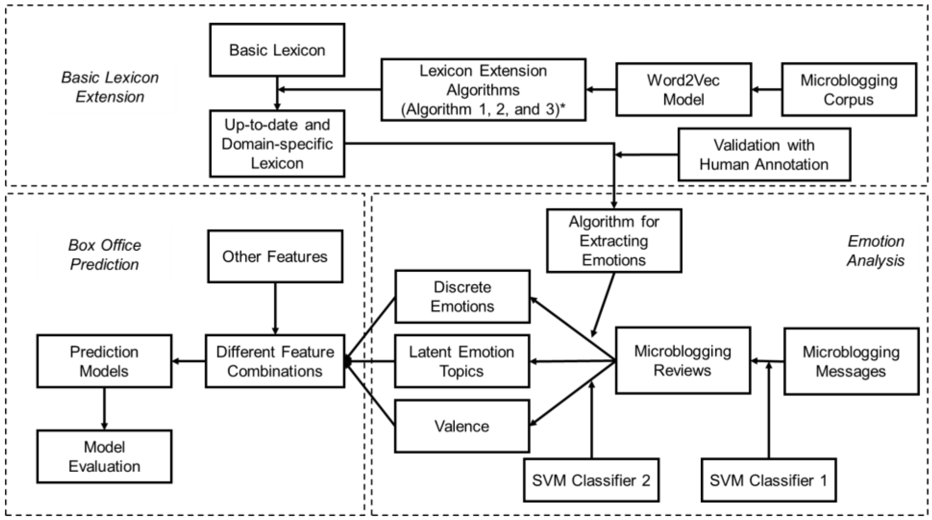
[6] http://tech.sina.com.cn/i/2018-06-24/doc-iheirxye9028421.shtml
[7] http://finance.eastmoney.com/news/1355,20180110820566856.html
[8] https://www.douban.com/group/topic/50452206/

obtained 499 movies. The number of movie and date combinations is 12,993, and the number of microblogging messages is 3,257,871. The average total box office revenue for a movie is 286.5 million RMB. The average box office revenue for a movie per day is 11.2 million RMB, and the average number of a movie's screening days is 16.9.

The scale of this empirical test is the largest among the extant studies on microblogging messages and movie box office sales, which usually involve fewer than 100 movies or have less than a one-year observation period (Hennig-Thurau et al., 2015; Liu, 2006; Rui et al., 2013; T. Song et al., 2019; X. Yu et al., 2012). Such a sample mimics a practitioner's situation, whereby they would like to use all of the available data to achieve the best predictive accuracy. In addition, I downloaded the Natural Language Processing and Information Retrieval (NLPIR) microblog corpus (over 5 million microblogging messages). I combined the NLPIR corpus and 3.26 million collected movie-related microblogging messages as training data for the Word2Vec model.

3.1.2 *Algorithms and Empirical Analyses*

I present the empirical framework in **Error! Reference source not found.**. This method i nvolves three parts: basic lexicon extension, emotion analysis, and box office sales predictions. First, I train the Word2Vec model, with which I extend the basic lexicon into an up-to-date and domain-specific emotion lexicon. Second, I conduct emotion analysis on microblogging reviews. To obtain microblogging reviews, I drop non-review messages with a pre-trained SVM classifier. Then, I conduct valence analysis, discrete emotion analysis, and latent emotion topic analysis on the reviews. Third, I combine emotion features along with other features, including historical box office sales, microblogging review volume, and screening days and weekends, to generate box office sales predictions and investigate the predictive power of different emotion features.



Note: *Algorithms 1, 2, and 3 are detailed in the following section.

Figure 3.1. Empirical Framework

This approach requires an existing emotion lexicon as a basic lexicon. Ren-CECps is an emotion lexicon based on 1,487 Chinese blog texts and contains 22,406 unique emotion words (Quan & Ren, 2010). I filter out 7,179 words that do not appear in any of the collected microblogging messages and obtain a lexicon that contains 15,227 words. Ren-CECps focuses on eight discrete emotions, i.e., surprise, joy, anticipation, love, anxiety, sadness, anger, and disgust. Each word w_i in Ren-CECps is associated with an eight-dimension emotion-intensity vector $v_i = \{e_j^i\}_{j=1}^{N=8}$, where $e_j^i \in [0,1]$ is the manually annotated intensity of the jth discrete emotion.

Ren-CECps is selected for the following three reasons. First, these eight discrete emotions are most commonly expressed in Chinese blog texts; using these emotions decreases confusion in emotion-category selection (Quan & Ren, 2010). Second, the texts are annotated based on Chinese blogs. Blogs and microblogging messages are online user-generated content. Hence, Ren-CECps is more suitable for this context than are lexicons based on general Chinese texts (e.g., NTUSD, HowNet, DUT). Finally, Ren-CECps is manually annotated and statistically validated (Quan & Ren, 2010).[9] Compared to the algorithm-constructed emotion lexicons (e.g., Yang et al., 2016), the results of Ren-CECps are more precise and reliable.

[9] Eleven annotators participated in the annotation work. According to Quan & Ren (2010), the authors spent two months on the joint training of annotators and developed annotation instructions. They also used a Kappa statistic to measure the pairwise agreement among the 11 annotators. The Kappa coefficient of the agreement is a statistic adopted by the computational linguistics community as a standard measure for such a purpose. The agreement for emotional words is 0.785. Given the complexity of this annotation task, I believe that the annotations are reliable and valid.

Next, I trained a Word2Vec model by using 3.26 million Chinese movie-related microblogging messages from the microblogging platform and 5 million microblogging messages from the NLPIR microblog corpus (Appendix A.1). The Word2Vec model maps each word in the training text to a high-dimensional vector, termed a word vector, which contains the semantic information of the word (Mikolov et al., 2013). The Word2Vec model defines the similarity between two words as the cosine similarity between their associated vectors (Mikolov et al., 2013; Y. Song et al., 2018).[10]

Algorithm 1: Lexicon Extension	**Remarks**
Input a basic lexicon $L_0 = \{(w_i, v_i)\}_{i=1}^{N}$;	w_i is an emotion word and v_i is an eight-dimension vector that represents emotional intensities of w_i.
Randomly split L_0 into training (80%), validation (10%), and test (10%) sets: L_0^{tr}, L_0^{va}, and L_0^{te};	
Get all words from corpus W;	W is the set of all unique words, except stopping words, in a target domain.
Input a pre-trained Word2Vec model;	
FOR each word $w_i \in (W - L_0^{tr})$:	
Get M words that are most similar to w_i: $\{w_{i,j}\}_{j=1}^{M}$;	M is selected as 100 (Xue et al., 2014).
IF $\exists j\ s.t.\ w_{i,j} \in L_0^{tr}$:	
Add $w_{i,j}$ into the Potential Emotion Word Set P;	
END IF	
END FOR	
Initiate lexicon $L_{old} = L_0^{tr}$, $L_{new} = L_0^{tr}$;	L_{new} will be the output as an extended lexicon.
Input an algorithm $f(w; L, \theta)$, a mapping from a word w to emotion intensities $(e_1, e_2, \ldots, e_8)$, where $e_i \in [0,1]$;	f is defined conditional on a parameter θ and a lexicon L. The details of function f are provided in Algorithm 2.
WHILE TRUE:	An iterative process.
FOR each word $w_i \in P$:	
$v_i \leftarrow f(w_i; L_{old}, \theta) = (e_1, e_2, \ldots, e_8)$;	
IF $\sum_i e_i > 0$:	If a word has at least one positive intensity value, add it to the extended lexicon.
$L_{new} \leftarrow L_{new} \cup \{(w_i, v_i)\}$;	
END IF	
END FOR	
IF $\lvert L_{new} \rvert = \lvert L_{old} \rvert$:	Check convergence.
BREAK	
ELSE:	
$L_{old} \leftarrow L_{new}$;	Update the basic lexicon and further extend the lexicon.
RETURN L_{new};	

[10] The natural language processing module *Gensim* in Python is used to construct the model. The dimension of a word vector is set to 200. This parameter is a default setting suggested by Gensim and used in the state-of-the-art implementation of Chinese word embeddings, which are validated across different Chinese natural language processing tasks (Y. Song et al., 2018).

Algorithm 1 presents the lexicon extension algorithm. First, I randomly split the basic lexicon $L_0 = \{(w_i, v_i)\}_{i=1}^{N}$ into training (80%), validation (10%) and test (10%) sets, i.e., L_0^{tr}, L_0^{va}, and L_0^{te}. Second, I construct W as a set of all unique words in the target domain, except stopping words. From W, I construct a subset P that contains potential emotion words. Specifically, for each word $w_i \in W$, if $w_i \notin L_0^{tr}$, I get M words that are most similar to w_i: $\{w_{i,j}\}_{j=1}^{M}$ by using the pre-trained Word2Vec model. If $\exists j\ s.t.\ w_{i,j} \in L_0^{tr}$, $w_{i,j}$ is considered a potential emotion word (Xue et al., 2014), and I add $w_{i,j}$ to P. Third, I iteratively extend the basic lexicon L_0^{tr}. I define an algorithm $f(w; L, \theta)$ that maps a word w to an emotion-intensity vector $v_i = (e_1, e_2, \ldots, e_8)$, where $e_i \in [0,1]$. As I detail in Algorithm 2, f is conditional on parameter θ and lexicon L. For each word $w_i \in P$, I use f to map w_i to an emotion-intensity vector v_i. If v_i has at least one positive intensity value, I add a word-intensity pair (w_i, v_i) to the basic lexicon. After one iteration, I check whether the number of words in the basic lexicon increases. If so, I repeat the iteration until the number of words converges.

Algorithm 2: Mapping Word to Emotional Intensities (Implementing f)	**Remarks**
Input a lexicon L;	L can be, for example, L_0^{tr} in Algorithm 1.
Input parameter $\theta = (K, \alpha)$; $K \in Z^+, \alpha \in [0,1]$;	Z^+represents the set of all positive integers.
Initiate outcome $v = (e_1, e_2, \ldots, e_8)$; $\forall i, e_i \leftarrow 0$;	
Input a pre-trained Word2Vec model;	
Input word w;	
Get a set of K most similar words to w in the Word2Vec model: $S_w \leftarrow \{w_j\}_{j=1}^{K}$;	
$S_w^L = S_w \cap L$;	Find all of the most-similar words that are in the lexicon.
IF $\lvert S_w^L \rvert = 0$: RETURN $v = (e_1, e_2, \ldots, e_8)$; END IF	If the interaction is an empty set, return an all-zero vector and end the algorithm.
FOR each word $w_j \in S_w^L$:	
$d_j \leftarrow$ semantic similarity between w_j and w;	The similarity is measured by cosine distance between the word vectors of w_j and w (Mikolov et al., 2013).
$v_j = \left(e_1^j, e_2^j, \ldots, e_8^j\right) \leftarrow$ retrieve emotional intensities of w_j from L;	
END FOR	
FOR $i \in \{1,2, \ldots, 8\}$:	
$e_i \leftarrow \frac{\sum_{j=1} d_j e_i^j}{\sum_j d_j}$;	Estimate the emotional intensities of w by the weighted average of its most similar words.
IF $e_i < \alpha$: $e_i \leftarrow 0$;	If the magnitude is less than a threshold, the value will be compressed to zero to reduce noise.
END IF	

END FOR RETURN $v = (e_1, e_2, \ldots, e_8)$;

Next, in Algorithm 2, I illustrate how I map a word to emotional intensities (f). The rationale of f originates from the K-nearest neighbor algorithm, but I customize it in the task and introduce hyperparameter $\alpha \in [0,1]$ to control for noise. In particular, for word w, I construct set S_w that contains K-most similar words by using the Word2Vec model. I then check the intersection of S_w and emotion lexicon L, denoted as S_w^L. If S_w^L is non-empty, I average the emotion intensities of emotion words in S_w^L to determine the emotion intensities of w. I compress emotion intensities that are lower than α to zero to reduce noise.

Algorithm 3: Validate and Test *f*	**Remarks**
Input a basic lexicon $L_0 = \{(w_i, v_i)\}_i^N$; Randomly split L_0 into training (80%), validation (10%), and test (10%) sets: L_0^{tr}, L_0^{va}, and L_0^{te};	L_0 is the same as that in Algorithm 1.
Draw a set Θ of $\theta = (K, \alpha)$, where K is randomly selected from $\{1,2, \ldots, 10\}$ and $\alpha \sim uniform(0, 0.2)$;	Random parameter search.
FOR each $\theta \in \Theta$: FOR each $w_i \in L_0^{va}$: $v_i^0 = (e_1^0, e_2^0, \ldots, e_8^0)$ ←retrieve true values of emotional intensities of w_i from the validation set;	Find the best hyperparameters with validation set.
$v_i \leftarrow f(w_i; L_0^{tr}, \theta) = (e_1, e_2, \ldots, e_8)$; Get MAE $\epsilon_i \leftarrow \sum_{i=1}^{8} \frac{\lvert e_i - e_i^0 \rvert}{8}$; END FOR Calculate average validation MAE $\epsilon_\theta \leftarrow \sum_i \frac{\epsilon_i}{\lvert L_0^{va} \rvert}$; END FOR Find $\theta^* \leftarrow \underset{\theta \in \Theta}{\operatorname{argmin}} \, \epsilon_\theta$;	Get predicted values with the training set.
FOR each $w_i \in L_0^{te}$: $v_i^0 = (e_1^0, e_2^0, \ldots, e_8^0)$ ←retrieve true values of emotional intensities of w_i from the test set; $v_i \leftarrow f(w_i; L_0^{tr}, \theta^*) = (e_1, e_2, \ldots, e_8)$; Get MAE $\epsilon_i \leftarrow \sum_{i=1}^{8} \frac{\lvert e_i - e_i^0 \rvert}{8}$; END FOR Calculate average test MAE $\epsilon_{\theta^*} \leftarrow \sum_i \frac{\epsilon_i}{\lvert L_0^{va} \rvert}$; RETURN ϵ_θ and ϵ_{θ^*};	Calculate test errors with the test set.

I use validation set L_0^{va} to optimally choose parameters K and α in f by using a random parameter search (Bergstra & Bengio, 2012) and then use test set L_0^{te} to evaluate the out-of-sample performance of f, as shown in Algorithm 3. I first draw a set Θ of $\theta = (K, \alpha)$, where K is randomly selected from $\{1,2, \ldots, 10\}$ and $\alpha \sim uniform(0, 0.2)$. For each θ, I use $f(w; L_0^{tr}, \theta)$ to generate a predicted emotion-intensity vector v for w. Then, I use mean absolute error (MAE) to evaluate the error between v and the human-annotated value v^0. After I obtain the optimal parameter θ^* that minimizes the validation MAE, I evaluate the out-of-sample MAE of $f(w; L_0^{tr}, \theta^*)$ based on test set L_0^{te}. The optimal parameter $\theta^*$is (K^*, α^*) = (5, 0.15), and the corresponding validation MAE is 0.072 (for full parameter-search results, see Appendix A.2). The MAE for the test set is 0.073. Given that the MAE can vary from 0.0 to 1.0, this result demonstrates both the in-sample and out-of-sample validity of the algorithm. Alternatively, Algorithm f can be constructed by using the mean model [11] or directly training more sophisticated state-of-the-art models, i.e., RF and XGB, on the 200-dimensional word vectors. As I detail in Appendix A.3, however, additional analysis shows that, although RF and XGB can outperform the mean model, the proposed algorithm achieves the best performance.

Next, I use Algorithm 1 and f obtained in Algorithms 2 and 3 to extend the basic lexicon. To maximize the model accuracy, I combine the training and validation sets as a basic lexicon. I use the test set to evaluate the out-of-sample MAE after each iteration in Algorithm 1. As shown in Figure 3.2, Algorithm 1 stops after 11 iterations, and a total of 6,710 new emotion words are obtained. This shows the value of lexicon extension, without which at least 30.6% (6,710 out of 21,937) of the emotional expressions would not be captured. During the iterations, the test MAEs are consistently between 0.067 and 0.068, demonstrating the validity of the iteration process.

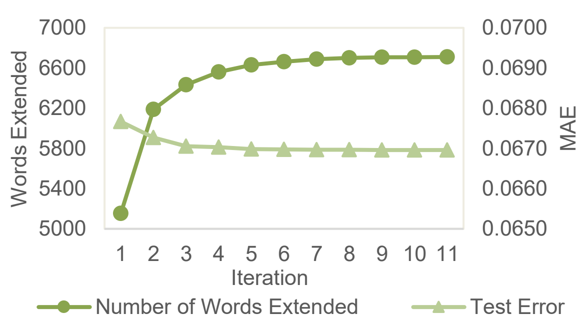


Figure 3.2. Number of Words Extended and Test Errors

To further ensure the validity of the newly mined emotion words, I randomly select 1,000 newly mined words and follow the manual annotation procedure by Quan & Ren (2010) to evaluate the accuracy of the estimated emotion intensities of these words. Because the emotional expression of a word depends on its context, for each word, I randomly retrieve 10 different

[11] As a benchmark model, the prediction of a new word in the mean model is produced by using the average emotion intensities of all words in the training set, i.e., $f_{mean}(w; L_0^{tr}) = \sum_i v_i / |L_0^{tr}|$, where $v_i \in L_0^{tr}$.

movie reviews that contain the word. Then, I recruit and train ten research assistants to annotate the emotional intensities of these words based on their contexts. For each word w_i ($i \in \{1,2,\ldots,1000\}$) in each review r_{ij} ($j \in \{1,2,\ldots,10\}$), two research assistants independently annotate intensities of the eight discrete emotions. I average the two assistants' annotation results to ensure that the results are not biased toward either research assistant's subjectivity. I denote the averaged annotation as $v_{ij} = \left(e_{ij}^1, e_{ij}^2, \ldots, e_{ij}^8\right)$, where $e_{ij}^k \in [0,1]$ represents the intensity for the kth discrete emotion. Next, I average over all the context j of word w_i and obtain $\overline{v_i} = \left(\frac{\sum_j e_{ij}^1}{10}, \frac{\sum_j e_{ij}^2}{10}, \ldots, \frac{\sum_j e_{ij}^8}{10}\right) \triangleq (e_i^1, e_i^2, \ldots, e_i^8)$ as ground truth emotional intensities for w_i. Finally, I denote the estimated emotional intensities for w_i as $\widehat{v_i} = (\hat{e}_i^1, \hat{e}_i^2, \ldots, \hat{e}_i^8)$ and calculate the mean absolute error as $MAE = \frac{\sum_i \sum_k \left|e_i^k - \hat{e}_i^k\right|}{1000\times 8}$. I find that the MAE is 0.080, which is close to the out-of-sample testing error of 0.073, demonstrating the validity of the newly mined emotion words.

Table 3.1 provides examples of emotion words and their estimated emotional intensities to further illustrate the performance of the proposed algorithm. First, the algorithm correctly distinguishes the subtle differences between similar words; it predicts that "Admire" is used to express love, whereas "Awe" is a blend of love and anxiety. "Grief" has a higher level of sadness compared to "Sad." "Shocking" is predicted to have a higher level of surprise than is "Pleasantly surprised," whereas the latter is a blend of surprise and joy. Second, the algorithm can detect domain-specific words. For example, the algorithm detects the word "Photogenic," a word specific to the movie review domain, and predicts that it can, to some extent, express love.

Table 3.1. Examples of Emotion Words and Estimated Emotional Intensities

Word in Chinese	**Word in English**	**Surprise**	**Sadness**	**Love**	**Joy**	**Disgust**	**Anticipation**	**Anxiety**	**Anger**
敬佩	Admire	0.00	0.00	0.26	0.00	0.00	0.00	0.00	0.00
敬畏	Awe	0.00	0.00	0.24	0.00	0.00	0.00	0.15	0.00
悲伤	Sad	0.00	0.68	0.00	0.00	0.00	0.00	0.00	0.00
悲痛	Grief	0.00	0.81	0.00	0.00	0.00	0.00	0.00	0.00
惊喜	Pleasantly surprised	0.21	0.00	0.00	0.16	0.00	0.00	0.00	0.00
震惊	Shocking	0.74	0.00	0.00	0.00	0.00	0.00	0.00	0.00
镜头感	Photogenic	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00

Note. The translation is produced by Google Translate *(translate.google.com)*.

Valence Analysis. I conduct a two-stage data processing approach to determine the valence of microblogging messages (Hennig-Thurau et al., 2015; T. Song et al., 2019). First, I conduct content classification to identify 1,214,310 microblogging review messages from 3,257,871 microblogging messages and filter out non-review messages, such as spam and unrelated messages. In particular, four coders, who were extensively trained for this annotation task, manually code 20,000 randomly selected microblogging messages into review or non-review messages. Every microblogging message is individually labeled by two different coders. A total of 16,244 messages receive consistent annotations across different coders, 7,193 of which are coded as reviews. I randomly divide the 16,244 messages into a training set, validation set, and test set. With the training set (80% of the annotated 16,244 messages), I conduct Chinese word segmentation with a commonly used tool, *Jieba*, built in Python. I leverage the feature hashing approach and chi-square feature selection method to obtain 5,000 features (T. Song et al., 2019) and train an SVM model based on the selected features. With the validation set (10% of the annotated 16,244 messages), the best parameters are selected as an RBF-kernel with penalty coefficient C = 300 by random parameter searching. With the test set (another 10% of the annotated 16,244 messages), I find that the out-of-sample F1-score is 86.3%. Second, the two coders divide 7,193 microblogging review messages into positive and negative,[12] and I follow the same procedure to obtain another RBF-kernel SVM model, with optimal C equal to 2,500 and an out-of-sample F1-score equal to 90.8% (for accuracy, 94.8%). I then sort 1,214,310 reviews into positive and negative ones with the SVM model. I denote the number of positive reviews for movie m on day t as $Pos_{m,t}$ and total reviews as $ReviewVol_{m,t}$. Then, the review valence is calculated as the ratio of positive reviews to total reviews, i.e., $ReviewVal_{m,t} = \frac{Pos_{m,t}}{ReviewVol_{m,t}}$ (Hennig-Thurau et al., 2015). The mean and standard deviation for $ReviewVol_{m,t}$ are 93.5 and 164.6, respectively. The mean and standard deviation for $ReviewVal_{m,t}$ are 0.8 and 0.2, respectively.

Discrete Emotion Analysis. Following the document-level emotion expression space model (Quan & Ren, 2010), I map each review r_j into an eight-dimensional vector $\{E_k(r_j)\}_{k=1}^{N=8}$, where $E_k(r_j)$ is determined by emotion, negation, and degree words contained in review r_j, representing the intensity of the kth discrete emotion. Negation and degree words are frequently

[12] I used a binary-score annotation instead of a continuous-score annotation for valence because it is the standard procedure of the literature on online review sentiment analysis (Hennig-Thurau et al., 2015; Rui et al., 2013; T. Song et al., 2019). It helps to build a baseline for us to understand the extent to which our approach can outperform the standard procedure. Further, compared to a binary-score annotation, a continuous-score annotation for valence could lead to noise in the annotated results, as it is difficult to achieve agreement among different annotators when they are rating in a continuous manner (Pang & Lee, 2005). Finally, although it is binary at the review level, it is continuous at the daily or weekly level, at which I conduct prediction. This allows valence to vary continuously in our prediction model.

used in Chinese and, thus, are helpful for accurate emotion analysis (Quan & Ren, 2010). I adopt a negation-word dictionary provided by TextMind, a Chinese language psychological analysis system developed by the Chinese Academy of Sciences. The dictionary contains 31 frequently used negation Chinese words. I adopt a degree-word dictionary that contains 60 degree words and their annotated values from Shi (2017). For example, the degree value of the word meaning "the most" in Chinese is annotated as 1.5, and that of the word meaning "kind of" is annotated as 0.8. I use a forward-sliding window to capture these negations and degree words. If the algorithm finds an emotion word in the review, it checks the three words[13] before the emotion word to capture negation and degree words. Then I have:

$$E_k(r_j) = \frac{1}{n_k}\sum_{i=1}^{n_k}(-1)^{m_i} \times DegV_i \times I_k(w_i), \tag{3.1}$$

where $\{w_i\}_{i=1}^{n_k}$ are the emotion words[14] in both review r_j and the lexicon and express the kth discrete emotion. If n_k = 0, then $E_k(r_j)$ is set to zero. $I_k(w_i)$ refers to the kth discrete emotion intensity of w_i. m_i is the total number of negative words that appear in the sliding window of w_i. $DegV_i$ is the average degree value of all degree words that appear in the sliding window of w_i. Let $v_{m,t}$ denote the total number of reviews on day t for movie m. Let $e_{m,t,k}$ represent the kth discrete emotion intensity of movie m on day t, and let $r_{m,t,i}$ denote the ith review on movie m on day t. Then I have:

$$e_{m,t,k} = \frac{\sum_{i=1}^{v_{m,t}} E_k(r_{m,t,i})}{v_{m,t}}, k \in \{1,2,\dots,8\}. \tag{3.2}$$

Latent Emotion Topic Analysis. I am motivated by Yu et al. (2012), who adopt a topic modeling approach (latent semantic indexing [LSI]) to capture latent emotion topics and use them for sales predictions. I follow their methodological framework but use a more advanced

[13] I choose three as the sliding-window size because, in Chinese, negative and degree words are usually within three words before the corresponding emotion word.

[14] One might be concerned that emotion words in movie reviews may be used not only to subjectively express emotions but also to describe the movie plot. The latter may induce noise in understanding consumer emotion when using a lexicon-based approach. I thus conduct an additional study to manually check the extent to which emotion words are plot-description related or subjective-expression related. In particular, I randomly select 200 movie reviews and invite two research assistants to read the reviews, identify all emotion words in these reviews, and sort these emotion words into a plot-description-related class or a subjective-expression-related class. There are 607 emotion words identified from the reviews, 544 (89.6%) of which are identified as subjective-expression-related emotion words. This result indicates that a majority of the emotion words in online movie reviews are used to express one's subjective feelings toward movies and justifies that it is appropriate to adopt a lexicon-based approach to analyze movie viewers' emotions. Further, I note that even plot-description-related emotion words can be helpful in predicting movie sales. For example, these emotion words signal the genre information of a movie. Such information is related to box office revenue and may be useful in a machine learning model.

topic modeling approach, latent Dirichlet allocation (LDA). The LDA approach can overcome the overfitting issue in LSI and provide better latent topic presentations (Blei et al., 2003). I follow the framework by Yu et al. (2012) to formalize the problem: For a given set of reviews $R = \{r_1, r_2, \ldots, r_N\}$ and a set of emotion words from lexicon $L = \{w_1, w_2, \ldots, w_M\}$, the review data can be described as an $N \times M$ matrix $C = \{c(r_i, w_j)\}_{ij}$, where $c(r_i, w_j)$ is the word frequency of w_j in r_i. As such, each row in C is an emotion-word frequency vector that corresponds to a review. I assume that each observation (r_i, w_j), the emotion expression in the review, is generated by an unobservable latent emotion $Z \in \{z_1, z_2, \ldots, z_k\}$.

Similar to LDA, for which hidden factors represent the "topics" of the document (Blei et al., 2003), Z represents the type of latent emotion that the writer would like to express through word w_j (X. Yu et al., 2012). Then, an LDA with the online variational Bayes algorithm is implemented with the Python machine learning tool *sklearn* to estimate the probability that latent emotion Z is embedded in review r, i.e., $P(Z|r)$ and the probability that the emotion word w would be used when the latent emotion is Z, i.e., $P(w|Z)$. Then, each review r_i is mapped into a k-dimensional latent emotion vector LaE_i, where $LaE_i = \left(P(z_1|r_i), P(z_2|r_i), \ldots, P(z_k|r_i)\right)^T$, which is the probability distribution of latent emotions embedded in r_i. Similar to Equation (2), I can summarize LaE_i to the daily level. I use $LaE_{m,t,k}$ to denote the kth latent emotion topic of microblogging review messages on movie m on day t. A key parameter to select is the number of latent emotions, i.e., k. Perplexity, defined as $e^{-\bar{v}}$, where $\bar{v}$ is the log-likelihood per word, is commonly used to select this parameter (Blei et al., 2003). The lower the perplexity that an LDA model produces, the better the fitness of the model to the data. Figure 3.3 shows that the perplexity is minimized when $k = 2$, i.e., the optimal number of latent emotions is 2.[15] The latent emotion estimated by Yu et al. (2012) is not interpretable. In this work, however, I conduct a deeper analysis and derive some insight into these two latent emotions. Formally, given a latent emotion z_k $(k = 1,2)$, the expected intensity of a discrete emotion e_s $(s = 1,2, \ldots, 8)$ can be

[15] I clarify that the perplexity is minimized based only on the training data. In particular, first, I use the lexicon constructed by using only the reviews before January 1, 2018, to capture the emotion words that appear in all of the reviews. Then, I use only emotion word frequency in the reviews before January 1, 2018, to train the LDA model and select the number of topics. Then, for reviews after January 1, 2018, I use the LDA model to predict the latent emotion topics of the reviews. Finally, the predicted latent emotion topics are used as features to predict the movie sales performance after January 1, 2018. This mimics a practitioner's situation, whereby the practitioner trains the LDA and prediction model with all of the historical data available at a time point and then uses the trained models to predict the outcomes after the time point.

derived as $E(e_s|z_k) = \sum_{w \in L} E(e_s|w) \cdot P(w|z_k)$, where $E(e_s|w)$ is given by lexicon L, and $P(w|z_k)$ is from the LDA model. I present the results of $E(e_s|z_k)$ in Figure 3.4.

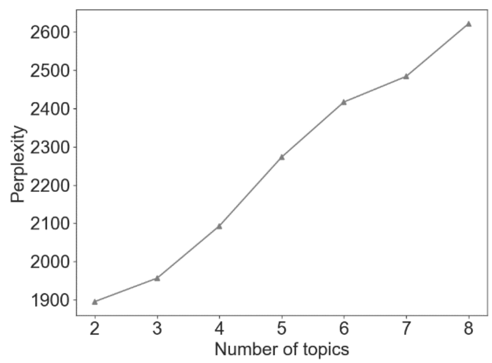


Figure 3.3. Perplexity of the LDA Model across the Number of Topics

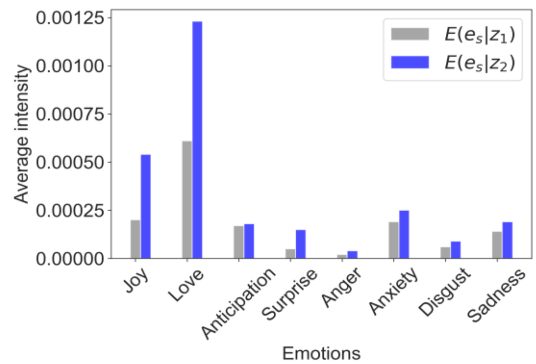


Figure 3.4. Average Discrete Emotion Intensities in Latent Emotions

One might expect that the two latent emotions are simply "positive" and "negative," but I find that the message sent by $E(e_s|z_k)$ involves more than these valences. First, I find that z_2 is more positive than z_1 because, for positive discrete emotions (love and joy), z_2 has an intensity about two times as high as z_1, whereas, for negative discrete emotions (anger, anxiety, disgust, and sadness), the differences between z_1 and z_2 are much smaller. Further, I observe that, for each discrete emotion s, $E(e_s|z_1)$ is less than $E(e_s|z_2)$, which means that, to express latent emotion z_2, a writer would use emotion words that are of high emotional intensity. In other words, z_2 is of higher arousal than z_1. Thus, based on the results, I infer that the two prevalence emotions in movie reviews are (i) positive and activated, with strong love and joy, which resembles "excited" or "delighted" and (ii) less positive and of low arousal, which resembles "bored" or "disappointed."

Box Office Sales Prediction. I adopt a monthly expanding window approach for box office sales predictions, which is commonly used by earlier predictive research (Geva et al., 2017; T. Song et al., 2019). I choose such an approach because it mimics a practitioner's situation, whereby every month, a cinema manager collects all of the available historical data to train a prediction model. In particular, I assume that the time point that a manager adopts the approach is January 1, 2018. Then, for each month t after January 1, 2018, the data samples of the preceding months (month 1 to $t-1$, where month 1 is January 2012) are used as a training set, and the data samples of month t are used as the test set. The lexicon is created by using only the reviews before January 1, 2018. As such, information about patterns of the future will not be used to "predict" the past data points. I use historical daily box office revenue data, emotions in reviews, review volume, screening days, and weekends as predictive variables (Table 3.2), consistent with the existing box office prediction literature (T. Song et al., 2019; X. Yu et al., 2012).

Table 3.2. Description of Variables

Variable	**Description**
$BoxOffice_{m,t}$	The box office of movie m on day t (RMB yuan)
$ReviewVol_{m,t}$	Microblogging review message volume of movie m on day t
$ReviewVal_{m,t}$	Microblogging review message valence of movie m on day t
$e_{m,t,k}$	The kth ($k \in \{1,2,\ldots,8\}$) discrete emotion intensity of microblogging review messages on movie m on day t
$LaE_{m,t,k}$	The kth ($k \in \{1,2\}$) latent emotion topic of microblogging review messages on movie m on day t
$Age_{m,t}$	Total number of movie m screening days from its release day to day t, controlling for the tendency of the box office
$IsWeekend_{m,t}$	Dummy variable for indicating whether day t is the weekend for movie m, controlling for the seasonality of the box office induced by weekends

Extant research notes that many movie-specific time-invariant factors could be influential for total box office sales, including genre, pre-release buzz, production budget, and star and director effects (Hennig-Thurau et al., 2015). With a time-series approach, the effects of these factors can be absorbed in the historical box office variables.

To compare the predictive power of different features, I define feature sets as presented in Table 3.3. The parameter p is the maximum number of time lags. To avoid generating too many undetermined parameters in the prediction model, I used the same time lag effect p for historical box office sales, review volume, valence, and emotion variables. I build prediction models with four different algorithms, i.e., LR, RF, SVR, and XGB, with default settings recommended by *sklearn*. For a given feature set, I first use a default RF model to conduct feature selection on a training set and then use the selected features to predict box office sales with one of four algorithms.

Table 3.3. Feature Sets Variables

Feature Set	Variable
Base	Historical daily box office data, historical review volume, and control variables: $\{BoxOffice_{m,t-1}$, $BoxOffice_{m,t-2}$, …, $BoxOffice_{m,t-p}$, $ReviewVol_{m,t-1}$, $ReviewVol_{m,t-2}$, …, $ReviewVol_{m,t-p}$, $Age_{m,t}$, $IsWeekend_{m,t}\}$.
Val	Historical review valence variables: $\{ReviewVal_{m,t-1}, ReviewVal_{m,t-2}, \dots, ReviewVal_{m,t-p}\}$.
emo8	Historical eight types of discrete emotion variables: $\{e_{m,t-1,k}, e_{m,t-2,k}, \dots, e_{m,t-p,k}\}$, $k \in \{1,2, \dots, 8\}$.
Anger	If I use only one type of discrete emotion as a feature set, the feature set will be named after this kind of emotion. For example, let k = 1 denote the emotion "anger"; then, the *Anger* feature set is: $\{e_{m,t-1,1}, e_{m,t-2,1}, \dots, e_{m,t-p,1}\}$. Similarly, by choosing different values of the parameter k, I can define feature sets *Anxiety*, *Disgust*, *Sadness*, *Anticipation*, *Surprise*, *Love*, and *Joy*.
Latent	Historical latent emotion topic variables: $\{LaE_{m,t-1,k}, LaE_{m,t-2,k}, \dots, LaE_{m,t-p,k}\}$, $k \in \{1,2\}$.

Finally, I evaluate prediction model performance with the mean absolute percentage error (MAPE) between the predicted and observed box office sales on any given day, consistent with the literature (T. Song et al., 2019; X. Yu et al., 2012). Specifically, MAPE $= \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$, where A_t is the actual sales of a movie-date pair $t \in \{1,2, \dots, n\}$ $(n = 929)$ in the test set, and F_t is the predicted sales. I use the LR model and base feature set to determine time lag p. Among $p \in \{1,2, \dots, 7\}$, the MAPE is minimized when $p = 3$. Thus, I use $p = 3$ for further analysis.

3.1.3 *Descriptive and Predictive Results*

I present the correlational analysis of discrete emotions and other predictors in Table 3.4. All of the absolute correlations between discrete emotions and other box office predictors are no more than 0.21. The low correlations show the orthogonality of emotions and other predictors, indicating that emotions provide important complementary information for the existing predictors. I also have several observations. First, all emotions are slightly positively correlated with review volume. This is possibly because more-popular movies also would have a higher volume of reviews, and the more-popular movies are usually better at evoking audience emotions. Second, the four negative emotions, i.e., anger, anxiety, disgust, and sadness, are slightly negatively correlated with review valence. The two positive emotions, i.e., love and joy, are slightly positively correlated with valence. Surprise and anticipation are considered mixed emotions in the existing literature (Nguyen et al., 2020) and can be either positive or negative in valence. For example, a negative review that contains *surprising* could be, "This movie is surprisingly bad." The results show that, on average, surprise and anticipation are slightly negatively correlated with valence. Third, all emotions, except anticipation, are slightly positively correlated with the weekend timing of a movie. This is not unexpected because (i) on weekends, there will be a larger audience and more reviews, and (ii) review volume is positively correlated with emotions. In addition, the slightly negative correlation of anticipation can be because, on weekdays, people usually express their expectations for movie watching in the upcoming weekend. Finally, all emotions are slightly negatively correlated with the number of screening days since a movie was released ($Age_{m,t}$). If viewers are enthusiastic about a movie, they usually watch the movie immediately after its release. The later adopters are usually less emotional than are the earlier adopters, as the results suggest. All of these observations are consistent with common experience and, thus, show the validity of discrete emotion variables.

Table 3.4 Correlations between Emotions and Other Box Office Sales Predictors

Correlation	$ReviewVol_{m,t}$	$ReviewVal_{m,t}$	$IsWeekend_{m,t}$	$Age_{m,t}$
$Joy_{m,t}$	0.09	0.00	0.02	-0.10
$Love_{m,t}$	0.12	0.05	0.01	-0.08
$Surprise_{m,t}$	0.09	-0.15	0.02	-0.08
$Anticipation_{m,t}$	0.10	-0.08	-0.00	-0.08
$Anger_{m,t}$	0.03	-0.16	0.02	-0.04
$Anxiety_{m,t}$	0.10	-0.19	0.02	-0.08
$Disgust_{m,t}$	0.14	-0.21	0.01	-0.05
$Sadness_{m,t}$	0.12	-0.08	0.01	-0.07

Note. All correlations are statistically significant ($p < 0.01$).

I present the correlations between each pair of the eight emotions in Table 3.5. No correlation is no more than 0.31, which validates the orthogonality across discrete emotions and justifies the necessity of investigating all of the emotions.

Table 3.5. Correlations between Discrete Emotions

Correlation	$Joy_{m,t}$	$Love_{m,t}$	$Surprise_{m,t}$	$Anticipation_{m,t}$	$Anger_{m,t}$	$Anxiety_{m,t}$	$Disgust_{m,t}$	$Sadness_{m,t}$
$Joy_{m,t}$	1.00							
$Love_{m,t}$	0.07	1.00						
$Surprise_{m,t}$	0.02	0.04	1.00					
$Anticipation_{m,t}$	0.03	0.02	0.17	1.00				
$Anger_{m,t}$	0.08	-0.03	0.06	0.03	1.00			
$Anxiety_{m,t}$	0.11	0.15	0.12	0.12	0.09	1.00		
$Disgust_{m,t}$	0.07	0.02	0.10	0.09	0.30	0.15	1.00	
$Sadness_{m,t}$	0.09	0.16	0.10	0.17	0.04	0.31	0.13	1.00

Note. All correlations are statistically significant ($p < 0.01$).

To demonstrate the value of constructing an up-to-date and domain-specific lexicon, I compare the prediction performance when using the extended and basic lexicons. First, I construct an up-to-date and domain-specific lexicon based on the reviews from 2012 to 2017 (Lexicon 2017). I construct another domain-specific lexicon based only on reviews up to 2012 (Lexicon 2012). Second, I use the basic lexicon (built in 2008), Lexicon 2012, and Lexicon 2017 to analyze eight types of discrete emotional features in the reviews of 2018. The sets of eight discrete emotions generated by the basic lexicon and Lexicon 2012 are denoted as emo8_basic and emo8_2012, respectively (emo8 is generated by Lexicon 2017). Table 3.6 and Table 3.7 present the model prediction results. Rows (a.0) and (b.0) show the MAPE of two benchmark feature sets, i.e., emo8 and base+emo8 (that combine the base feature set and emo8; in Table 3.6 and Table 3.7, the plus sign indicates the combining of different feature sets), under four prediction models. The lowest MAPE (39.99) is achieved by base+emo8 under SVR.[16] For Rows (a.1) to (a.22), MAPE increase compared to emo8 is derived by $\frac{1}{n}\sum_{t=1}^{n} 100\left(\left|\frac{A_t-F_{tj}}{A_t}\right| - \left|\frac{A_t-F_{t0}}{A_t}\right|\right)$, where F_{t0} is the predicted value of emo8; F_{tj} is that of feature set in Row $(\mathrm{a}.j)$. A positive MAPE increase in Row $(\mathrm{a}.j)$ indicates that using emo8 produces a lower MAPE than using the feature set in Row $(\mathrm{a}.j)$. The standard deviation of the MAPE increase (as shown in parentheses in the table) is the standard deviation of $100\left(\left|\frac{A_t-F_{tj}}{A_t}\right| - \left|\frac{A_t-F_{t0}}{A_t}\right|\right)$ for $t \in \{1,2,\dots,n\}$ divided by $\sqrt{n}$ due to the central limit theorem. Then, a one-sided t-test is used to determine the significance level of the MAPE increase. The same procedure is applied to Rows (b.1) to (b.23), where F_{t0} indicates the predicted value of base+emo8, and F_{tj} indicates the predicted value of the feature sets in Row $(\mathrm{b}.j)$. The feature sets in Row (b.j) are derived by combining the base feature set with the feature set in Row (b.j). I mainly focus on interpreting the results presented in Table 3.7, because the base features are expected to be used in practice. Table 3.6 serves as a benchmark and robustness check to Table 3.7. The MAPE increase in Table 3.7 is smaller than that in Table 3.6. It implies that the predictive advantage, which emo8 has over the other emotion feature sets, partially results from its association with features in the base feature set (historical sales, review volume, seasonality, and timing). Notably, base+emo8 is clearly the method of

[16] Predicting movie sales is a difficult task. This MAPE has largely outperformed the prior research that uses online review sentimentsentiments for box office predictionpredictions (X. Yu et al., 2012) (in the same setup with the time lag as 3, the prior research's MAPE is over 80). I do believe that, for a practitioner, with more data available, such as advertising expenditure data, the MAPE can be further reduced.

choice in this setting, because its predictive power is always superior or at least equivalent to that of the other alternative feature sets studied in this work.

Table 3.6. MAPE of Emotion Feature Sets across Different Prediction Models

No.	Feature Set	LR	RF	SVR	XGB	No.	Feature Sets	LR	RF	SVR	XGB
a.0	**MAPE of emo8**	**116.76**	**124.90**	**118.21**	**135.42**						
	MAPE Increase compared to emo8						**MAPE Increase compared to base+emo8**				
a.1	emo8_basic	11.65*** (3.45)	33.82*** (10.81)	6.82*** (1.72)	25.32** (14.78)	a.12	Sadness	15.28*** (5.74)	186.74*** (61.13)	10.23*** (3.98)	33.26** (17.70)
a.2	emo8_2012	9.39*** (3.29)	18.55** (9.60)	1.40** (0.65)	12.78 (11.05)	a.13	Surprise	15.95*** (5.29)	170.96*** (46.27)	12.09*** (3.87)	81.72*** (27.62)
a.3	latent	19.74*** (6.33)	193.74*** (62.58)	32.78*** (8.82)	86.73*** (33.93)	a.14	val+Anger	12.65*** (4.66)	106.4*** (33.39)	14.65*** (4.48)	85.03*** (34.82)
a.4	latent+emo8	4.87*** (1.65)	20.87** (9.01)	8.04*** (2.29)	14.57** (7.92)	a.15	val+Anxiety	14.46*** (5.05)	149.76* (90.84)	16.84*** (5.13)	84.52*** (31.57)
a.5	val	12.98*** (4.74)	71.72*** (22.57)	18.44*** (5.65)	56.04*** (22.29)	a.16	val+Anticipation	15.89*** (5.35)	59.2*** (20.75)	17.12*** (5.03)	48.19*** (18.38)
a.6	Anger	10.28*** (4.07)	83.61*** (33.73)	4.17** (2.09)	39.40** (18.31)	a.17	val+Disgust	18.15*** (6.27)	72.12** (34.76)	16.21*** (4.99)	47.99** (26.63)
a1.7	Anxiety	11.92*** (4.39)	83.79*** (22.09)	7.83*** (2.81)	36.31*** (13.40)	a.18	val+Joy	13.95*** (5.64)	51.22*** (17.90)	14.95*** (5.20)	76.20*** (31.94)
a.8	Anticipation	14.02*** (4.88)	86.01*** (32.10)	9.82*** (3.33)	37.33** (19.88)	a.19	val+Love	21.38*** (6.16)	69.68*** (27.08)	21.51*** (5.80)	56.73*** (23.62)
a.9	Disgust	14.30*** (5.34)	15.34* (9.98)	7.97*** (2.90)	-14.38 (11.43)	a.20	val+Sadness	15.88*** (5.70)	106.95*** (35.79)	16.50*** (5.35)	33.26** (17.90)
a.10	Joy	14.00** (6.42)	66.95*** (25.41)	9.66** (4.98)	67.27 (55.55)	a.21	val+Surprise	17.30*** (5.82)	74.01*** (21.34)	18.87*** (5.59)	96.66*** (33.07)
a.11	Love	30.48*** (10.11)	154.72*** (61.86)	15.97*** (4.45)	65.40*** (26.34)	a.22	val+emo8	22.02*** (6.64)	72.88*** (29.99)	17.66*** (5.09)	60.71** (26.87)

Note. Standard deviations in parentheses. MAPE = $\frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t-F_t}{A_t}\right|$, where A_t is the actual value of a movie-date pair $t \in \{1,2,\dots,n\}$ $(n = 929)$ in the testing set, and F_t is the predicted value. For Row (a.1) to (a.22), the MAPE increase is derived by $\frac{1}{n}\sum_{t=1}^{n} 100\left(\left|\frac{A_t-F_{tj}}{A_t}\right| - \left|\frac{A_t-F_{t0}}{A_t}\right|\right)$, where F_{t0} is the predicted value of emo8, and F_{tj} is that of the feature set in Row (a.j). The standard deviation of the MAPE increase is the standard deviation of $100\left(\left|\frac{A_t-F_{tj}}{A_t}\right| - \left|\frac{A_t-F_{t0}}{A_t}\right|\right)$ divided by $\sqrt{n}$ due to the central limit theorem. Then, a one-sided *t*-test is used to determine the significance level of the MAPE increase. *p < 0.10, **p < 0.05, ***p < 0.01.

Table 3.7. MAPE of Emotion and Base Feature Sets across Different Prediction Models

No.	Feature Set	LR	RF	SVR	XGB	No.	Feature Sets	LR	RF	SVR	XGB
b.0	**MAPE of base+emo8**	**44.40**	**44.87**	**39.99**	**41.00**						
	MAPE Increase compared to emo8						**MAPE Increase compared to base+emo8**				
b.1	base+emo8_basic	0.05 (0.13)	1.22 (1.08)	0.06** (0.04)	1.78* (1.22)	b.13	base+Surprise	1.33** (0.66)	3.12** (1.45)	0.70 (0.67)	0.36 (1.66)
b.2	base+emo8_2012	0.20 (0.19)	0.45 (0.98)	0.03 (0.04)	1.02 (1.04)	b.14	base+val+Anger	1.37** (0.67)	1.40 (1.54)	1.30** (0.68)	0.79 (1.04)
b.3	base+latent	1.35** (0.66)	2.41* (1.58)	1.67*** (0.68)	1.94 (1.68)	b.15	base+val+Anxiety	1.36** (0.64)	4.95*** (1.86)	1.38** (0.69)	1.82 (1.58)
b.4	base+latent+emo8	0.33*** (0.11)	0.00 (1.33)	0.59*** (0.15)	2.31** (1.14)	b.16	base+val+Anticipation	1.32** (0.66)	4.67*** (1.87)	1.34** (0.68)	2.13* (1.48)
b.5	base+val	1.26** (0.67)	5.12*** (1.86)	1.02* (0.72)	0.28 (1.37)	b.17	base+val+Disgust	1.26** (0.64)	4.34*** (1.70)	1.30** (0.69)	1.17 (1.43)
b.6	base+Anger	1.37** (0.66)	1.37 (1.63)	0.68 (0.67)	0.85 (1.26)	b.18	base+val+Joy	1.31** (0.66)	2.30* (1.55)	1.32** (0.69)	2.94** (1.53)
b.7	base+Anxiety	1.34** (0.65)	3.41*** (1.43)	0.72 (0.67)	1.69 (1.72)	b.19	base+val+Love	1.33** (0.67)	1.22 (1.19)	1.27** (0.67)	2.18* (1.45)
b.8	base+Anticipation	1.31** (0.66)	3.68*** (1.34)	0.69 (0.66)	2.98** (1.56)	b.20	base+val+Sadness	1.42** (0.66)	5.23*** (1.79)	1.36** (0.68)	2.05* (1.58)
b.9	base+Disgust	1.24** (0.64)	4.42*** (1.71)	0.69 (0.67)	0.71 (1.25)	b.21	base+val+Surprise	1.29** (0.66)	2.82** (1.51)	1.32** (0.69)	1.48 (1.45)
b.10	base+Joy	1.30** (0.66)	3.13** (1.41)	0.70 (0.67)	3.35** (1.75)	b.22	base+val+emo8	0.15 (0.14)	0.38 (0.90)	0.40*** (0.10)	2.74*** (1.13)
b.11	base+Love	1.33** (0.66)	4.76*** (1.45)	0.66 (0.65)	0.91 (1.46)	b.23	base	1.41** (0.74)	8.80*** (2.01)	6.19*** (1.06)	6.54*** (2.07)
b.12	base+Sadness	1.39** (0.66)	4.27*** (1.62)	0.70 (0.67)	1.19 (1.64)						

Note. Standard deviations in parentheses. The same procedure used in Table 3.6 is applied to Rows (b.1) to (b.23), where F_{t0} indicates the predicted value of base+emo8, and F_{tj} indicates the predicted value of the feature sets in Row $(b.j)$. *p < 0.10, **p < 0.05, ***p < 0.01.

Value of Lexicon Extension. Rows (b.1) and (b.2) of Table 3.7 show that emo8_basic and emo8_2012 consistently generate a higher MAPE than does emo8. Rows (a.1) and (a.2) of Table 3.6 show that these results are qualitatively consistent and statistically more significant. The MAPE differences between emo8 and emo8_2012 result only from constructing an up-to-date lexicon, as they are both generated by domain-specific lexicons. The MAPE differences between emo8 and emo8_2012 are mostly smaller than those between emo8 and emo8_basic, which indicates that, in addition to constructing an up-to-date lexicon, constructing a domain-specific lexicon plays an important role in reducing the MAPE. The MAPE differences between base+emo8_2012 and base+emo8 are insignificant in Table 3.7, implying that constructing a domain-specific lexicon is more effective in reducing the MAPE than is constructing an up-to-date lexicon. In other words, although emotional expressions within the domain of movie reviews evolves from 2012 to 2017, the difference in emotional expressions caused by such an evolution is smaller than the difference caused by domain specificity.

Discrete Emotions and Latent Emotion Topics. I then investigate the predictive power of latent emotions as compared to discrete emotions. Row (b.3) of Table 3.7 shows that discrete emotions outperform latent emotion topics. Row (b.4) of the table shows that combining discrete emotions and latent emotion topics achieves a worse prediction result than using discrete emotions alone. Rows (a.3) and (a.4) in Table 3.6 show that these results are qualitatively robust and statistically significant. The results can be explained by discrete emotion theory, which argues that complex emotions are mixtures of discrete emotions. The information of the latent emotions, as complex emotions, should already be absorbed by the discrete emotions, and, thus, latent emotions bring redundant information to predictions. Combining emo8 and latent emotions may lead to overfitting issues. SVR and XGB are more flexible models and are more likely to suffer from overfitting issues than are LR and RF. Accordingly, base+latent+emo8 has a larger MAPE increase under SVR and XGB than LR and RF in Row (b.4) of Table 3.7. Further, latent emotion topics are generated based on the data-driven approach, i.e., topic modeling (Blei et al., 2003; X. Yu et al., 2012). There is no existing theory, to my knowledge, guaranteeing that they are unbiased measures to consumer emotions. Therefore, measurement errors may contaminate the prediction.

Discrete Emotions and Valence. Next, I investigate the predictive power of valence, discrete emotions, and the combination of the two sets of features. Rows (a.5) and (b.5) show that discrete emotions outperform valence in predictive power. Rows (a.22) and (b.22) of the table show that combining valence and discrete emotions cannot increase the predictive power compared to using discrete emotions alone, which suggests that discrete emotions and valence are substitutive rather than complementary. Compared to discrete emotions, valence is relatively a simple and biased proxy of consumer sentiments (Lerner et al., 2015). For example, consumers' emotions with similar valence (sadness and disgust, anxiety and anger) are linked to distinct cognitive and behavioral reactions (Lerner et al., 2004; Yin et al., 2014). Further, if valence is just a biased predictor, the orthogonality of valence and emotions would not benefit the prediction in a meaningful way. Therefore, similar to latent emotions, the issues of overfitting and measurement errors may happen to valence as well. For each discrete emotion, I have similar conclusions. Under SVR (the model with the highest accuracy and lowest variance), Rows (a.5) to (a.21) and Rows (b.5) to (b.21) show that each discrete emotion generates a lower MAPE than in combination with valence or when using valence alone. Third, the values of MAPE increase in Rows (a.6) to (a.13) and Rows (b.6) to (b.13) are all positive (except disgust in Row (a.9) under XGB), indicating that combining all discrete emotions (emo8) outperforms using each discrete emotion alone. Indeed, discrete emotion theory (Lerner et al., 2015; Plutchik & Kellerman, 1982; Tomkins, 1962) argues that discrete emotions are relatively independent of each other (consistent with the results in Table 3.5). To combine all of the emotions means to combine orthogonal and meaningful information. The above exception (disgust) indicates that disgust is of the highest predictive power among the eight discrete emotions. This is also consistent with the experimental results, presented later, where I show that the effects of disgust on perceived review helpfulness and purchase intention are greater than the effects of most of the other emotions. Nevertheless, the best performance is achieved by using base+emo8, which indicates other discrete emotions still provide complementary predictive power to disgust.

Additional Baseline Models. To further establish the relevance of the machine learning approach and all of the machinery involved, I test two more baseline models against the best model, base+emo8 under SVR. First, I use only the previous-day sales variable as the predictor. I find that the MAPE for the first baseline model is 46.73, which is significantly higher than that of base+emo8 (46.73 vs. 39.99, paired t-test p-value < 0.001). Second, I use the Word2Vec representation for reviews and combine this with the base feature set to predict next-day sales. Specifically, I map each word in a review to a 200-dimension word vector. Then, the Word2Vec representation for a review is the average of all word vectors. Research has empirically shown that this is a successful and efficient way of obtaining sentence semantic information (Arora et al., 2017; Kenter et al., 2016). It is also a common practice in deep learning to first learn low-level high-dimension and latent representations for unstructured textual data and then to combine them with the traditional machine learning models, such as the SVR, to predict the outcomes. Using this baseline model achieves a statistically significantly lower MAPE than using the base feature set (45.39 vs. 46.18, paired t-test p-value = 0.093), showing that latent semantic representations provide useful information beyond the base feature set. Nevertheless, the MAPE of this baseline model is still significantly higher than that when using base+emo8 (45.39 vs. 39.99, paired t-test p-value < 0.001).[17]

Weekly Predictions. I include an additional analysis of weekly predictions. I summarize variables to a weekly level and follow the same procedure to make weekly predictions. I find that the MAPE by the base feature set is 57.53%, which is the same as that by combining the Base and Val feature sets. This indicates that the valence feature is not predictive of weekly box office sales and is dropped by the RF model in the feature selection process. Combining discrete emotions and the base feature set, in contrast, produces a significantly lower MAPE, at 53.54% (using a paired t-test, p-value = 0.003). These results are consistent with the daily prediction results.

3.2 STUDY 2: A RANDOMIZED EXPERIMENT ON CAUSALITY AND ITS MECHANISM

3.2.1 *Motivation*

The predictive results further motivate this work to study the effects of eight types of discrete emotions in an experimental study, as Study 1 suggests that discrete emotions are more fundamental than are valence and complex emotions in terms of understanding emotions in online reviews. An important question is whether the relationship between discrete emotions in

[17] This result can be due to the following. First, although the Word2Vec representations may comprehensively contain semantic information not limited to emotional expressions, much of this semantic information can be irrelevant to consumer decisions, which merely adds noise to the prediction model. Second, the prediction task becomes challenging when directly adding word vectors to the model due to their high-dimension nature. I have 499 movie samples, which is comparable to the dimensions of word vectors. Thus, the model performance may be undermined due to dimensionality and overfitting. Third, when I construct discrete emotion variables, I incorporate complementary information from the basic lexicon (Ren-CECps). The basic lexicon contains human-annotated information about how the word is associated with discrete emotions. Such information is not included in a word vector. This additional information has been discussed in cognitive psychological theories (Klauer, 1997; Schwarz, 1990; Van Kleef, 2009) and shown by our experimental study to have significant impacts on consumer decisions, thereby helping to improve prediction accuracy.

online reviews and sales is causal or driven by confounding factors. If the relationship is causal, I can expect the predictive relationship to be robust. Otherwise, the predictive power may vanish when the confounding factors change. In this sense, a causal examination can complement the results of predictions (Hofman et al., 2017). More importantly, the predictive results of machine learning algorithms are not interpretable. An interpretable causal mechanism would complement my predictive study in several ways. First, from a theoretical perspective, it adds to our understanding of how discrete emotions in online reviews affect individuals' purchase intention. Second, understanding the mechanism would improve the external validity of the predictive results. If the effects of emotions originate from enhancing review perceived processing fluency, which is not context-specific, the effects of emotions can be generalized to similar contexts, such as online reviews of restaurants, books, and music. Further, if a mediating effect exists, I can expect that the predictive relationship is stronger when consumers face information overload, and that the predictive relationship may be weaker when the information that consumers need to process is relatively simple. This also would generate additional managerial implications. For example, firms may decide whether to adopt predictive analytics with emotion analysis based on the level of information complexity in reviews that their targeted customers encounter. Another implication is that online review platforms may design features to make information processing more fluent, which contributes to review helpfulness. Finally, the effects of emotions may be different across cultures. Examining this relationship in the Chinese and U.S. movie markets can increase the cross-culture generalizability of the results.

3.2.2 *Stimulus Materials*

To manipulate eight discrete emotions in online reviews, I prepared stimulus materials using real-world online movie reviews. As shown in Table 3.8, first, I identify two reviews that are without specific expression of the eight discrete emotions as positive and negative baseline reviews.

Table 3.8. Stimulus Materials

Group	**Review**
Positive Baseline	The three-act narrative structure is clear and the second act is interesting, full of sly humor. The actor and actress's performance is good as well. The film succeeds in terms of character development.
Surprise	<u>I am very surprised after watching this movie.</u> The three-act narrative structure is clear and the second act is interesting, full of sly humor. The actor and actress's performance is good as well. The film succeeds in terms of character development.
Anticipation	<u>I anticipated the movie to be good before I watched it, and indeed,</u> the three-act narrative structure is clear and the second act is interesting, full of sly humor. The actor and actress's performance is good as well. The film succeeds in terms of character development.
Love	<u>I love the movie very much!</u> The three-act narrative structure is clear and the second act is interesting, full of sly humor. The

Group	Review
	actor and actress's performance is good as well. The film succeeds in terms of character development.
Joy	It was very joyful to watch the movie! The three-act narrative structure is clear and the second act is interesting, full of sly humor. The actor and actress's performance is good as well. The film succeeds in terms of character development.
Negative Baseline	What kept this film from being great was the failure to show us the purpose behind the entire drive and desire to start this project in the first place. They didn't show what the end of the project had looked like, what the investors felt, how actress and her company ended up, and most of all how the two main characters continued with life.
Anger	I got very angry after watching the movie! What kept this film from being great was the failure to show us the purpose behind the entire drive and desire to start this project in the first place. They didn't show what the end of the project had looked like, what the investors felt, how actress and her company ended up, and most of all how the two main characters continued with life.
Anxiety	I felt very anxious toward the movie. If the director makes another movie like this, I'd be concerned! What kept this film from being great was the failure to show us the purpose behind the entire drive and desire to start this project in the first place. They didn't show what the end of the project had looked like, what the investors felt, how actress and her company ended up, and most of all how the two main characters continued with life.
Sadness	I felt very sad after watching the movie. What kept this film from being great was the failure to show us the purpose behind the entire drive and desire to start this project in the first place. They didn't show what the end of the project had looked like, what the investors felt, how actress and her company ended up, and most of all how the two main characters continued with life.
Disgust	I felt very disgusted after watching the movie! What kept this film from being great was the failure to show us the purpose behind the entire drive and desire to start this project in the first place. They didn't show what the end of the project had looked like, what the investors felt, how actress and her company ended up, and most of all how the two main characters continued with life.

To identify reviews that were positive or negative in valence, but without specific expressions of the eight discrete emotions, I focused on 11 movies (*Wonder Park*, *Five Feet Apart*, *No Manches Frida 2*, *Captive State*, *Dominirriquenos 2*, *The Mustang*, *Faith*, *Hope and Love*, *The Aftermath*, *Ash is Purest White*, *The Hummingbird Project*, and *Combat Obscura*) that were released on March 15, 2019. I scanned all 706 reviews of the 11 movies on IMDb and deleted redundant reviews and those of extreme length. I removed (1) the movie title, (2) emoji images, and (3) the emotional content that directly indicates reviewers' discrete emotions, such as "I love this movie," and "Disgusting experience," and collected 18 non-emotional reviews as baseline samples (nine positive baseline reviews and nine negative baseline reviews) (Yin et al., 2014). Then, 50 participants were hired from MTurk to rate the emotions in reviews, based on the following question, "In your opinion, to what extent does each of the following words describe how the reviewer felt when he/she wrote the above review? (Surprise, Anticipation, Love, Joy, Anger, Anxiety, Disgust, Sadness), using a 9-point Likert-type scale of 1 = not at all to 9 = very much) (Yin et al., 2014). The results of an ANOVA show that the eight reviews demonstrate no significant difference among the various discrete emotions ($ps > 0.052$). I checked the eight reviews and chose two (one positive baseline review and one negative baseline review) that are feasible in regard to priming different discrete emotions.

Second, following the practice of Yin et al. (2014) and the definition of eight discrete emotions (Table 3.9), I add emotional stimuli sentences at the beginning of the baseline reviews (the underlined sentences in Table 3.8) to manipulate the eight discrete emotions.

Table 3.9. Definition of Eight Discrete Emotions

Discrete Emotion	Definition
Surprise	Surprise has one core appraisal, which is whether something is novel or unexpected (Scherer, 2001).
Anticipation	Anticipation occurs in a situation of looking for the purpose; it implies that the situation is predictable (Plutchik, 1984).
Love	Love is conceived of as a blend of joy and acceptance (Plutchik, 1984).
Joy	Joy, according to the Oxford English Dictionary, is a feeling of great happiness.
Anger	Anger is an emotional state that motivates a person to alleviate personal harm attributed to others and is characterized by states of heightened arousal or activation (Yin et al., 2014).
Anxiety	Anxiety is an unpleasant affective state characterized by uneasiness and uncertainty (C. A. Smith & Ellsworth, 1985).
Sadness	Sadness arises from loss and helplessness (Lazarus, 1991) and evokes the implicit goal of changing one's circumstances (Lerner et al., 2004).
Disgust	Disgust revolves around the appraisal theme of being too close to an indigestible object or idea (Lazarus, 1991).

3.2.3 *Procedure*

I conduct a between-subject experiment, for which a total of 700 participants are hired from MTurk,[18] an online experiment platform. The ten reviews in Table 8 define the ten experimental groups, and each participant is randomly assigned to one group. The two groups with baseline reviews are the control groups, and the rest are the treatment groups. After removing the samples with an unreasonably short completion time (less than 4 seconds per question) and those that could not pass the attention check,[19] I achieved a valid sample of 634 participants. Of all participants, 59.9% are female; 53.1% are between 21 and 39 years old, and 46.7% are 40 years old or above; and 75.2% consider online reviews before they go to the movies, and 21.3% may consider online reviews (I provide balance checks in Appendix A.4). After reading the review,[20] the participant rates (1) emotion embedded in reviews (Yin et al., 2014); (2) processing fluency (Storme et al., 2015); (3) perceived helpfulness of the reviews (Sen & Lerman, 2007); (4) purchase intention for the movie (Lee et al., 2014); and (5) perceived writer rationality (Xiao et

[18] I recruited only MTurk Master Workers in our study to ensure the quality of the responses. According to MTurk, "A Master Worker is a top Worker of the MTurk marketplace that has been granted the Mechanical Turk Masters Qualification. These Workers have consistently demonstrated a high degree of success in performing a wide range of human intelligence tasks across a large number of Requesters."

[19] I add an instruction for the attention check, "Please leave this question blank," to ensure the quality of the responses (Desimone et al., 2015).

[20] Consistent with the practice in IS research on online reviews, I ask the participants to read one review and then report outcomes (Xiao et al., 2018; Yin et al., 2014, 2017) because inviting the participants to read multiple reviews may induce confounding factors. For example, the recall of emotions in the last review may affect the effectiveness of emotions in the current review. Future research could investigate the case in which participants read a portfolio of emotional reviews.

al., 2018).[21] All questions are the same as in previous studies (Table 3.10) and are measured using the 9-point Likert-type scale.

Table 3.10. Variables Measured in the Experiment

Variable	Measurement
Emotion (Yin et al., 2014)	In your opinion, to what extent does each of the following words describe how the reviewer felt when he or she wrote the above review? (Surprise, Anticipation, Love, Joy, Anger, Anxiety, Disgust, Sadness) (1 = not at all, 9 = very much)
Perceived Processing Fluency (Storme et al., 2015)	To what extent do you agree the following statements when you read the review? • *I have trouble fully understanding the meaning.* • *I get the meaning easily.* • *I understand the message without any problem.* • *I find it is complicated to get the message.* • *I have no difficulty understanding the meaning.*
Perceived Writer Rationality (Xiao et al., 2018)	To what extent do you agree the following statements when you read the review? • *The writer presented a justification for his or her information.* • *The writer used logic to express his or her viewpoints.* • *The writer explained the reasoning behind his or her review.*
Perceived Helpfulness (Sen & Lerman, 2007)	Using the scale below, how would you describe the above consumer review? • *not at all helpful/very helpful* • *not at all useful/very useful* • *not at all informative/very informative*
Purchase intention (Lee et al., 2014)	• *The likelihood of your watching the movie* • *The probability that you would consider watching this movie* • *Your willingness to watch the movie*

3.2.4 *Results*

I first conduct a manipulation check to ensure that each type of emotional content is successfully targeted, as shown in Table 3.11. In the two control groups, I expect that there is no difference among the emotions (the null hypothesis). I conduct ANOVA tests on the two control groups (for positive baseline, $p = 0.342$, $n = 73$; for negative baseline, $p = 0.877$, $n = 51$), and the results show that I cannot reject the null hypothesis. Further, in the eight treatment groups, the t-test shows that each treated emotion in the treatment group is significantly higher than that in the corresponding control group ($ps < 0.10$), indicating that stimulus materials are effective. I present the reliability and validity check of three major constructs in Table 3.12. Cronbach's alphas for perceived processing fluency (PF), perceived review helpfulness (PH), purchase intention (PI), and perceived writer rationality (PR) are 0.92, 0.96, 0.97, and 0.81, respectively, indicating adequate construct reliability. Given that the four scales were used in previously validated

[21] Perceived writer rationality is studied in Xiao et al. (2018) and is shown to be relevant to anger in reviews. As I detail later, this helps us to understand the mechanism of the effect of anger.

research, I conduct confirmatory factor analysis to assess convergent and discriminant validity. The average variances extracted (AVEs) for the four constructs are all above 0.7, demonstrating convergent validity; the AVE of each of the constructs is higher than the highest squared correlation with any other latent variable (0.64), demonstrating discriminant validity (Fornell & Larcker, 1981).

Table 3.11. Manipulation Check

Emotion	**Mean of Control Group**	**Mean of Treatment Group**	***p*-value of *t*-test**
Surprise	5.685 (73)	6.362 (69)	0.052 (142)
Anticipation	6.123 (73)	6.985 (68)	0.004 (141)
Love	5.877 (73)	6.576 (59)	0.035 (132)
Joy	6.274 (73)	6.767 (60)	0.073 (133)
***p*-value of ANOVA**	**0.342 (73)**	--	--
Anger	5.510 (51)	7.629 (62)	0.000 (113)
Anxiety	5.451 (51)	6.576 (66)	0.000 (117)
Sadness	5.490 (51)	6.918 (61)	0.001 (112)
Disgust	5.804 (51)	7.292 (65)	0.001 (116)
***p*-value of ANOVA**	**0.877 (51)**	--	--

Note. *Sample size in parentheses.

Table 3.12. Reliability and Validity of Constructs

			Squared Correlation of Factors			
Construct	**Cronbach's Alpha**	**AVE in CFA**	**PF**	**PH**	**PI**	**PR**
PF	0.92	0.78	1.00			
PH	0.95	0.88	0.52	1.00		
PI	0.97	0.91	0.32	0.42	1.00	
PR	0.81	0.77	0.50	0.64	0.26	1.00

Note. PF = Perceived processing fluency; PH = Perceived review helpfulness; PI = Purchase intention; AVE = Average variance extracted; PR = Perceived writer rationality; CFA = Confirmatory factor analysis.

Table 3.13. Regression Results

Variable	(1) DV: PI	(2) DV: PH	(3) DV: PF	(4) DV: PH	(5) DV: PI	(6) DV: PH	(7) DV: PF	(8) DV: PH	(9) DV: PR	(10) DV: PH	(11) DV: PI
Anticipation	0.48*** (0.11)	0.40** (0.16)	1.02*** (0.08)	-0.02 (0.13)							
Joy	0.24*** (0.08)	0.13 (0.09)	0.78*** (0.08)	-0.19** (0.08)							
Love	1.22*** (0.11)	0.94*** (0.11)	1.34*** (0.09)	0.38*** (0.10)							
Surprise	0.86*** (0.12)	0.93*** (0.08)	0.93*** (0.10)	0.54*** (0.08)							
Anger					-2.49*** (0.30)	-0.53** (0.24)	0.34* (0.18)	-0.67*** (0.19)	-0.49*** (0.05)	-0.61 (0.37)	
Anxiety					-0.53*** (0.02)	-0.63*** (0.04)	-0.46*** (0.02)	-0.44*** (0.04)			
Disgust					-2.48*** (0.26)	-1.67*** (0.22)	-0.54*** (0.18)	-1.44*** (0.16)			
Sadness					-2.60*** (0.28)	-0.39* (0.20)	-0.28* (0.16)	-0.27* (0.14)			
PH											0.41*** (0.15)
PF				0.42*** (0.07)				0.42*** (0.04)			
PR										0.53** (0.24)	
Frequency	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviews	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intercept	5.08*** (0.19)	5.38*** (0.14)	6.47*** (0.16)	2.68*** (0.51)	5.80*** (0.63)	5.63*** (0.60)	5.29*** (0.29)	3.40*** (0.48)	5.92*** (0.13)	4.44*** (1.28)	2.69*** (0.58)
N	329	329	329	329	305	305	305	305	113	113	634
Adj. R^2	0.03	0.05	0.05	0.21	0.33	0.12	0.06	0.31	0.11	0.16	0.16

Note. Each emotion variable is a dummy variable that indicates whether a participant belongs to the corresponding treatment group; PI = Purchase intention, PH = Perceived review helpfulness, PF = Perceived processing fluency, PR = Perceived writer rationality. Frequency, Age, Gender, Reviews control for subjects' frequency of movie watching, ages, genders, and whether they consider online reviews before purchasing a movie ticket, respectively. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3.13 presents the results of ordinary least square (OLS) estimation. Each emotion variable in the table is a dummy variable that indicates whether a participant belongs to the corresponding treatment group and estimates the effect of the corresponding treatment condition. I control participants' frequency of going to the cinema (Freq), ages (Age), genders (Gender), and whether they consider online reviews before purchasing a movie ticket (Reviews) in all regressions.

In Columns (1) to (4), I use the samples in the positive baseline group and four positive emotion groups (n = 329). Columns (1) to (3) show that each positive emotion has a positive effect on PI, PH, and PF, respectively. Other than joy on PH, all coefficients are statistically significant. Online reviews with joyful expressions may be perceived as less helpful because overt expressions of joy in content tend to signal the writer's overconfidence (Jiang et al., 2019). Column (4) shows the result of adding the mediator PF to the regression of PH. Compared to the results in Column (2), all coefficients of positive emotions in Column (4) are smaller, indicating that PF mediates the effects of all positive emotions on PH (Baron & Kenny, 1986). In Columns (5) to (8), I use the samples in the negative baseline group and four negative emotion groups (n = 305). Columns (5) to (7) show that each negative emotion has a significantly negative effect on PI, PH, and PF (except anger on PF), respectively. Column (8) shows the result of adding the mediator PF to the regression of PH. Compared to Column (6), all coefficients of negative emotions in Column (8) are smaller in effective size (except anger), indicating that PF mediates the effects of anxiety, sadness, and disgust on PH. Anger is different from other negative emotions because it reflects an individual's irrationality and low coping efforts (Kalamas et al., 2008; Xiao et al., 2018). A review reader is likely to attribute anger to the writer's irrationality and then to stop cognitively processing the review information (Junyong Kim & Gupta, 2012; Xiao et al., 2018). This is consistent with my experimental results, in which anger can increase perceived processing fluency but decrease perceived review helpfulness, as the reader may quickly judge the angry review as of low informative value. To further confirm this deduction, I conduct an additional analysis with samples in the anger group and negative baseline group. Column (9) shows that, indeed, anger negatively affects PR. Column (10) shows that, when adding the mediator PR to the regression of PH on anger, the effect of anger becomes insignificant, indicating a complete mediation effect of PR on anger (Baron & Kenny, 1986). Column (11) confirms the positive effect of PH on PI. Notably, because my analysis contains multiple statistical inferences simultaneously, the risk of Type 1 error (false significance) increases.

To address this concern, I conduct a Bonferroni correction (Simes, 1986). The results show that my findings are robust. In statistics, the Bonferroni correction is a method to counteract the multiple comparisons problem, i.e., the risk of false significance increases when considering multiple statistical tests simultaneously. Although the classical Bonferroni correction can mitigate the concern of false significance, it is a conservative method that gives the chance of failure to reject a false null hypothesis, because it ignores potentially valuable information (Simes, 1986; Holm, 1979). Therefore, in what follows, I employ an improved Bonferroni procedure (Simes, 1986). Formally, let $P_{(1)}, \ldots, P_{(n)}$ be the ordered p-values ($P_{(1)}$ is the smallest value) for testing hypotheses $H_{(1)}, \ldots, H_{(n)}$. $H_{(i)}$ ($i \in \{1, \ldots, n\}$) is rejected provided that $P_{(i)} < i\alpha/n$ and $H_{(1)}, \ldots, H_{(i-1)}$ have all been rejected, where α is the desired overall alpha level and n is the number of hypotheses. In my model, $H_{(i)}$ is the hypothesis that a certain regression coefficient is significantly different from zero and n is the total number of coefficients in 11 regressions in

Table 3.13.

Finally, I normalize the treatment effect of each discrete emotion on PI and PH to the same scale as shown in Table 2.1. The normalized effect of surprise, anticipation, love, joy, anger, anxiety, sadness, and disgust on PI (PH) is 1.27 (1.38), 0.55 (0.47), 1.75 (1.35), 0.49 (0.26), -1.18 (-0.25), -0.47 (-0.56), -1.82 (-0.28), and -1.66 (-1.12), respectively. All effect sizes are higher than the existing non-emotional factors listed in Table 1, which demonstrates the economic importance of discrete emotions in online reviews.

I review the literature to derive theoretical insight into the heterogeneous effects of emotions on PI. First, because movie consumers tend to be arousal-seeking (Xie & Lee, 2008), high-arousal positive emotional experiences (i.e., love and surprise) associated with movie consumption are more rewarding than low-arousal ones (i.e., joy and anticipation). Similarly, high-arousal negative emotional experiences (i.e., anxiety and anger) are less detrimental to PI than low-arousal ones (i.e., sadness and disgust). Second, low-arousal positive/negative emotions are of similar effectiveness, but high-arousal positive/negative emotions are not. The difference between high-arousal emotions can be explained based on extant research. Love is an emotion implying high levels of certainty, trust, and acceptance (Lazarus, 1991; Plutchik, 1984), whereas surprise is a more neutral emotion compared to love (Nguyen et al., 2020). Therefore, a writer's expressions of love tend to generate a greater endorsement toward a movie than do expressions of surprise. Expressions of anger are especially likely to be noticed, encoded, recalled, and can directly affect a consumer's attitude (Yin et al., 2021). Anxiety in reviews tends to increase perceived uncertainty of product quality (Yin et al., 2014), which may only indirectly affect purchase intention through a reader's risk-averse tendency. Therefore, anger tends to be more effective in decreasing the purchase intention than does anxiety.

3.3 Discussion

This chapter makes several contributions to the information systems literature in regard to emotions in online reviews and sales. First, I find that discrete emotions are the best representation for emotional information embedded in online reviews for predicting sales as compared to valence (Hennig-Thurau et al., 2015; T. Song et al., 2019) and latent emotion topics (X. Yu et al., 2012). I also demonstrate that discrete emotions are substitutive for valence and latent emotion topics. With discrete emotion analysis as outperforming valence analysis, the prevalent practice in industry and academia, my work suggests that researchers and managers can use discrete emotions to advance the understanding of consumer emotions embedded in large-scale online user-generated content. My results further provide empirical evidence that supports discrete emotion theory (Lerner et al., 2015; Plutchik & Kellerman, 1982; Tomkins, 1962). It is also valuable to find that theory-driven discrete emotional features are more informative than are data-driven latent emotion topics.

Second, my work demonstrates the economic importance of discrete emotions in online reviews. I show that the effectiveness of discrete emotions is comparable to or higher than that of the existing factors in the literature, including valence, arousal (East et al., 2017; Junyong Kim & Gupta, 2012; Yin et al., 2017), review quality, credibility (Teng et al., 2017), and reviewer characteristics (A. H. Huang et al., 2015; Ngo-Ye & Sinha, 2014; Yin et al., 2016). Further, this chapter is among the first to show that the effectiveness of discrete emotions in online reviews is robust across different cultures.

Third, my work takes an initial but important step to understand the effects of discrete emotions from a new perspective that is overlooked in the information systems literature, i.e., the mediating effect of perceived processing fluency. Differing from the existing mechanisms, i.e., perceived writer cognitive effort (Yin et al., 2014) and perceived writer rationality (Xiao et al., 2018), which explain only the effects of anxiety and anger, the mediating effect of processing fluency is prevalent in all types of discrete emotions, except anger. Future research could involve more comprehensive investigations of this theoretically interesting and practically important topic. For example, it would be valuable to determine how to use empirical methods to predict perceived processing fluency in reviews as well as to determine whether processing fluency is associated with other factors beyond emotions in online reviews, e.g., a writer's review writing ability.

Fourth, through advancing existing work (Xue et al., 2014), my work proposes an approach that enables automatic domain-adaptive emotion lexicon construction and multi-dimensional emotion detection in texts, which adds to the literature on affective computing (Picard, 2003) and domain-driven data mining (Cao, 2010). In particular, the previous work by Xue et al. (2014) can be viewed as a special case of my Algorithm 1 when f is implemented by a K-nearest neighbor algorithm, without the iterative process, and when there are only two emotional dimensions (positive and negative). I contribute to their work as follows. First, I modify f by introducing another hyperparameter α to achieve better accuracy. Further, I show that using a KNN-type algorithm that uses the cosine similarity between word vectors achieves better performance than does directly training the state-of-the-art models on the 200-dimension word vectors. Second, I propose an iterative process that enables us to mine as many words as possible in the domain lexicon. my method detected 30.2% (1,556 out of 5,154) more emotion words than did theirs. More importantly, I show that test errors will remain consistently low as the number of iterations grows, which demonstrates the validity of the process. Third, I validate the newly mined words with out-of-sample testing and human annotation. Fourth, I propose a more general framework for lexicon extension with statistical language modeling. Finally, I construct and validate a new emotion lexicon specific to microblogging movie reviews for future research and application.

This work has several managerial implications. First, this paper highlights the economic importance of discrete emotions in online reviews. Online review platforms, such as Amazon, Yelp, and IMDB, currently ask users to rate only their positive/negative attitudes (i.e., valence) toward consumption experiments. My work calls for attention to design features that facilitate discrete emotional expressions. Two notable exceptions that consider such features are (i) Facebook.com, which had a major expansion of its emoji responses from "Like" to multiple emotions/emojis (e.g., angry, funny, love, like), and (ii) Rappler.com, a news website that, at the end of each piece of news, asks its audience to express their feelings by choosing one out of eight discrete emotions.

Second, the mediating effect of processing fluency can be leveraged to enable writers for helpful review writing. When the platform detects that a consumer is writing a positive review of low readability (e.g., with excessive length, with information presented in a complicated way), it can suggest that the consumer explicitly express his or her positive emotions to increase review processing fluency. When the platform detects that a consumer is writing a negative review, it can advise the consumer to write a review of high readability (and with more rational expression if the consumer's emotion is anger), instead of more negative and emotional expression. Further, the platforms may leverage my algorithm to add emotional tags to highlight positive discrete emotions in reviews, which can enhance the readers' fluency in processing review information.

Third, my results help movie marketers to develop online movie marketing strategies. My methodology enables marketers to detect online reviews with different discrete emotions. Thus, to boost box office sales, movie marketers could shift resources to design, share, reply, or control online reviews in regard to their embedded emotions (T. Song et al., 2019). In addition, because my methodology involves only public data, it can help marketers to predict the performance of other competitive movie products that are released in the same period and facilitate competitive responses.

Fourth, my approach helps cinema managers to predict box office sales more accurately, thereby helping them to optimally select movies, arrange screenings (T. Song et al., 2019; X. Yu et al., 2012), and make operational decisions of inventory management. Finally, my methodology can be easily adapted in other domains (e.g., books, music, restaurants, car sales) to detect emotions in consumer reviews and to predict future business performance.

I acknowledge some limitations of this work, which may motivate future research. First, I did not assess the recall of the emotion words in my lexicon because it is extremely difficult to obtain the ground truth of how many unique emotion words there are in 3.26 million microblogging messages. Second, in Study 2, I used the participants' self-reported data to measure purchase intention. To mitigate self-report biases, future research can offer participants the option of buying a movie ticket as a means to observe their actual purchase behaviors. Third, as online platforms, including Facebook, are adopting more pre-designed emotion-expression options for users, researchers can develop an experiment to evaluate the business value of such emotion-expression-related designs. Fourth, the mechanisms through which perceived processing fluency affects perceived review helpfulness can be further explored. Finally, I used a Word2Vec model to map words into vectors. Future research may investigate whether more sophisticated models, e.g., transformer-based models, can produce more accurate word vectors and thereby generate better emotion lexicons. Further, in Algorithm 2, I used a weighted average of the emotional intensities of the most similar words to predict the emotional intensities of the newly extended words. The weights are based on semantic similarity, which is reasonable but not necessarily optimal. The attention-based architecture may be leveraged to learn the optimal weights.

Chapter 4. UNDERSTANDING VIDEO ADVERTISEMENT EFFECTIVENESS USING DEEP LEARNING ALGORITHMS AND ECONOMETRIC METHODS

This chapter goes beyond text-mining algorithms and customer emotions. In the context of video advertising, this chapter applies a variety of computer vision, natural language processing, clustering, and visualization algorithms in understanding customer behaviors. In addition, this chapter shows that algorithmic output can be integrated into econometric methods to better understand behaviors.

In particular, this chapter presents a large-scale investigation to understand the effectiveness of outstream video ads in the context of online shopping. Specifically, the first research question is: *How effective are outstream video ads at driving CTR?* Although practitioners are turning to video ads, management researchers argue that the intrusive nature of planting ads on shopping pages may hurt CTR by inducing consumers' annoyance and avoidance (Todri et al., 2020). Thus, it is important to empirically answer the question with large-scale evidence.

Second, on online shopping websites, such as Amazon, one unique feature of outstream video ads is that they exist in competitive environments. Millions of e-commerce sellers are under fierce market competition every day, and they need to know whether video advertisement is an effective tool to help their products succeed in markets with high rivalry, in which products are of low differentiation from each other. I theorize that outstream video ads may help capture consumer attention in competitive markets (Cetin & Bingol, 2014), and propose the second research question: *How does the effectiveness of outstream video ads differ according to the extent of product differentiation?* To my knowledge, this work presents the first such study in the literature of video marketing and analytics (for a review, see Zhou et al. (2021)) to investigate this question.

Third, successful ad content provides informational, entertaining, and educational value to customers, whereas uninformative and uninteresting ad content may backfire (Li & Xie, 2020). Insight into what video content features increased advertisement effectiveness is of high managerial importance to advertisers. Despite this importance, management scholars have just begun to explore how to derive insights from large-scale video content with video analytics techniques, for example, from online classroom videos (Zhou et al., 2021). The previous literature on video content analysis has relied largely on manual coding (Goodrich et al., 2015; A. N. Smith et al., 2012; Teixeira et al., 2014), which is labor-intensive and prohibitively costly to scale. A lot is unknown regarding which are the important video content features of outstream ads. Therefore, I propose the third research question: *What video content features are associated with outstream video advertisement effectiveness?*

Collaborating with a leading e-commerce platform, I collect a unique click-stream data set of 200,000 randomly selected sessions. A session is defined as a continuous period of user activity in the browser, where successive events are separated by no more than 30 minutes. I collect the product queries contained in each of the sessions. For each product query, I collect data on all products listed on the shopping page, including products in organic search results and sponsored products with video ads. For each product, I collect data on its characteristics, video ad content (if any), and consumer clicking behaviors toward each of the products.

I leverage a deep representation learning approach, product embedding (Chen et al., 2020; Grbovic & Cheng, 2018; J. Wang et al., 2018), to measure product differentiation of a query. Then, I use reduced-form models-that is, Mahalanobis Distance matching, panel logistic regression, and panel linear probability regression with controlling for query-level fixed effects-to demonstrate the effectiveness of video ads on CTR, and its heterogeneity when the levels of product differentiation vary. Further, motivated by research on social media image analytics (Shin et al., 2020) and the elaboration likelihood model (ELM) (Petty & Cacioppo, 1986), I conduct video analytics to extract a variety of interpretable video ad features to learn the effect of video characteristics on video advertisement effectiveness. A two-stage logistic model is proposed to understand consumers' decisions regarding clicking, in which I first model consumers' latent attention toward a product, and then, conditional on latent attention, I model consumers' likelihood of clicking the focal product.

Several different model specifications consistently show a more than 500% increase in CTR for products advertised by outstream videos, compared to otherwise very similar products, which indicates high economic significance. The two-stage model shows that outstream video ads increase CTR first through effectively attracting consumer attention. Such effectiveness is higher when products are less differentiated in a market. Further, contingent on consumer attention, my model shows that two types of video content features lead to a higher likelihood of clicking. Features of the first type facilitate efficient consumer learning of a product, including short length, high content consistency, low object-level complexity, high ad-product relevance, and few words. Features of the second type signal high costs and effort invested in making the video content, which may further signal unobservable product quality (Kihlstrom & Riordan, 1984). These features are celebrity endorsement, the use of a professional video-making technique (the two shot, a filmmaking technique that encompasses two people in the frame), high visual aesthetics (including the rule of thirds and colorfulness), and high pixel-level complexity.

This chapter has important theoretical implications. To begin, this work is among the first to demonstrate the effectiveness of outstream video ads. This work contributes to the literature of video advertising (Campbell et al., 2017; Goodrich et al., 2015; Joa et al., 2018) by demonstrating that the effects of outstream video ad features are highly distinct from our knowledge of the effects of instream video ad features. Second, this work adds to the literature of competition, advertising, and consumer attention (Cetin & Bingol, 2014; Nelson, 1970; Telser, 1964) by showing that advertising effectiveness on consumer attention is negatively associated with product differentiation in a market, in the context of outstream video advertising. To overcome empirical challenges and achieve this contribution, I collect a large-scale, cross-market, and click-stream level dataset, leverage deep representation learning techniques to measure product differentiation levels, and estimate unobservable consumer attention. Third, my work contributes to the literature on video analytics in marketing (Zhou et al., 2021) by illustrating a systematic, interpretable, automatic, and scalable video ad content analytics framework.

My work also provides significant managerial implications for practitioners. First, for outstream video ad makers, I show a variety of video content features that lead to a higher CTR. More generally, whereas instream video ads aim to provide entertaining and affective value to viewers (Campbell et al., 2017; Goodrich et al., 2015; Joa et al., 2018), directors should make outstream video ads that aim to efficiently convey key product information to consumers and signal product quality. Second, my work helps sellers on e-commerce platforms to form advertising strategies. Although outstream video ads are effective in all markets, their effectiveness is higher in markets where product differentiation is lower (more competitive markets). Sellers in these markets can consider allocating more resources to bid on video ad placement. Third, the video ad feature framework enables e-commerce platforms to identify a variety of important features that are associated with the CTR of ads and help platforms to refine their advertising recommendation systems. Finally, consumer attention to a video ad is unobservable but of high managerial relevance, for example, for brand awareness. my two-stage model helps the platforms to infer latent consumer attention to video ads and provide more dimensions of advertising feedback to advertisers.

4.1 CONCEPTUAL FRAMEWORK

I present my conceptual framework in Figure 4.1. First, based on the literature of the two-stage model of advertising, outstream video ads may affect both consumer attention and clicking. Further, attention is a pre-condition of clicking. Second, based on the literature on advertisement, competition, and limited attention, I propose to investigate how the level of product differentiation would shape the effectiveness of video advertising in attracting consumer attention. Third, after the consumer pays attention to a product with a video ad, I examine what video ad content features would moderate video ad effectiveness on consumer clicking. I note that the content features can take effect only in the second stage, because if a consumer pays no attention to the content, no content features will affect their clicking. Based on ELM and video content analytics, I construct video content features and divide them into features in the peripheral route and central route, respectively.

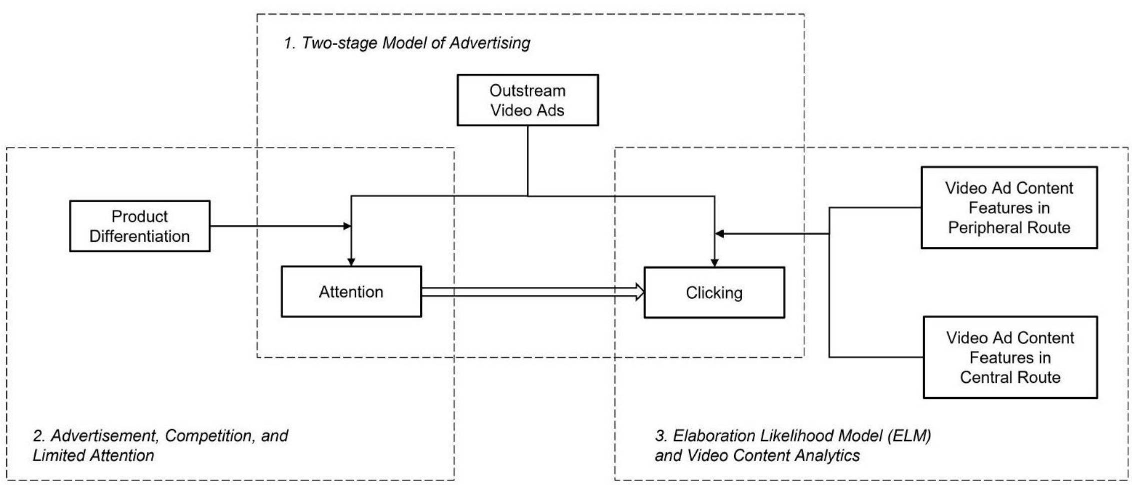


Figure 4.1. Conceptual Framework

4.2 CONTEXT AND DATA

In this new form of digital advertising, when consumers search products with a query and as they scroll down the query result webpage, they may encounter a sponsored product presented with a video ad (See an example shown in Figure 4.2). The video ad will auto-play when it appeals on the user's screen. I confirmed with the platform that, due to the randomness and complication of its advertising recommendation systems, it is very difficult from the consumers' perspective to predict whether and where they would encounter a video ad; thus, the setting is quasi-experimental.

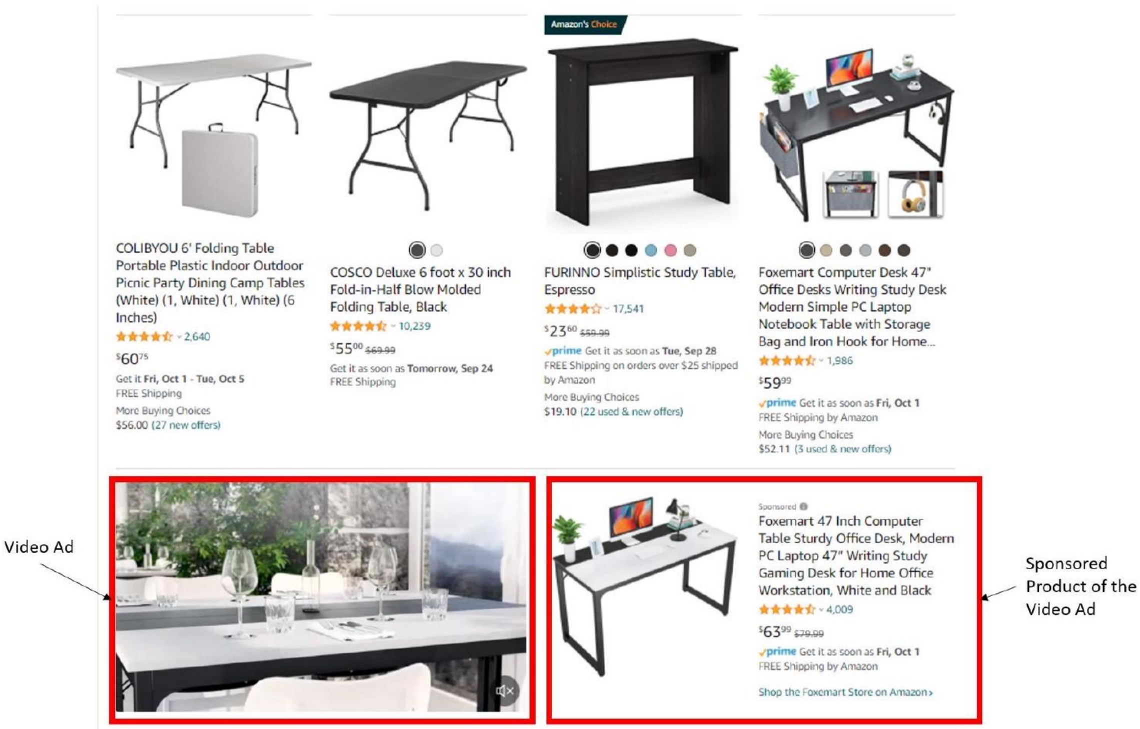


Figure 4.2. An Example of Video Ads on Shopping Websites

I collected a dataset of 200,000 sessions. It is a random sample of all sessions starting at 10:00 am Pacific Time (PT), June 13, 2021. One session can contain multiple queries. In total, I have 278,272 queries contained in the 200,000 sessions. For each query, I choose to collect the first page of its products because, according to the platform, only a very low proportion of consumers would navigate to the second page or further pages. For each product listed in the query, I collect data on product information (name, price, tags[22], volume and ratings of reviews, position on the shopping page), video ad information (whether associated with a video ad, and if so, whether the video gets impressed, and the video content), and user behaviors (i.e., whether the consumer clicked the product). Video impression is defined as the event in which over 50% of the pixels of the video ad are shown on the user's screen for at least 2 seconds, as a standard measure used by the platform. To simplify the dataset, I choose not to include audio data, because, in my context, ads play automatically on mute when they appear on the screen and very few consumers turn on the audio of a video ad. Notably, to understand how advertising effectiveness is affected by market competition, one important empirical challenge is to obtain large-scale, cross-market data by which variation in competitive levels can be utilized for statistical identification. This dataset enables me to overcome this challenge.

In what follows, I introduce the variables that I construct from my sampled data. In particular, I first measure product differentiation in a query with the product embedding approach. Second, I construct video features of interest based on video content analytics. Third, I construct important control variables.

4.2.1 *Product Differentiation*

Empirically measuring product differentiation for a large number of markets is challenging. Thanks to the recent development of representation learning, marketing researchers have started studying competition with product embedding vectors (Chen et al., 2020). Product embedding vectors are generated by deep learning models in computer science literature (Biswas et al., 2017). They are high-dimensional numerical vectors that preserve important product information. Notably, product embedding is the state-of-the-art technology that has been widely adopted in e-commerce systems, such as search ranking systems of Airbnb (Grbovic & Cheng, 2018) and recommendation systems of Alibaba (J. Wang et al., 2018).

A multi-modal deep neural network model, extensively pre-trained based on billions of product samples and validated by applied scientists and annotators at Amazon (C. Wang et al., 2019), produces a 144-dimension numerical vector for each of the products in my sample. The high-dimensional embedding vectors summarize the latent characteristics revealed by all the

[22] Tags are used by the platform to provide additional information for a product, for example, whether this product is a sponsored product. The concrete information of product tags is not disclosed due to confidentiality.

available textual and image information of their corresponding product, which includes product images, product name, brand information, product description, and usage instruction.

With the embedding vectors, I measure product differentiation within a query. I leverage the idea of "within-cluster distance" from the K-means clustering algorithm-the average distance from observations to the cluster centroid, as a measure of the variability of the observations within the group. Similarly, a query with a smaller average distance has less differentiated products than does a query with a larger average distance. Specifically, for product t in query j, I denote its embedding vector as $v_{jt} \in \mathcal{R}^k$, where $k = 144$ and $\|v_{jt}\| = 1$, where $\|\cdot\|$ is the l^2-norm. Then, the within-cluster distance (as a measure of product differentiation) is calculated as $\text{ProdDif}_j = \frac{1}{n_j}\sum_{t=1}^{n_j} D(v_{jt}, \bar{v}_j)$, where $\bar{v}_j = \frac{1}{n_j}\sum_{t=1}^{n_j} v_{jt}$ is the cluster centroid, and n_j is the total number of products in query j. Function D returns the euclidean distance between two embedding vectors in the space of $\mathcal{R}^k$.

4.2.2 *Video Impression and Features*

I define a dummy variable $WithVideoAd_{jt}$ to indicate whether product t in query j is presented with a video ad. I note that the following video variables are defined only when $WithVideoAd_{jt} = 1$. First, $VideoImpression_{jt}$ is a dummy variable indicating video ad impressions (i.e., whether 50% of the pixels of the video ad are shown on the user's screen for at least 2 seconds). Then, based on the ELM and visual content features in the existing literature, I construct video features in the central route and the peripheral route, respectively.

To begin, I construct the features in the central route, that is, the features that help consumers to evaluate the pros and cons of the advertised product. First, I focus on video length, measured by the number of minutes of a video ad, denoted as $VideoLength_{jt}$. As a basic property of video ads, length is argued to be an important factor regarding video advertising effectiveness (Zhou et al., 2021). Early research by lab experiments found that longer instream (pre-roll) video ads are more effective because they are perceived by consumers as less intrusive and increase ad recall (Goodrich et al., 2015). Industrial experience[23], in contrast, argues that shorter social media video ads are more effective. Short length may imply that the director has a more concise and compelling way of storytelling, making it easier for consumers to understand

[23] https://business.linkedin.com/marketing-solutions/native-advertising/video-ads/video-ads-best-practices

the merits of the advertised product, whereas long video ads can create information overload and increase the difficulty of evaluating the advertised product. Thus, it is interesting to understand with a large-scale field test whether shorter or longer outstream video ads are more effective.

Second, I define the feature of continuity editing, which relates to content consistency. Content consistency is defined as the consistency of specific content with respect to previous content in an ad. It is an important factor studied in the literature of email marketing (Fong, 2017) and social media image advertising (Shin et al., 2020), which can help a consumer learn new products. It can either positively or negatively affect the process of consumer product evaluation, depending on whether the consumer is variety- or consistency-seeking (Fong, 2017; Shin et al., 2020). As for video ads, content consistency is the consistency of content of a specific frame with respect to its previous frames. A related technique in film making and video creation is "continuity editing" (Bordwell et al., 1993). As part of the "classical-Hollywood-style" developed by early European and American directors, continuity editing is about keeping the consistency of on-screen elements over the course of a scene.[24] Because in the management literature content consistency is currently studied only in texts and images and because both positive and negative effects of content consistency are found (Fong, 2017; Juho Kim et al., 2014; Shin et al., 2020), it is interesting and important to investigate whether continuity editing in videos is of business value regarding driving advertising effectiveness. To measure continuity editing, I use Amazon Rekognition, a video analytics tool developed by Amazon[25], to detect scene cuts. Scene cuts are induced by video transitions. High-frequency transitions (i.e., low content consistency) may increase viewer engagement (Zhou et al., 2021), but they may not give viewers enough time to comprehend what is presented in the ad (Juho Kim et al., 2014). I use the time between scene cuts (calculated in the log-transformed average number of seconds, denoted as $LogSecPerScene_{jt}$) to measure the degree of continuity editing, following the existing study on online classroom videos (Zhou et al., 2021).

Third, object-level complexity measures the diversity of objects detected in visual content and has a positive impact on consumer engagement toward social media image posts (Shin et al., 2020). Therefore, I expect that object-level complexity has an impact on product evaluation and shapes the effectiveness of video ads. But because the ways that images and videos show objects

[24] https://en.wikipedia.org/wiki/Continuity_editing
[25] https://aws.amazon.com/rekognition/

are fundamentally different (static versus dynamic), I do not know ex-ante whether the diversity of objects would positively affect video ad effectiveness. To measure object-level complexity, I extend the existing measure in image studies (Shin et al., 2020), dynamically detect objects in the video ads based on computer vision and deep learning techniques, and calculate the Shannon diversity index of these objects, denoted as $ObjCmp_{jt}$. The value of the measure is greater when a video ad, on average, contains a higher number of unique objects during every 5 seconds (See Appendix B.1 for technical details).

Fourth, I measure the semantic relevance of the ad content and the product information, also referred to as product relevance (Li & Xie, 2020), denoted as $ProdRel_{jt}$. Existing advertising research shows that ad-product relevance (also known as congruence) moderates the effectiveness of image ads (Miniard et al., 1991). Irrelevant content in ads creates extra difficulty for readers to comprehend the product information and therefore leads to low ad recall and less favorable consumer attitudes, as shown in studies on print ads (Heckler & Childers, 1992) and social media images (Li & Xie, 2020). I am interested in whether this can be generalized to video ads, in which much richer information is presented to consumers compared to image ads and print ads. One might argue that video ads with low product relevance could offer more diverse product information or other entertainment values to consumers. For example, occasionally, customers see some video ads in which the product remained unclear (until revealed at the end of the ads). To formally investigate the question, I measure the ad-product relevance of a video ad, leveraging Amazon Rekognition's object and text detection techniques and the popular neural network language model, Word2Vec (Mikolov et al., 2013) (See Appendix B.2 for technical details). The higher the value of $ProdRel_{jt}$ (ranging from -1.0 to 1.0), the more semantically relevant the video ad and the product.

Fifth, I analyze the amount of textual information embedded in video content. An industrial report on Facebook ads[26] argues that when image ads are text-heavy, the browsing experience can become noisy and distracting. I am therefore motivated to investigate whether text-heavy video ads would also lose their effectiveness to some extent. Amazon Rekognition is used to capture words and their time stamps in the video. Then, I use the log-transformed average number of words per second to measure the volume of text in the video, denoted as $LogWordsPerSec_{jt}$.

[26] https://repurposehouse.com/hack-text-heavy-videos-for-facebook-ads/

Next, I construct the features in the peripheral route, in other words, features that do not relate to the pros and cons of a product. First, with human and face detection techniques provided by Amazon Rekognition, I measure social presence and celebrity endorsement in a video ad. I use $HasPerson_{jt}$ and $HasCelebrity_{jt}$ as dummy variables to indicate whether humans are detected and whether celebrities are detected in the video ad, respectively.

Second, I detect the presence of Two Shot (denoted as dummy variable $TwoShot_{jt}$). In cinematography, a two shot is an important technique of encompassing two people (the subjects) in the frame (Thompson and Bowen 2009). Two shots can be used for showcasing romantic relationships, highlighting intimate actions, comparing and contrasting characters, and revealing information.[27] A two shot is argued to be one of the most effective techniques for successful video ad making, according to an industrial report from Linkedin.[28] I use Amazon Rekognition to detect person paths in video ads, that is, capturing the position and time of each person who is moving in the video. I measure whether there exists at least one frame that contains two moving people in the video.[29]

Third, visual aesthetics is a representative peripheral route visual feature that increases consumer engagement toward social media content (Shin et al., 2020) and in online classrooms (Zhou et al., 2021) and increases Airbnb property demand (Zhang et al., 2021). I first measure video aesthetics regarding composition, the way in which objects in the video are arranged. Among the most representative aesthetic composition techniques is the rule of thirds (ROT) (Amirshahi et al., 2014). An image can be divided into nine equal parts with (imaginary) horizontal and vertical third lines. The ROT states that the main visual elements should be placed close to the four intersections of the imaginary lines (Krages, 2012). Amazon Rekognition detects objects in the videos and returns a bounding box of each object. To measure ROT, I calculate the euclidean distance between the centroid of each bound box and each of the four imaginary intersections. Then, I take the minimum of the four distances as ROT distance. For all objects in the video, I take the average of their ROT distance, denoted as $ROTdis_{jt}$. Videos with smaller values of $ROTdis_{jt}$follow ROT to a larger extent. Second, I measure video aesthetics

[27] https://www.nfi.edu/two-shot/

[28] https://business.linkedin.com/marketing-solutions/native-advertising/video-ads/video-ads-best-practices

[29] In my sample, some videos are made by presenting static pictures one-by-one, similar to playing slides. Sometimes, two persons may appear in a static picture. Such a case does not fit the definition of the two shot in cinematography, because the two people do not have any dynamic interaction. Therefore, I code only a frame with two moving persons as a two shot.

regarding colorfulness, or color complexity, which is constructed by the entropy of color distribution (Shin et al., 2020), denoted as $ColorCmp_{jt}$. The larger the variable, the more colorful the video (See Appendix B.3 for technical details).

Finally, I detect pixel-level complexity of the video content, that is, the variation of video frames at the pixel level (Pieters et al., 2010), denoted as $PixelCmp_{jt}$. As a commonly accepted peripheral route visual cue, pixel-level complexity can evoke low-level visual processes and affect visual advertising effectiveness (Pieters et al., 2010). Pixel-level complexity can increase a viewer's physiological arousal (Shin et al., 2020). Arousal can further positively affect consumer engagement (Berger and Milkman 2012). I follow the validated method to measure pixel-level complexity, in other words, the compressed file size of an image, which represents the minimal computer storage required to store the image (Pieters et al., 2010; Shin et al., 2020). If the variation at the pixel level increases, the minimal storage size needed to store the image becomes greater. For a certain video, I average the compressed file size of its frames as a pixel-level complexity measure of the video.

I show the result of correlation analysis on the video features in Appendix B.4. I find that the maximum value of correlation is 0.51 between $PixelCmp_{jt}$ and $ColorCmp_{jt}$, because the color is part of the information stored in pixels. The second-largest correlation is 0.41 between $LogWordsPerSec_{jt}$ and $ProdRel_{jt}$, possibly because text-heavy video ads are likely to be designed for introducing the products in detail, which leads to high product relevance. Other correlations are consistently lower than 0.3. In summary, the results show that the video features are relatively orthogonal to each other.

4.2.3 *Other Control Variables*

In this section, I introduce important control variables. First, as the high-dimensional product embedding vectors summarize product characteristics, they can be used to control product idiosyncratic effect in the econometric analysis. But the high-dimensional nature of the vectors can cause difficulty in statistical inference. I thus use the principal component analysis (PCA) model, a commonly used dimension reduction algorithm, to reduce the dimension of the product embedding vectors from 144 to 16 (Minka, 2000). The 16 variables are denoted as $ProdEmb_{jt}^{\kappa}$, where $\kappa \in \{1,2,\dots,16\}$. In Appendix B.5, I illustrate the validity of the 16 -dimension vectors with additional analysis, following a common practice of the literature (Grbovic & Cheng, 2018; J. Wang et al., 2018). Second, for each product, I define $Position_{jt}$, a positive integer variable that indicates the position of product t in query j. The product listed first in the retrieved results of the search query is coded as 1, and the second as 2, and so forth. This controls for position bias of consumer behaviors (Blunch, 1984), meaning that consumers are more likely to pay attention to products in the top positions than those in the lower positions regardless of product quality and relevance. Third, I control for logarithm-transformed product price ($LogPrice_{jt}$), logarithm transformed number of customer reviews ($LogReviews_{jt}$), and ratings ($Ratings_{jt}$) for product t in query j, as they are important factors in the consumer decision process regarding clicking. Fourth, I control product tags that may affect the CTR. Finally, based on the search-experience-credence (SEC) good classification (Ekelund et al., 1995), I control the product type of a query. I include two dummy variables $ExpGoods_j$ and $SearchGoods_j$ to indicate whether query j contains experience, search, or credence goods. In this context, one query can contain only one type of good.

4.3 REDUCED-FORM EVIDENCE

In this section, I provide reduced-form evidence to show the effectiveness of outstream video ads and how the effectiveness changes when product differentiation varies. Before specifying the reduced-form models, I use Mahalanobis Distance Matching (MDM) as a pre-processing step to filter samples at the query level. Queries with and without video ads could be potentially different, and matching is an effective way to adjust for selection bias (Rosenbaum & Rubin, 1983).

4.3.1 *Query-level Mahalanobis Distance Matching*

I use the following covariates for query-level matching. First, I match for the log-transformed average price, the log-transformed average number of customer reviews, and average ratings of the products in the query, which may potentially correlate with a marketer's decision about bidding for a video ad. Second, I match for product differentiation variable ProdDif for query j, as marketers in a more competitive market could be more likely to leverage new forms of advertising to attract attention. Third, I match for the product category of the query, in other words, whether it contains search, experience, or credence goods, as marketers' decisions about different categories could be potentially heterogeneous. Fourth, I match for the device that is used to submit the query, that is, at the mobile versus PC end, because the presentation of video ads at the mobile end is slightly different from that at the PC end. Further, consumers at different ends can be heterogeneous.

With the Mahalanobis distance, I follow a one-to-one matching procedure to match each query containing a video ad with the query that is of the shortest distance to the focal query and is without a video ad (as detailed in Appendix B.6). After matching, I obtained 23,996 queries in total (half of them contain a video ad). The total number of products contained in these queries is 605,271. Table B.3 of Appendix B.6 shows that MDM eliminates query-level imbalance after matching (with all percentage bias less than 0.23%, derived by the percentage difference of variables of the queries without a video ad and those of the queries with a video ad).

4.3.2 *Fixed Effects Model Estimation*

Although I balance for cross-query covariates, query-level fixed effects and product characteristics within the query may also confound the effect of video ads on CTR. A conditional logistic regression (CLR, also known as fixed effects logistic regression) model is specified to identify the effect of video ads on CTR. Specifically, for query j and all the products $t \in \{1,2,\dots,T_j\}$ listed in the query j, the dummy variable y_{jt} indicates whether product t is clicked by the consumer, as a random realization of a Bernoulli distribution with probability p_{jt} :

$$y_{jt} \sim \text{Bernoulli}\left(p_{jt}\right), \log \frac{p_{jt}}{1-p_{jt}} = \alpha_j + X_{jt}\beta$$

where α_j captures the query-level (unobservable) fixed effects, including individual fixed effects of the user who submits the query, and query-specific effects (e.g., product category, average price, average product ratings, level of product differentiation, etc.). Further, as all queries in my samples happened at a fixed time point, time fixed effects are also absorbed in α_j.

X_{jt} is a vector that includes product-specific variables. To begin with, I include $WithVideoAd_{jt}$ to indicate whether product t in query j is presented with a video ad. Then, I include my variable of interest, $VideoImpression_{jt}$, indicating video ad impression. Further, I include the interaction term $VideoImpression_{jt} \times ProdDif_j$ to measure the effect of product differentiation on video ad effectiveness.

Next, I include product-specific control variables. First, I include $Position_{jt}$, the position of product t in query j, to control for position bias of consumer behaviors. Second, I control for logarithm-transformed product price ($LogPrice_{jt}$), logarithm-transformed number of customer reviews ($LogReview_{jt}$), and ratings ($Ratings_{jt}$) for product t in query j, as they are important factors in the consumer decision process regarding clicking. Third, I control for product tags. Finally, I include 16 product embedding variables to control for latent product characteristics. A conditional maximum likelihood estimator (MLE) is used to estimate β (Hosmer Jr et al., 2013).

Although logistic regression models are commonly used for dummy outcome variables, I also specify a linear probability model with fixed effects (LPMFE) for the following reasons. First, the parameter estimates of LPMFE can be directly interpreted as the "mean marginal effect" of covariates on the CTR, whereas parameter estimates of CLR are interpreted as the effects on the log-odds ratio of clicking, which are less intuitive than marginal effects. Second, when the rate of the outcome-CTR in my context-is low, LPMFE may achieve estimates with lower bias and more accurate standard errors compared to the logistic regression model according to Monte Carlo simulations (Deke, 2014). Third, LPMFE can perform as a robustness check for the estimates of CLR. Formally, for query j and all the products $t \in \{1,2,\dots,T_j\}$ listed in the query j, the probability of t being clicked is denoted as p_{jt} :

$$p_{jt} = \alpha_j + X_{jt}\beta + \epsilon_{jt},\ E(\epsilon_{jt} \mid X_{jt}) = 0$$

Finally, I also run a linear probability model with random effects (LPMRE) to check the robustness of the model without query-level fixed effects term α_j. Notably, in LPMFE and LPMRE, for confidentiality, the coefficients are reported as the lift of CTR. Namely, the ratio of increase or decrease of CTR when there is a one-unit change in an independent variable compared to a baseline CTR. Specifically, lift of CTR is calculated as (new CTR-baseline CTR)/baseline CTR. The baseline CTR is derived by the model with dummy variables taking the value as zero and continuous variables taking their average values.

Table 4.1. Reduced-Form Estimation Results

	CLR Log-odds of Clicking		LPMFE Lift of CTR		LPMRE Lift of CTR	
	Parameter	Std. Err.	Parameter	Std. Err.	Parameter	Std. Err.
$VideoImpression_{jt}$	2.685***	0.322	5.276***	0.450	5.445***	0.383
$VideoImpression_{jt} \times ProdDif_j$	−1.265**	0.452	−4.935***	0.804	−4.969***	0.679
$WithVideoAd_{jt}$	−1.271***	0.230	−0.951***	0.162	−1.055***	0.149
$Position_{jt}$	−0.064***	0.002	−0.049***	0.001	−0.039***	0.001
$LogPrice_{jt}$	−0.111***	0.020	−0.114***	0.024	−0.109***	0.016
$LogReviews_{jt}$	0.019**	0.006	0.033***	0.008	0.008	0.005
$Ratings_{jt}$	−0.040***	0.010	−0.049***	0.013	−0.039***	0.010
$ProdEmb_{jt}$	Yes		Yes		Yes	
Product Tags	Yes		Yes		Yes	
Query-level Fixed Effects	Yes		Yes		No	
# groups	8,228		23,996		23,996	
# obs.	242,856		605,271		605,271	
Log-Likelihood	-29,243		NA		NA	
R-square	NA		0.009		0.009	

Note:* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. NA = Not Applicable. Constant terms are not disclosed due to confidentiality.

Table 4.1 shows the estimation results of the three reduced-form models. First, the models consistently show that outstream video ads are significantly effective in driving CTR. In particular, based on CLR, $VideoImpression_{jt}$ increases the log-odds ratio of clicking by $2.685\ (p < 0.001)$. Based on LPMFE and LPMRE, $VideoImpression_{jt}$ increases CTR by $527.6\%\ (p < 0.001)$ and $544.5\%\ (p < 0.001)$, respectively, demonstrating economic significance. Second, their effectiveness is higher when products in a query are less differentiated, as the coefficients of the interaction term $VideoImpression_{jt} \times ProdDif_j$ are consistently and significantly negative. The coefficients (from -493.5% to -496.9%) are smaller than those of $VideoImpression_{jt}$, indicating that even in a market with high product differentiation ($ProdDif_j$ is close to its maximum value, which is 1), video ads would still have a positive effect. In addition, the similar results of LPMFE and LPMRE indicate that query-level fixed effects do not confound the estimation of video ad effectiveness, implying that Mahalanobis Distance Matching has corrected for potential selection issues (if any) at the query level.

4.4 Model Analysis

The reduced-form models, however, do not consider the two-stage nature of customer choice, namely, attention and clicking. Variables such as video impression, product position, and tags are linked to the attention stage, whereas other variables such as video characteristics and product reviews can only take effect in the second stage provided that the consumer has paid attention to the product. Thus, the model could be misspecified if I directly link all variables to clicking behaviors.

Consumer attention to a product is unobservable to the advertisers and researchers. A two-stage logistic model is proposed to estimate the latent attention level to a product before estimating the consumer's likelihood of clicking. Notably, the aforementioned query-level matching procedure is conducted before estimating the two-stage logistic model. Formally, I use a latent dummy variable A_{jt} to indicate whether a consumer pays attention to product t in query j, which follows a Bernoulli distribution with a mean of p_{jt}^A. Further, based on the assumption of logistic regression, the log-odds ratio of p_{jt}^A is linked to a query-specific effect term α_j^A and a vector Z_{jt} containing attention-related variables:

$$A_{jt} \sim \text{Bernoulli}\left(p_{jt}^A\right),\ \log \frac{p_{jt}^A}{1 - p_{jt}^A} = \alpha_j^A + Z_{jt}\beta^A.$$

In Z_{jt}, I first include $VideoImpression_{jt}$, as products with video ads occupy high-visibility placements and have a higher probability of attracting attention. Notably, one cannot guarantee that consumers pay attention to the sponsored product even if $VideoImpression_{jt} = 1$ (i.e., over 50% of the pixels of the video ad are shown on the user's screen for at least 2 seconds). I only know that the video ad occupies a fraction of the screen for a short time period. It is possible that a consumer ignores the video ad due to ad avoidance tendency but pays attention to other products. Indeed, researchers have found the effect of “banner blindness,” which means that even if a banner is shown on the screen, a consumer may completely ignore it (Benway, 1998). Similarly, I cannot rule out the possibility of "video ad blindness" ex-ante. Therefore, in my model, I link video impressions to consumer attention in a probabilistic manner. Second, I include $VideoImpression_{jt} \times ProdDif_j$, because when products are less differentiated in a query, a video ad may make the corresponding product more unique and more easily recognized. Third, I include $Position_{jt}$, because the position bias theory argues that more attention is paid in the top positions, and it can decay as consumers scroll down the shopping page (Blunch, 1984). Fourth, I add product tags to control their potential effects on attracting attention.

Next, I assume that at the clicking stage, the dummy variable of clicking $\left(y_{jt}\right)$ is a random realization of a Bernoulli distribution with a mean of p_{jt}^C, conditional on $A_{jt} = 1$. Again, I assume that the log-odds ratio of p_{jt}^C is linked to a query-specific effect term α_j^C and a vector X_{jt} containing variables associated with the likelihood of clicking:

$$y_{jt}\big|_{A_{jt}=1} \sim \text{Bernoulli}\left(p_{jt}^C\right),\ \log\frac{p_{jt}^C}{1-p_{jt}^C} = \alpha_j^C + X_{jt}\beta^C.$$

Intuitively, $\alpha_j^C + X_{jt}\beta^C$ captures the consumer's latent interest levels toward product t in query j. Further, if a product does not receive attention, it will not receive clicking with probability 1; in other words, $\mathcal{P}\left(y_{jt} = 0 \mid A_{jt} = 0\right) = 1$.

In X_{jt}, I first include $VideoImpression_{jt}$ and its interaction terms with video content features, because if a product is without a video ad, or its video ad is not displayed on the consumer's screen, video content features will not have any effect on clicking. Notably, $HasCelebrity_{jt}$ and $TwoShot_{jt}$ are meaningful only if there are persons in the video content, that is, $HasPerson_{jt} = 1$. Therefore, instead of directly interacting them with $VideoImpression_{jt}$, I interact them with $VideoImpression_{jt} \times HasPerson_{jt}$. Second, I include $LogPrice_{jt}$, $LogReviews_{jt}$, $Ratings_{jt}$, and the vector of $ProdEmb_{jt}$. Price, number of reviews, ratings, and product-specific characteristics (product description, brand information, etc.) are important factors that a consumer uses to evaluate a product. Evaluation happens after attention because attention is about allocating cognitive resources to a specific product, whereas evaluation is about leveraging the allocated cognitive resources to judge whether the price is affordable, whether the rating and online review information support the purchase decision, and whether the product characteristics meet the consumer's preference.

The query-specific effect terms $\alpha_j^Q (Q \in \{A, C\})$ cannot be canceled out by the conditional maximum likelihood estimator that I used in CLR, because such an estimator applies only to ordinary CLR, not to general logistic-style regressions. However, as I have shown that query fixed effects estimation and random effects estimation yield consistent results, I model query-specific effects as random effects. Specifically, I provide structures for query-specific effects with query-level covariates M_j and random errors ϵ_j^Q. In M_j, I include two dummy variables *ExpGoods* $_j$ and *SearchGoods* $_j$ to indicate whether query j contains experience goods or search goods, controlling for the heterogeneity of product categories. I also include product differentiation measure *ProdDif* to control levels of competition. Formally, $\alpha_j^Q = M_j\gamma^Q + \epsilon_j^Q$, where ϵ_j^Q are assumed to follow normal distributions $\mathcal{N}(0, \sigma_Q^2)$, with σ_Q as parameters to be estimated. Intuitively, the parameters σ_Q account for the unobservable query heterogeneity. I separately model α_j^A and α_j^C to allow the different query-specific effects at the attention and the clicking stages. For computational feasibility and mathematical trackability, I assume ϵ_j^A and ϵ_j^C to be independent. I use maximum likelihood estimation for parameters of interest $\theta = (\beta^Q, \gamma^Q, \sigma_Q), Q \in \{A, C\}$ (as detailed in Appendix B.7).

4.5 EMPIRICAL RESULTS AND FINDINGS

Table 4.2. Model Estimation Result (DV: y_{jt})

	(1) Without Video Features		(2) Full Model	
Variables	Parameter	Std. Err.	Parameter	Std. Err.
Attention Stage (Z_{jt})				
VideoImpression $_{jt}$	2.541^{***}	0.028	2.624^{***}	0.098
VideoImpression $_{jt}$ × ProdDif $_j$	-0.712^{***}	0.247	-0.748^{***}	0.276
WithVideoAd $_{jt}$	-1.331^{***}	0.100	-1.331^{***}	0.127
Position $_{jt}$	-0.061^{***}	0.002	-0.061^{***}	0.001
Product Tags	Yes		Yes	
Attention Stage (M_j)				
ProdDif$_j$	-0.988^{***}	0.102	-1.022^{***}	0.124
ExpGoods $_j$	-0.142	0.092	-0.141	0.087
SearchGoods $_j$	-0.068	0.070	-0.064	0.067
σ_j^A	0.149	0.128	0.149	0.174
Clicking Stage (X_{jt})				
VideoImpression $_{jt}$	0.281	0.163	2.126^{***}	0.094
VideoImpression $_{jt}$ × VideoLength $_{jt}$			-2.380^{***}	0.187
VideoImpression $_{jt}$ × LogSecPerScene $_{jt}$			0.303^{**}	0.119
VideoImpression $_{jt}$ × ObjCmp$_{jt}$			-2.084^{***}	0.106
VideoImpression $_{jt}$ × ProdRel $_{jt}$			0.999^{***}	0.367
VideoImpression $_{jt}$ × Log WordsPerSec $_{jt}$			-0.609^{**}	0.251
VideoImpression $_{jt}$ × HasPerson $_{jt}$			-0.809^{***}	0.268
VideoImpression $_{jt}$ × HasPerson $_{jt}$ × HasCelebrity $_{jt}$			1.212^{***}	0.261

VideoImpression $_{jt}$ × HasPerson $_{jt}$ × TwoShot $_{jt}$			0.433^{*}	0.211
VideoImpression $_{jt}$ × ROTdis $_{jt}$			-1.790^{***}	0.329
VideoImpression $_{jt}$ × ColorCmp p_{jt}			1.977^{***}	0.363
VideoImpression $_{jt}$ × PixelCmp $_{jt}$			0.733^{***}	0.146
LogPrice $_{jt}$	-0.435^{***}	0.040	-0.432^{***}	0.041
LogReviews$_{jt}$	-0.078^{***}	0.024	-0.077^{***}	0.024
Ratings $_{jt}$	-0.674^{***}	0.068	-0.667^{***}	0.074
ProdEmb $_{jt}$	Yes		Yes	
Clicking Stage (M_j)				
ProdDif $_j$	1.115^{***}	0.366	1.331 $***$	0.273
ExpGoods $_j$	-0.014	0.179	-0.053	0.158
SearchGoods $_j$	-0.608^{***}	0.201	-0.666^{*}	0.293
σ_j^C	0.021	0.340	0.031	0.225
# groups	23,996		23,996	
# obs.	605,271		605,271	
Log-Likelihood	$-48{,}444$		$-48{,}436$	

Table 4.2 summarizes the estimation results of the two-stage model. In Column (1), I run a model without video features, and then in Column (2), I run a full model. The estimation results of Columns (1) and (2) are consistent across the two models. Therefore, I mainly discuss the results of Column (2).

At the attention stage, first, Column (2) shows that video ad impression ($VideoImpression_{jt}$) significantly attracts consumer attention ($\beta = 2.624, p < 0.001$) . Outstream video ads and their sponsored products are listed along with many similar products on the shopping website. Successfully attracting consumer attention is a pre-condition for consumer clicking. For instream video ads, in contrast, because they play before a viewer's intended video content, the viewer's attention has already been paid to the video content and the ads (Campbell et al., 2017). Further, differing from banner ads, which consumers tend to avoid (Benway, 1998; Todri et al., 2020), my model suggests that consumers are more likely to pay attention to outstream video ads, highlighting the difference between the two types of ads.

Second, the coefficient of the interaction term of video ad impression and product differentiation ($VideoImpression_{jt} \times ProdDif_j$) is statistically negative ($\beta = -0.748, p <$

0.001), which suggests that the effect of video ad impression on consumer attention is stronger when products are less differentiated in a query. The result is consistent with the analytical proposition by Cetin and Bingol (2014). On a webpage filled with almost identical products, a video ad gives the advertised product a significantly different appearance than other products, and therefore more easily captures a consumer's attention.

For control variables at the attention stage, first, the coefficient of $Position_{jt}$ is significantly negative ($\beta = -0.061, p < 0.001$), which indicates that the products listed higher on the page attract more consumer attention, consistent with the position bias (Blunch, 1984). Second, the coefficient of $ProdDif_{j}$ is significantly negative ($\beta = -1.022, p < 0.001$), which indicates that when products are less differentiated, the likelihood that a consumer pays attention to an average product is higher-possibly because when faced with a large number of undifferentiated products, a consumer needs to examine each product to evaluate their differences. Third, the coefficients of *ExpGoods* and *SearchGoods* are not significant ($p > 0.05$), indicating that there is no significant heterogeneity across product categories at the attention stage. Finally, σ_j^A is not significantly different from zero ($p > 0.05$). This suggests that there is no significant unobservable query heterogeneity at the attention stage.

At the clicking stage, first, without modeling the effects of video features, the coefficient of $VideoImpression_{jt}$ is positive but not significant ($\beta = 0.281, p > 0.05$) in Column (1). As I incorporate video content features in Column (2), the coefficient of $VideoImpression_{jt}$ becomes significantly positive. It indicates that although video ads occupy high-visibility placements and significantly attract consumer attention, they may not necessarily persuade consumers to click. This result echos theoretical propositions that ads attract limited attention (Cetin and Bingol, 2014) but on average they do not differentiate products (Telser, 1964). Only well-designed video ads persuade consumers to click, possibly because they are informative in differentiating products (Nelson, 1970), or signal unobservable product quality (Tirole, 1988).

As for the central route features, given video impression ($VideoImpression_{jt} = 1$), video ads with shorter length ($\beta = -2.380, p < 0.001$), higher content consistency ($\beta = -0.303, p < 0.01$), lower object-level complexity ($\beta = -2.084,\ p < 0.001$), higher product relevance ($\beta = 0.999, p < 0.001$), and fewer words ($\beta = -0.609, p < 0.01$) are more effective in driving CTR. All these relationships consistently indicate that if outstream video ad content can help a consumer to more efficiently learn product information, it will more effectively drive CTR. Specifically, shorter length implies that the director of the video has a more concise and compelling way of storytelling. Consistent content facilitates consumers to comprehend the key information conveyed by ads (Kim et al., 2014). Low object-level complexity indicates that the video content has fewer unrelated objects in addition to the main product. High product relevance indicates that the main product is highlighted in the video content. Both low object-level complexity and high ad-product relevance allow efficient learning, leading to high ad recall and favorable attitudes (Heckler & Childers, 1992; Li & Xie, 2020). Finally, it is much slower and harder for a consumer to learn from excessive words than from video content. According to the principle of minimum effort (Watts & Zhang, 2008), excessive words set barriers for efficient learning, and people tend to choose the most cost-saving way to learn new information, namely, through visual information.

I further highlight that the above findings on central route features are distinct from the conclusions drawn on instream video ads. In contrast to outstream, first, longer instream video ads are found to be more effective (Goodrich et al., 2015). Second, whether instream video ads are relevant to the sponsored products is less important than is their entertainment value in predicting advertising effectiveness (Campbell et al., 2017; Goodrich et al., 2015; Joa et al., 2018). The difference can result from the following aspects. First, instream video ads inevitably interrupt viewers by stopping them from watching their intended content, which tends to make consumers feel intruded upon and induce irritation (Campbell et al., 2017). Longer videos with more affective content are perceived as less intrusive and annoying because they would provide more entertainment value, and therefore increase viewership (Campbell et al., 2017; Goodrich et al., 2015; Joa et al., 2018). Second, instream and outstream video ads have different viewer groups. Instream video ads are more likely to be viewed by “prosumers,” who actively produce and consume content on social media, for example, famous YouTube personalities (Joa et al., 2018). Prosumers (e.g., YouTubers) view instream video ads because they may learn more novel ideas to create their own video content, not necessarily because they are interested in the sponsored products. Therefore, in instream video ads, they tend to pay attention to entertaining elements, ignore the product-related information, and seek content variety instead of consistency. In contrast, outstream video ad viewers on online shipping websites are goal-driven consumers. They search on the platform to find a product that meets their preferences. They need to efficiently learn product characteristics and make a clicking decision. Exposed to a website filled with a massive number of similar products, their cognitive resources are too limited for them to process long video ads with inconsistent content, complex objects, irrelevant information, and an excessive number of words.

For the peripheral route, given video impression, I find that in contrast to traditional ads (Fortin & Dholakia, 2005), social presence in video ads is not associated with a higher CTR ($HasPerson_{jt}, \beta = -0.809, p < 0.001$). Given social presence, CTR is predicted to be higher only when persons in the video ads are celebrities ($\beta = 1.212, p < 0.001$) and/or are arranged with the two shot technique ($\beta = 0.433, p < 0.05$). Further, video ads are more effective with higher levels of aesthetic design, particularly, conforming to ROT ($\beta = -1.790, p < 0.001$) and having higher color complexity ($\beta = 1.977, p < 0.001$). Finally, higher pixel-level complexity ($\beta = 0.733, p < 0.001$) is linked to a higher CTR.

Instead of helping consumers to directly learn product information, peripheral route features help consumers to efficiently make inferences about product quality. My model suggests that outstream video ads are more effective when their peripheral route features show that high costs and effort are invested in making the video content. The high costs and effort further signal product quality and drive CTR. In particular, it is much more costly to invite celebrities to endorse the sponsored products than to invite non-celebrities. Further, professional film-making techniques involved in video content (two shot, visual aesthetics, and pixel-level complexity) signal high costs and effort in making the ads. For example, many firms pay a high salary to invite famous directors to produce professional video ads.[30] I note that although visual aesthetics may also drive instream video ad viewership, existing literature suggests that viewers are more likely to enjoy the aesthetic value of the ads (Campbell et al., 2017; Joa et al., 2018), instead of using these features to make inferences about product characteristics. In addition, high pixel-level complexity induces high physiological arousal (Shin et al., 2020). Arousal can act as a source of information and the consumer infers that much effort has been made to induce her high arousal state (Yin et al., 2017).

For control variables at the clicking stage, conditional on product characteristics ($ProdEmb_{jt}$), first, a lower price is linked to a higher CTR ($\beta = -0.432, p < 0.001$). Second, the number of reviews and average ratings have a negative coefficient ($\beta = -0.077, p < 0.001$;

[30] https://www.premiumbeat.com/blog/best-commercials-by-famous-directors/

$\beta = -0.667, p < 0.001$), possibly because I focus on products listed on the first page of the search results. The number of reviews and average ratings for these products are sufficiently high, and increasing the numbers does not positively affect consumer behavior. Consistent with the literature, an unusually high number of reviews and average ratings may raise the concern of deception reviews (Anderson & Simester, 2014; Ott et al., 2012), reducing review credibility (Kusumasondjaja et al., 2012). Third, for queries in which products are more differentiated, CTRs tend to be higher ($\beta = 1.331, p < 0.001$), possibly because consumers can more easily identify and click on the favorable products. Fourth, compared to credence goods, CTRs for search goods are lower ($\beta = -0.666, p < 0.05$), whereas there is no significant difference between credence and experience goods ($\beta = -0.053, p > 0.05$). Finally, σ_j^C is not significantly different from zero ($p > 0.05$), suggesting that there is no significant unobservable query heterogeneity at the clicking stage.

4.6 ROBUSTNESS CHECKS

I run a battery of robustness checks to further validate my analyses and results. First, to validate my estimation procedure, I run a simulation study to illustrate that my model can successfully recover pre-defined true parameters. The results are available upon request. Second, I check the model fitness of my two-stage model. I run a two-stage model that includes only the constant terms, and I find that the log-likelihood is $-51{,}132$. Compared to the log-likelihood of $-48{,}444$ for Column (1) and $-48{,}436$ for Column (2) in Table 2, the results show that my model helps to explain the variance of data. The pseudo R-squares for Column (1) and Column (2) are 5.257% and 5.273%, respectively.

Third, I analyze additional composition aesthetic features beyond ROT. The other two techniques in composition documented in the existing marketing literature are diagonal dominance (arranging objects on the diagonal lines of the frame) and visual symmetry (arranging objects such that their centroid is close to the middle of the frame) (Zhang et al., 2021). I follow the existing study to construct the two variables. I find that the correlation between ROT and diagonal dominance and the correlation between ROT and visual symmetry are 0.93 and 0.92, respectively. The high correlations imply that if an ad designer professionally applies ROT for video composition, s/he is also likely to use other techniques as well. Thus, to avoid the multicollinearity issue, I choose to include only ROT as a proxy of composition aesthetics in my main model.

4.7 Discussion

Although the new format of video ads, outstream, is becoming increasingly popular in the industry, extant research focuses mainly on instream video ads (Akpinar & Berger, 2017; Campbell et al., 2017; Goodrich et al., 2015; Joa et al., 2018; Tellis et al., 2019). Combining video analytics, machine learning, and econometric analysis, I designed a large-scale, query-level observational study on click-stream data from a leading e-commerce platform to understand the effectiveness of outstream video ads. Several different model specifications consistently show that products in outstream video ads receive a more than 500% increase in CTR compared to otherwise very similar products, demonstrating economic significance. Outstream video ads increase CTR first through capturing consumer attention. Such effectiveness on attention is higher when products are less differentiated in a market. Contingent on consumer attention, video content features that facilitate efficient consumer learning or signal product quality significantly increase consumers' likelihood of clicking.

This work provides important theoretical implications. The paper is among the first to demonstrate the effectiveness of outstream video ads on consumer attention and clicking, contributing to the video marketing literature. Notably, I find that video features of outstream ads have highly distinct effects on advertising effectiveness compared to the effects of video features of instream ads documented in the literature, possibly because the two types of video ads are presented in different contexts and with different viewer groups.

Second, I show that outstream video ad effectiveness is higher when product differentiation in a market is low, which provides empirical evidence to the long-standing and controversial question in the literature of competition and advertising-whether and how competition shapes advertising effectiveness (Telser, 1964; Nelson, 1970; Tirole, 1988; Cetin & Bingol, 2014). I achieved this contribution by addressing three major empirical challenges. To begin, I leverage a large-scale and cross-market dataset from a leading e-commerce platform, in which variation in product differentiation levels is meaningful for statistical identification. Further, I quantitatively measure product differentiation for a large number of markets with product embedding, the state-of-the-art deep representation learning technique. Finally, I use a two-stage model to infer unobservable consumer attention. I show that competition is linked to higher effectiveness of video impression on attention attraction, which provides empirical evidence to the analytical proposition by Cetin and Bingol (2014). Moreover, I show that the outstream video ad effectiveness goes beyond capturing consumer attention. Well designed video ads further persuade consumers to click the advertised products, which supports the proposition that ads can be informative to consumers (Nelson, 1970; Tirole, 1988), rather than being uninformative (Telser, 1964) or inducing annoyance and avoidance (Todri et al., 2020).

Third, this work contributes to the literature of video analysis in digital marketing, by illustrating a systematic, interpretable, automatic, and scalable video ad content analytics framework. The framework bridges the research on social media image posts (Shin et al., 2020) and the literature of film-making and cinematography (Bordwell et al., 1993; Thompson & Bowen, 2009). The majority of marketing research relies on manual coding to understand video content (e.g., Smith et al., 2012; Teixeira et al., 2014; Goodrich et al., 2015). Such an approach is prohibitively costly for analyzing the large-scale video ad data on today's e-commerce systems. The few exceptions in video marketing literature focused mainly on consumer emotion/reaction analysis captured by cameras (Teixeira et al., 2012; Lu et al., 2016), or instructor emotion analysis for online classroom videos (Zhou et al., 2021). Outstream video ad content, which is designed to advertise products on online shopping websites, is fundamentally different from the above types of videos, and a new framework is needed. Motivated by Shin et al. (2020), I incorporate the ELM model (Petty & Cacioppo, 1986) into the framework. I provide empirical evidence that video features in both central and peripheral routes have a significant impact on ad effectiveness.

This work provides important managerial implications for practitioners. First, this work provides implications for outstream video ad making. My results indicate that a video ad with a compelling and concise way of storytelling (i.e., shorter length, higher content consistency, lower object-level complexity, high product relevance, and fewer words), with professional design techniques (ROT, color- and pixel-level complexity, and two shot), and with celebrity endorsement is effective in persuading clicking actions. These results imply that managers should efficiently convey key product information and invest more effort to signal product quality in video ad making. Second, this work helps advertisers to form advertising strategies. For sellers in a more competitive market, that is, of low levels of product differentiation, outstream video advertisement is suggested by my model to be a more powerful tool to attract consumer attention. Thus, these sellers can consider allocating more resources to bid on video ad placement. I note, though, that outstream video ads are also effective for markets of high product differentiation. Third, the video feature framework helps the platforms to refine their advertising recommendation systems. CTR prediction is an integral part of advertising recommendation. A high CTR indicates a better consumer experience with the ad and increases the revenue of the platform. This work provides a variety of significant video content features for CTR prediction and advertisement recommendation. Finally, my model helps the platform to construct more metrics on video ads. CTR is not the only measure of interest to understand ad effectiveness. Advertisers may also bid on ads for consumer attention and brand awareness. Consumer attention, however, is unobservable. My two-stage model based on observable data helps the platform to evaluate the unobservable consumer attention and provide more dimensions of feedback to advertisers regarding advertising outcomes.

Chapter 5. WHEN EMOTION AI MEETS STRATEGIC USERS

After learning and outperforming human reasoning in various domains (Silver et al., 2017), emerging AI technologies are learning and interpreting human emotion. AI leverages this knowledge to improve interactions between organizations and individuals, including customer service automation, online reputation management, chatbot humanization, and mental healthcare. These technologies are referred to as “emotion AI” or “empathetic AI” (M.-H. Huang & Rust, 2018; Y. Yu et al., 2023).

In application, emotion AI detects a user’s (patient or customer) emotions and determines the allocation of limited healthcare or marketing resources to the user (McStay, 2018). For example, working with the Department of Veterans Affairs and the Massachusetts General Hospital, CompanionMx, an accompanying mental health monitoring application, listens to someone speaking into their phone and analyzes the speaker's voice and phone use for signs of anxiety.[31] Beyond voice analysis, practitioners are increasingly adopting smart wearable devices to detect patients’ depression and stress. If a patient is predicted to be in an unhealthy emotional state, doctors may be allocated by the platform to follow up with the patient. The AI technology from Cogito helps call centers identify the customer’s emotions on the phone and make customer service decisions. Firms implement emotion AI that screens large-scale user-generated content on social media and online review platforms to identify negative reviews (Song et al., 2019; Yu et al., 2023). Then, the AI follows the firm's policies to decide whether to allocate marketing resources (e.g., personalized managerial responses, monetary compensations) to handle negative customer emotions.[32]

5.1 THE EMOTION AI PARADOX

I focus on the adverse selection problem in the interaction of emotion AI and strategic human agents (e.g., patients and customers). The firm wants to match resources with agents' internal emotional states (e.g., patients' depression or customers' dissatisfaction). The firm will incur a loss if its resource allocation and agents' internal emotional states are mismatched, termed misallocation loss. A naive emotion AI would allocate more resources to more negative agents.

[31] https://mitsloan.mit.edu/ideas-made-to-matter/emotion-ai-explained
[32] https://go.affectiva.com/affdex-for-market-research

Knowing the mechanism of emotion AI, rational agents have the incentive to game the system, i.e., expressing negative emotions more intensely than they naturally do, to gain more desirable allocation outcomes.

Hereafter, I use the term intensity of natural emotion expression (for brevity, natural intensity) to refer to the intensity of negative emotional expression that is consistent with one's internal emotional state. Put differently, it is one's intensity of negative emotional expression when incentives of misrepresenting their emotional state are absent. Unless otherwise noted, when I refer to emotion expression, I focus on its intensity because emotion can be expressed in different forms, including written, vocal, and facial, whereas intensity is the most common and fundamental element of various emotional expressions (Han et al., 2022; Yin et al., 2014; Y. Yu et al., 2023).

Notably, emotional misrepresentation (deviating from natural intensity) is a commonly used strategy in negotiation and bargaining (Campagna et al., 2016). Only 2% of negotiators reveal internal emotional states or facts when incentives of misrepresentation exist (Fulmer et al., 2009). As such, the organization that implements a naive emotion AI faces a paradox - emotion AI is designed to handle agents' negative emotions but eventually encourages more negative emotional feedback and generates misallocation losses.

Therefore, I ask when adopting emotion AI is economically valuable and how the associated allocation policy should be designed. Notably, escalating emotional intensities are costly for agents. They tradeoff between the benefits and costs of emotional escalation. As a screening device, sophisticatedly designed emotion AI may be valuable in estimating agents' true types based on their observed emotional expressions and then personalizing allocation that matches their natural intensities.

However, several challenges exist in identifying the economic value of emotion AI and designing allocation policies. First, algorithmic noise may undermine the value of screening (Barrett et al., 2019; Crawford et al., 2021). Second, agents' negative emotional expressions can have an undesired spillover effect on the organization, termed spillover loss. For example, negative online reviews would lower a potential customer's purchase intention and thus, negatively affect the firm's future sales performance (Song et al., 2019; Y. Yu et al., 2023). Negative emotion diffusion in social media and online depression communities lead to harmful consequences (Yu et al., 2020; Tang et al., 2021). AI may encourage negative emotion escalation and enlarge its spillover effect. Third, agents' natural intensities and abilities to misrepresent emotions (termed gaming ability) are potentially heterogeneous. If either dimension is homogeneous, using a screening device to fully separate agents is possible (Spence, 1973; Kartik et al., 2007). Under noisy signaling, spillover effect, and two-dimensional heterogeneity, whether AI can profitably automate emotion recognition and resource allocation processes is questionable and needs deep investigation.

I develop and analyze a game-theoretical model to answer this question. I explicitly model agents as strategic, and the firm is aware of the potential manipulation of emotional expression data, the existence of algorithmic noise, and the negative spillover effects of emotions. The firm moves first and decides on emotion AI adoption and resource allocation policy. Then, strategic agents respond to the firm's policy by determining their optimal emotional expression intensities based on their private information of natural intensities and gaming abilities. I first identify the conditions under which adopting emotion AI is economically valuable to the firm and leads to a unique optimal allocation policy and a stable market equilibrium. Second, I investigate the welfare impact of emotion AI. Finally, I compare adopting emotion AI with hiring human employees to recognize emotions and allocate limited resources.

5.2 KEY RESULTS AND INSIGHTS

5.2.1 *Economic Value of Emotion AI*

Managers are concerned that implementing a weak AI with noise in the system may backfire and be susceptible to gaming behavior. My first result, however, shows that the key factor that can undermine the value of emotion AI adoption is the spillover effect of negative emotions rather than the existence of algorithmic noise or gaming behavior. The spillover effect may be so high that any incentive in escalating emotions would lead to unaffordable loss to the firm, invalidating the rationale to adopt emotion AI.

However, if spillover loss is low relative to misallocation loss, I show that the screening benefit of adopting AI is first-order, and the loss associated with algorithmic noise and gaming behavior is second-order to the firm. Therefore, if the firm adopts the AI but optimally "underutilizes" the emotional expression data (i.e., making the allocation outcomes less sensitive to detected emotional expressions), the loss induced by AI adoption is minimal while the screening benefit remains sensible. Notably, if the spillover loss is negligible, it is always profitable for the firm to adopt emotion AI regardless of the bias, variance, or distribution of algorithmic noise. Further, AI systems with lower variance rather than bias in detecting emotional intensities are more profitable because the firm's optimal allocation policy can correct the AI biasedness in resource allocation. Only the variance affects equilibrium outcomes.

5.2.2 *Optimal Allocation Policy Design*

The space of optimal allocation policies is infinite. I restrict my attention to linear allocation policies because they are the most practical and interpretable policies for firms and in line with the academia's interest (Frankel & Kartik, 2019, 2022; Y. Yu et al., 2020, 2023). I show that linear allocation policies have the merit of being robust to the AI's biasedness in detecting emotions. I establish that, in most cases, the optimal linear allocation policy associated with a given AI system is unique, which implies that AI-induced market equilibria are stable and predictable.

A policy that commits to underutilizing the emotional expression data reduces gaming behavior and incentivizes agents to reveal natural emotional expressions. Beyond reducing information loss in signaling, such a policy also reduces the misallocation loss caused by algorithmic noise. Counterintuitively, underutilizing emotional expression data may not always be the optimal policy. It depends on the market composition. It is optimal when agents in the market have positively correlated natural emotional expressions and gaming abilities. However, when agents with low natural intensities tend to have high gaming abilities, the firm may find it more profitable to overutilize the emotional expression data even though it knows that the data is muddled with agents' gaming behavior and contains algorithmic noise. In this case, the firm puts agents in a "rat race" of emotional expression, where all agents escalate emotional intensities to costly signal their true types.

5.2.3 *Welfare Impacts*

As a screening device, emotion AI impacts agents' utility heterogeneously. Agents with high levels of natural emotional expressions and gaming abilities benefit from emotion AI adoption, whereas other agents are worse off. Adopting emotion AI and screening based on emotional expression data strictly reduces agents' utility on average and increases the firm's utility.

The welfare implication of algorithmic noise is more intricate. My result challenges the conventional wisdom that algorithmic noise is always harmful and a stronger AI is always desirable. I establish that marginally enlarging the variance of algorithmic noise (adopting a weaker AI) may increase social welfare. To explain this, I first focus on AI weakness and the agent's and firm's welfare. The variance of algorithmic noise biases the firm's allocation policy toward underutilizing emotional expression data and reduces the firm's allocation discrimination over agents. Agents are less incentivized to escalate emotional intensities, which reduces their signaling costs. In sum, when the variance of the noise marginally increases, agents will receive higher utility on average.

The impact of AI weakness on the firm's utility is less obvious. A weaker AI reduces the firm's screening power but also reduces agents' gaming behavior and makes their signals more informative. Mitigating emotional escalation also means reducing the spillover loss caused by negative emotions. Therefore, implementing a weaker AI has competing impacts on the firm's expected utility. To my surprise, it can be shown that the firm's expected utility always decreases when a weaker AI is implemented. Still, because the variance of algorithmic noise has competing impacts on the firm's utility, the game between the firm and agents are not zero-sum. In this vein, the firm's optimal level of utilizing the emotional expression data does not equal the socially optimal point. If the firm finds it profitable to implement AI, it will always overutilize the data compared to the socially optimal level.

A weaker AI may increase social welfare when agents' marginal benefit exceeds the firm's marginal loss while decreasing social welfare is also possible. I show that the direction of social welfare change depends on the curvature of the firm's loss function at its minimum point and a threshold determined by agents' stakes over the firm's allocation and their average gaming ability.

5.2.4 *Comparing AI and Human Service Systems*

When negative emotional content is massive in scale, AI outperforms employees in recognizing emotions because the marginal cost for AI to scale up is almost zero. Even regardless of scalability, I show that using AI to automate emotion recognition and resource allocation is profitable under a broad set of conditions. The key intuition behind this conclusion is that human employees exposed to negative emotions have behavioral responses. The firm needs to tradeoff between the loss due to algorithmic noise and the loss due to employees' emotional responses. Employees may be empathetic to agents' negative emotional experiences and over-allocate resources to them. Under-allocation is also possible when employees exposed to negative emotions feel emotionally abused. Negative emotions increase employee turnover rates. I show that when employee-specific loss (misallocation and turnover), agents' stakes over allocation, natural intensities, or gaming behavior are of sufficiently high levels, AI systems are preferred over human systems. The intuition is that AI systems can prevent employees from exposing to excessive negative emotions and reduce the firm's costs. Moreover, even if agents' stakes and employee-specific loss are low, human systems do not outperform AI systems when the heterogeneity in natural intensities is either minimal or sufficiently large.

5.3 IMPLICATIONS

I summarize the implications of this work. First, I identify the key tradeoffs in emotion AI adoption. My work mitigates the concern that algorithmic noise and gaming behavior may undermine the economic value of emotion AI (Barrett et al., 2019). Instead, I highlight that adopting emotion AI may not be valuable when it exacerbates the negative spillover effect of emotional expressions. My work provides practical solutions in deciding whether to adopt emotion AI. When adopting emotion AI systems, managers should minimize the variance of algorithmic noise rather than the biasedness. Compared to human service systems, the economic value of AI also stems from preventing employees from exposing to excessive negative emotions, relieving their burden of emotional labor, and maintaining a consistent allocation policy for the firm. Second, I provide optimal data-driven allocation policy design guidelines. Depending on the market composition, managers should decide on underutilizing or overutilizing emotional expression data. Linear allocation policies have merits regarding interpretability, robustness to emotion detection bias, and market equilibrium stability. Third, I call for regulating emotional data utilization but not for naively banning emotion AI. As firms tend to overutilize emotional data compared to the socially optimal point and create deadweight loss, regulators should restrict the firm's utilization of emotional data. It is in line with the call for protecting data privacy and underutilizing emotional data (Crawford et al., 2021). Finally, algorithmic noise may not necessarily be undesirable from a social designer's perspective. It may increase social welfare by implicitly regulating firms' utilization of emotional data and agents' gaming behavior.

5.4 MODEL

An agent has a type $(\eta,\gamma) \in \mathcal{R}^2_{++}$, where natural intensity η is the highest intensity that the agent can express at a minimum cost, and gaming ability γ parameterizes the agent's cost to escalate her intensity of negative emotional expression above the natural intensity. A higher γ means that the agent has a higher ability (i.e., lower marginal costs) to escalate negative emotional expression intensities. I allow (η,γ) to have a correlation and follow a joint distribution $F(\mu_\eta,\sigma^2_\eta,\mu_\gamma,\sigma^2_\gamma,\rho)$, where μ_η and σ^2_η are the mean and variance of η; μ_γ and σ^2_γ are the mean and variance of γ; $\rho \in (-1,1)$ is the correlation between η and γ, i.e., $\rho = \text{Cov}\,(\eta,\gamma)/(\sigma_\eta\sigma_\gamma)$.

The firm cannot observe the agent's type (η,γ). To make my model more general, I further assume that the firm can only observe the parameters $(\mu_\eta,\sigma^2_\eta,\mu_\gamma,\sigma^2_\gamma,\rho)$ without observing the functional form of F. The firm's primary goal is to match an allocation $y \in \mathcal{R}$ to the agent's type η, with a quadratic loss of $(y-\eta)^2$. For example, in the case of customer service, over-compensation $(y > \eta)$ incurs an unnecessary loss to the firm, whereas under-compensation $(y < \eta)$ makes the customer unsatisfied and increases the firm's probability of losing the customer. The firm chooses a policy $y = Y(x)$ to allocate the limited resources, as a function of the agent's intensity of expressed emotion $x \in \mathcal{R}$. In other words, my setting is a signaling game in which an agent sends a signal x to inform the firm about her private type η. In addition, the agents have potentially heterogeneous ability γ to game the allocation rule $Y(x)$. To differentiate η,γ, and x, I refer to η and γ as the agent's type and refer to x as the agent's action because η and γ are the agent's endowment whereas x is the result of the agent's decision.

If the firm implements an AI system to detect the agent's action x, the detected signal $\hat{x}$ contains noise (the prediction error) $e \in \mathcal{R}$ produced by the AI system ($\hat{x} = x + e$). I denote the mean of the noise as $\mu_e = \mathbb{E}[e]$ and the variance as $\sigma^2_e = \mathbb{E}[e^2] - \mu^2_e$. In reality, $\hat{x}$ is a potentially biased estimator of $x(\mu_e \neq 0$ and $\sigma_e > 0)$ because even the state-of-the-art AI technologies cannot perfectly detect emotional expressions due to the subtle and intricate nature of human emotions. I assume that μ_e and σ_e are known to the firm. Typically, the emotion AI is trained by the firm in a supervised manner using a labeled data set that contains the ground-truth labels of the intensity of emotional expressions. The firm can derive the mean and variance of the prediction error e based on the predicted $\hat{x}$ and the ground-true x in the labeled set. It is reasonable to assume that e is independent of the agent's type (η,γ) because the labeled set and the training process of emotion AI are not related to the agents' types or actions. For generalizability, I allow the distribution of e to be unknown to the firm and the agent because the emotion-detection algorithms are often based on deep neural network architectures, which are highly non-linear and intricate (Y. Yu et al. 2023). The asymptotic distribution of such an algorithm-generated estimator is difficult to characterize.

Remark 1. σ_η and σ_γ reflect the heterogeneity in agents' types and have important impacts on the equilibrium outcomes. If $\sigma_\eta = 0$, the firm knows that the market has only one type of agents (with $\eta = \mu_\eta$) and it can allocate the resources $y = \mu_\eta$ regardless of the agent's action x. In this case, no emotion AI needs to be implemented. If $\sigma_\gamma = 0$, meaning that all agents have the same marginal cost of increasing x and the lower type would always need to incur an additional cost to send the same x compared to the higher type, then a full separating equilibrium can be achieved (Kartik et al., 2007). I am particularly interested in the case where both agents' natural intensities and gaming ability are heterogeneous $(\sigma_\eta, \sigma_\gamma > 0)$, while the model includes the situations where $\sigma_\eta = 0$ and $\sigma_\gamma = 0$ as special cases. With the two dimensions of heterogeneity, the firm is faced with the situation of "muddle information" (Frankel & Kartik, 2019) where the information about the agent's natural intensity η is muddled with irrelevant information about her gaming ability γ.

Remark 2. ρ depicts the market composition that shapes the equilibrium outcomes. I can discretize η and γ and identify four types of agents (Frankel & Kartik, 2019), i.e., (1) high- η and high- γ, or the "the high-type gamer", (2) high- η and low- γ, or the "the high-type natural," (3) low- η and high- γ, or the "the low-type gamer," and (4) low- η and low- γ, or the "low-type natural." A positive ρ reflects that a market is composed of more high-type gamer and low-type natural agents, whereas a negative ρ reflects that a market is composed of more high-type natural and low-type gamer agents. Screening agents based on their types in a market with $\rho < 0$ (versus a market with $\rho > 0$) can be more challenging because in this case the cost for a low-type agent to mimic the high-type agents tend to be lower and the signal x sent by the agent is less informative to the firm regarding the agent's type.

5.4.1 *Agent's Problem*

The cost for an agent of type (η, γ) is denoted as $C(x; \eta, \gamma)$. I assume that the function $C(\cdot; \eta, \gamma)$ is continuous and twice-differentiable. Further, based on my context, I characterize a set of conditions that $C(\cdot; \eta, \gamma)$ needs to satisfy. First, C needs to be increasing and convex in x because more effort is needed to express a higher level of emotion $\left(\frac{\partial C}{\partial x} > 0\right)$ and, as the emotional expression level increases, the marginal cost of escalating the emotional expression level also increases $\left(\frac{\partial^2 C}{\partial x^2} > 0\right)$. Second, the marginal cost of escalating x should be lower for agents with higher natural intensity η in that they are naturally more negative than the other agents and can easily express more negative emotions $\left(\frac{\partial^2 C}{\partial x \, \partial \eta} < 0\right)$. Relatedly, the marginal cost of escalating x should be lower for the agents with higher gaming ability γ by definition $\left(\frac{\partial^2 C}{\partial x \, \partial \gamma} < 0\right)$. Third, the cost of expressing a natural level of emotion $C(\eta; \eta, \gamma)$ is normalized as zero.

The value function of an agent's action x is denoted as $V(x; Y, e, s)$, where x, e, and Y determine the firm's received signal $\hat{x}$ and allocation $y = Y(\hat{x})$. $s \in \mathcal{R}_{++}$ captures the agent's exogenous stakes over the firm's allocation y. Intuitively, s varies across service contexts. When y represents a doctor's time and effort allocated to a patient in the mental healthcare context, the patient's stakes for y are high (which leads to a large s). If y represents the monetary compensation that an E-commerce merchant pays a customer for failing to deliver a low-involvement product timely, the customer's stakes tend to be low (which leads to a small s). s can also vary across markets. For example, in a developing country where customers have lower income than those in a developed country, customers' stakes over a given monetary compensation tend to be higher.

I parameterize $V(x; Y, e, s) = sY(\hat{x}) = sY(x + e)$. A leading example of C that satisfies my conditions is $C(x; \eta, \gamma) = (x - \eta)^2/\gamma$, as commonly assumed in the related literature (Frankel & Kartik, 2019, 2022). An agent seeks to choose x based on her private type (η, γ) and maximize her utility $U_a \triangleq V(x; Y, e, s) - C(x; \eta, \gamma)$ in expectation:

$$\max_x \mathbb{E}[U_a] = \max_x \mathbb{E}[V(x; Y, e, s) - C(x; \eta, \gamma)] = \max_x \mathbb{E}[sY(x + e) - (x - \eta)^2/\gamma] \qquad (1)$$

5.4.2 *Firm's Problem*

Assume the firm chooses among linear allocation rules, i.e., choosing $(\beta, \beta_0) \in \mathcal{R}_+ \times \mathcal{R}$ such that

$$Y(x) = \beta x + \beta_0. \tag{2}$$

$\beta \geq 0$ is assumed because the firm cannot punish an agent for expressing her emotions. An important decision for the firm is whether to choose $\beta = 0$ or $\beta > 0$. When the optimal β is zero, the emotion AI does not provide economic value to the firm. In other words, it is not profitable for the firm to screen customers based on detecting their expressed emotional intensity x. The firm would simply allocate a fixed amount of $y = \beta_0$ to all agents.

I consider the agent's best response to the linear allocation rules. To simplify my notation, I denote stakes $m \triangleq s/2$. I denote x^* as the best response. From the first-order condition $\frac{\partial \mathbb{E}[U_a]}{\partial x} = 0$ (and $\frac{\partial^2 \mathbb{E}[U_a]}{\partial x^2} = -2 < 0$), I have:

$$x^* = \eta + \frac{s}{2}\beta\gamma \triangleq \eta + m\beta\gamma \tag{3}$$

Remark 3. The agent's best response to a linear allocation rule is also linear. The best response x^* contains two parts, the natural intensity η and the gaming behavior $m\beta\gamma$. When the firm decides to implement AI ($\beta > 0$), natural intensity η is muddled by gaming ability γ and stakes m. The gaming behavior increases when the exogenous stakes m, the incentive β, and the gaming ability γ are higher. It is in line with Goodhart's Law which states that when a measure (in my context, the expressed emotional intensity x) becomes a target (of limited resource allocation y), it ceases to be a good measure.

Remark 4. x^* is not affected by the algorithmic error e because e is independent to the agent's type (η, γ) and the policy parameter β. The terms related to $\mathbb{E}[e]$ in the agent's utility function is a sunk cost (or gain) to the agent. This resembles the situation that, in reality, the agents do not have to know the details of the algorithms before entering the game. They only know that they are more likely to obtain resources if expressing a higher intensity of negative emotion. In this vein, the model does not concern the topic of algorithmic transparency (see Q. Wangs et al. (2022) and references therein) but focuses more on the roles of the algorithmic noise in the game.

The firm's utility function contains two parts, the quadratic loss induced by resource misallocation $(Y(\hat{x}) - \eta)^2$ and the spillover loss which, aligned with the misallocation loss, is also assumed to be quadratic (λx^2). Intuitively, the spillover effect of extremely intense negative emotions could be marginally more harmful to the firm than that of low-intensity negative emotions.

3.2.1. Negative Spillover Effect and Implication of $\boldsymbol{\lambda}$ I use $\lambda \in \mathcal{R}_+$to capture the relative importance of the spillover loss due to the agent's emotional expression compared to the misallocation loss. If $\lambda = 0$, the firm only cares about the misallocation loss. The case $\lambda = 0$ is possible when the agent interacts with the firm through a private channel in which no negative spillover effect for the firm would be generated by the agent's negative emotional expression. For example, the agent initiates a private call to the firm's customer service system. However, there are many situations where the agent's interaction with the firm is not completely private, and x would have a significantly negative spillover effect on the firm's welfare ($\lambda > 0$). For instance, first, a customer's negative online review would be observed by potential customers. It is empirically shown that negative reviews have a significantly negative causal impact on the potential customers' purchase intentions and, therefore, the firm's future revenue (Y. Yu et al., 2023). In some cases, the loss due to negative reviews would be much higher than the misallocation loss ($\lambda \gg 1$), so the online seller may take the moral or legal risks to bribe the reviewers in exchange for biased and less negative reviews (Mostagir & Siderius, 2022). Second, negative emotions are contagious in online social networks (Y. Yu et al., 2020), especially those in online depression communities (ODCs) (Tang et al., 2021). Some members in ODCs may have the moral hazard to over-express their depression in exchange for more mental healthcare resources. Over-expressed depression puts the other member at higher mental health risks due to depression contagion. The increased diffusion of depression in online communities decreases the utility of ODC operation managers.

In sum, the firm's loss function is given by:

$$L = \mathbb{E}[(Y(\hat{x}) - \eta)^2 + \lambda x^2],\ \lambda \in \mathcal{R}_+ \tag{4}$$

I do not include the cost of implementing the emotion AI because once a model is trained, deploying it is of nearly zero marginal cost to the firm. Further, there are open-source algorithms and models for the firm to detect emotions (Yu et al. 2022). Even if there are fixed or marginal costs of using the emotion AI system, the tradeoff is trivial because the firm can compare the reduced loss calculated by the above equation with the costs of AI adoption. Therefore, I focus on the key tradeoffs.

3.2.2. Decomposition of Firm's Loss Function I decompose the loss function to derive insight into the sources of loss to the firm.

Proposition 1. Under the optimal β_0, the firm's loss function can be written as:

$$L(\beta) = \underbrace{\beta^2\sigma_e^2}_{\text{Loss due to noise}} + \underbrace{\lambda\mathbb{E}[x^2]}_{\text{Spillover loss}} + \underbrace{\mathbb{E}[(\mathbb{E}[\eta \mid x] - \eta)^2]}_{\text{Loss of estimating } \eta \text{ given the true } x} + \underbrace{\mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])^2]}_{\text{Misallocation loss given estimation}} \quad (5)$$

where $Y_0(x) = \beta x + \mathbb{E}[\eta - \beta x]$ is the optimal allocation rule given β when $e \equiv 0$.

Remark 5. Proposition 1 indicates that the biased term of the algorithmic error (μ_e) and the distribution of e do not affect the firm's loss and, therefore, the firm's decision of β.

The proof of Proposition 1 is provided in Appendix C.1. Proposition 1 highlights four sources of loss to the firm. First, Proposition 1 and Remark 5 indicate that compared to the case when the AI perfectly estimates the true signal $x(e \equiv 0)$, the only term related to e that affects the firm's loss is the variance of e and the policy parameter β. This result is counterintuitive because one would expect that an unbiased AI-generated estimator with $\mu_e = 0$ is generally preferred. However, given that the biased term (μ_e) is known, I show that under the linear allocation rules, the firm can always choose an optimal β_0 to correct the measurement error without knowing its distribution. This result speaks to the recent literature on de-biasing machine-learning generated variables with semi-/non-parametric methods (Chernozhukov et al. 2022), especially when these variables are used in (generalized) linear specifications (Qiao and Huang 2021). Further, the machine-learning practitioners always face the so-called "bias-variance dilemma," and my result implies that the firm would prefer the minimum-variance machine-learning generated estimator.

Second, the rest of the three parts in the loss function are the spillover loss due to x, the information loss due to estimating η provided that the firm receives the true signal x, and the misallocation loss given the estimation of η. Compared to the strategy $\beta = 0$, a firm may benefit from implementing $\beta > 0$ to better match its allocation to agents according to x and therefore reduce the cost of resource misallocation (the last term in the loss function). Increasing β may or may not reduce the information loss due to estimating η based on x. Increasing β makes the agents with a higher γ express a higher x. If the natural intensity η is positively correlated with γ, x could be a more efficient indicator of η in this vein, and vice versa. A positive β not only increases the costs due to the existence of the algorithmic error ($\sigma_e^2 > 0$) but the spillover loss because $\beta > 0$ unexpectedly encourages the gaming behavior $m\beta\gamma$ and, hence, increases $\mathbb{E}[x^2]$.

3.2.3. The Linear Allocation Assumption I provide more discussion on the rationale of following the linear allocation assumption (Frankel & Kartik, 2019, 2022) in my context. First, Proposition 1 implies that the linear allocation rules are robust to the bias of the AI-generated estimator. Second, linear policies are simple and practical for firms to implement. These policies preserve the order between the intensity of emotional expression and the allocation, which mimics the firm's goal of allocating more resources to an agent with a higher level of negative emotions. Third, the economic value of the detected emotional variables is usually studied in a linear model; therefore, my assumption is aligned with the practice of the related empirical studies (Y. Yu et al. 2020, 2023). Fourth, according to Equation 3, the agent's best response to a linear allocation rule is also linear, which provides a clear relationship among signal x, natural intensity η, and the gaming behavior $m\beta\gamma$.

Lastly, the parameter β in the linear allocation rules is not only interpretable but carries important economic implications. The parameter β captures how the firm would use the emotional expression data detected by the AI system. The optimal $\beta = 0$ means that the firm would strategically ignore emotional expressions either because they are not informative about an agent's natural intensity or because a positive β encourages the agents' gaming behavior which would induce a negative spillover effect unaffordable to the firm. Then there is no economic value in implementing emotion AI. The optimal $\beta \in (0,1)$ means that emotion-AI detected information is valuable in screening the agents' types. Still, the firm also realizes that the emotional expression is muddled information and that a high β may encourage gaming behavior and increase the negative spillover effect. In this vein, the firm decides to underutilize this data of x ($\beta < 1$) in its allocation policy. The case when the optimal $\beta > 1$ is counterintuitive but not impossible. It indicates a situation where the firm overutilizes the emotional expression data even though the firm knows that it is muddled information, and a high β increases the negative spillover effect. The firm would do so when the information of natural intensity η is not completely muddled by gaming ability γ. Then, every agent is encouraged by the firm to "over-express" their emotions ($x > \eta + m\gamma$) and costly signal their true type η.

Before I present the model analysis, I summarize the main notation of the paper in Table 5.1.

Table 5.1 Summary of Main Notation

Notation	Definition
η	Agent's natural intensity
γ	Agent's gaming ability
ρ	Correlation between η and γ
μ_i	Expected value of a random variable i (e.g., $i = \eta$ or γ)
σ_i^2	Variance of a random variable i
y	Firm's allocation
x	Agent's intensity of expressed negative emotion
e	Error in emotion detection
$\hat{x}$	Detected/estimated intensity of expressed negative emotion by AI, i.e., $\hat{x} = x + e$
m	Defined as $s/2$, where s is the agent's exogenous stakes over allocation
Y	Policy function that determines allocation, with policy parameters β and β_0
C	Agent's cost function
V	Agent's value function
U_a	Agent's utility function
λ	Relative importance of spillover loss compared to misallocation loss to the firm
L	Firm's loss function

5.5 ANALYSIS

5.5.1 *Equilibrium Analysis*

The firm decides on AI adoption and designs allocation policies before the agents respond to the firm's strategies. Therefore, I consider the case where the firm has the first-mover advantage by committing to an optimal pair of (β, β_0) and seeks a Stackelberg solution. Although there always exits a pooling Nash Equilibrium in which the firm chooses $(\beta, \beta_0) = (0, \mu_\eta)$ and agents choose to reveal their natural intensity η, the firm may choose an equilibrium with $\beta > 0$ if it is more profitable. I am interested in (1) under what conditions there exists an equilibrium with $\beta > 0$ because it justifies the economic value of adopting the emotional AI; (2) whether this (separating) equilibrium, if any, is unique because uniqueness would increase the predictive power of my results to the reality and rule out the possibility of mixed strategies; (3) what the roles of the algorithmic noise plays in shaping the equilibria; and (4) how should the best policy be designed, i.e., what conditions should the optimal β^* satisfy.

Formally, according to Proposition 1, the firm's problem can be written as:

$$\min_{\beta\geq 0,\beta_0} \mathbb{E}[(Y(\hat{x})-\eta)^2+\lambda x^2] = \min_{\beta\geq 0} L(\beta) \tag{6}$$

When the agent's best response is $x=\eta+m\beta\gamma$, $Y(\hat{x})=\beta(x+e)+\beta_0$ is quadratic in β, and $L(\beta)$ is quartic in β. Therefore, directly solving β^* is challenging, and the behavior of L with respect to β is intricate. Each optimal value of β^* pins down a Stackelberg equilibrium. The equilibrium outcomes are provided in Lemmas 1 and 2, visualized in Figure 5.1, and proved in Appendix C.2. Denote that $\lambda_0 \triangleq \frac{\mathrm{Var}\,(\eta)}{m\mathbb{E}[\eta\gamma]}$ and $\lambda_1 \triangleq \frac{2m\mathrm{Cov}\,(\eta,\gamma)-\mathrm{Var}\,(\eta)-\mathrm{Var}\,(e)}{m^2\mathbb{E}[\gamma^2]}$.

Lemma 1. For $\rho \in (-1,0)$,

(1) $\beta^*=0$ if $\lambda \geq \lambda_0$;

(2) $\beta^* \in (0,2)$ and unique if $\lambda < \lambda_0$.

Lemma 2. For $\rho \in [0,1)$,

(1) When $\lambda_0 \geq \lambda_1$, (1a) $\beta^*=0$ if $\lambda \geq \lambda_0$; (1b) $\beta^* \in (0,1)$ and unique if $\lambda < \lambda_0$;

(2) When $\lambda_0 < \lambda_1$, there exists a unique $\lambda_0^* \in (\lambda_0,\lambda_1)$ such that (2a) $\beta^*=0$ if $\lambda > \lambda_0^*$; (2b) $\beta^* \in (0,1)$ and unique if $\lambda < \lambda_0^*$; (2c) there exist two equilibria, $\beta^*=0$ and a unique $\beta^* \in (0,1)$, if $\lambda=\lambda_0^*$.

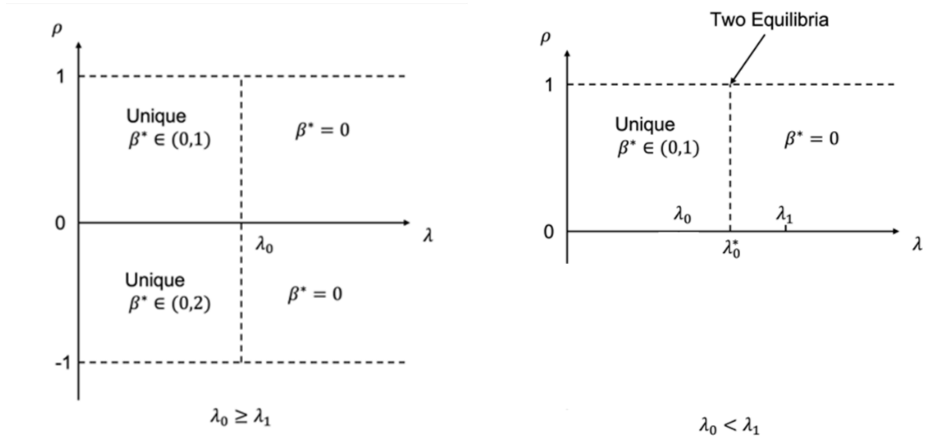


Figure 5.1 Equilibrium Outcomes

Lemmas 1 and 2 provide important managerial implications. First, they imply a sufficient condition for implementing AI ($\beta^* > 0$), that is, when λ is smaller than a threshold. In other words, the emotion AI is of economic value when the spillover loss of negative emotional expression is relatively of lower importance compared to the misallocation loss. It formally shows the key tradeoff that although AI can help the firm to screen the agent's type η, it incentivizes the agent to game the system and cause negative externality beyond the interaction. In most of the cases, the threshold of λ takes the simple form of $\lambda_0 = \frac{\text{Var}(\eta)}{m\mathbb{E}[\eta\gamma]}$. It closely relates to the economic value of screening. In particular, the threshold increases in $\text{Var}(\eta)$, the heterogeneity in the natural intensity. The more heterogeneity in the natural intensity, the more benefit of implementing a screening device. The threshold decreases in m and γ because the gaming behavior, quantified by $m\beta\gamma$, decreases the informativeness of the signal and, therefore, decreases the benefit of screening.

Remark 6. The second part of Lemma 2 implies that $\lambda < \lambda_0 = \frac{\text{Var}(\eta)}{m\mathbb{E}[\eta\gamma]}$ is not a necessary condition for the applicability of AI ($\beta^* > 0$). When $\lambda_1 > \lambda_0$, the emotion AI can have economic value even if $\lambda > \lambda_0$. It implies that $\text{Cov}(\eta, \gamma) > 0$ and, thus, $\rho > 0$. In addition, $\lambda_0 < \lambda < \lambda_1 = \frac{\sigma_\eta(2m\rho\sigma_\gamma - \sigma_\eta) - \sigma_e^2}{m^2(\mu_\gamma^2 + \sigma_\gamma^2)} < \frac{m^2\rho^2\sigma_\gamma^2}{m^2(\mu_\gamma^2 + \sigma_\gamma^2)} < \rho^2 < 1$. Therefore, $\lambda < \lambda_0 = \frac{\text{Var}(\eta)}{m\mathbb{E}[\eta\gamma]}$ is a necessary condition for $\beta^* > 0$ if $\lambda \geq 1 > \rho^2$. It follows that if the spillover loss is at least as important as the misallocation loss ($\lambda \geq 1$), $\beta^* > 0$ if and only if $\lambda < \lambda_0$.

Remark 7. The condition $\lambda < \lambda_0$, which guarantees the economic value of AI, is independent of the bias, variance, and the distribution of the noise e.

This is counterintuitive because firms are usually concerned that implementing a weak emotional AI may backfire. However, the loss due to the noise is second-order, and the gain of screening is first-order. If the firm commits to underutilizing the detected signal $\hat{x}$ by choosing a small and optimal β, Lemmas 1 and 2 guarantee the economic value of implementing the emotional AI. Compared to e, the agents' stakes are more influential in determining the value of implementing AI.

Remark 8. Holding other parameters constant, given $\lambda > 0$, $\beta^* = 0$ is the unique equilibrium when m is sufficiently large.

Note that as $m \to \infty$, I have both $\lambda_0 \to 0_+$ and $\lambda_1 \to 0_+$. Therefore, for any given $\lambda > 0$, it must hold that $\lambda > \max\{\lambda_0, \lambda_1\}$. According to Lemmas 1 and 2, $\beta^* = 0$ is the unique equilibrium. It implies that in the service contexts where the agents' exogenous stakes over the firm's allocation are sufficiently large, the gaming behavior would be so salient that it undermines the informativeness of the signal and imposes a highly negative spillover effect. In this case, the value of implementing AI is minimal.

Second, the two lemmas imply that when the firm-agent interaction channel is completely private ($\lambda = 0$), implementing an emotion AI system is always profitable. One would expect that implementing the emotion AI could sometimes have no economic value due to the existence of the gaming behavior and the algorithmic noise. Counterintuitively, the above statement holds regardless of the gaming behavior and the algorithmic noise. When there is no spillover loss, marginally increasing β from 0 provides a first-order gain to the firm due to screening and incurs a second-order loss due to the gaming behavior and the algorithmic noise. It can be shown that the derivative of the firm's loss function at $\beta = 0$ is $-2\text{Var}(\eta)$. In other words, the marginal gain of choosing a positive β is linearly related to the agent's natural intensity heterogeneity.

Third, the indicator of market composition ρ plays an important role in shaping the equilibrium. When $\rho \geq 0$, the market is composed of agents whose natural intensity and gaming ability are likely to be positively correlated. The optimal policy in this case, as Lemma 2 indicates, always under-utilizes the data of $\hat{x}$ ($\beta^* < 1$). This is because in this market (versus the market with $\rho < 0$), lower types (agents with lower η) tend to have a higher marginal cost to mimic the higher types. A small incentive β would be sufficient to separate the lower types from the higher types. Given the firm knows that $\mathbb{E}[x] > \eta$ due to agents' gaming behavior, it discounts the signal $\hat{x}$ by implementing $\beta < 1$. Another motivation for the firm to underutilize the signal beyond its reduced informativeness is to reduce the loss due to the noise (σ_e^2) and the spillover loss, as shown in Proposition 1.

Counterintuitively, according to the second part of Lemma 1, when $\rho \in (-1,0)$ the firm may commit to over-utilize the data of $\hat{x}$ by choosing a $\beta^* \in (1,2)$ even though it is aware that agents' gaming behavior muddles the data and that such a policy would further encourage the gaming behavior, increase the loss due to noise, and increase the spillover loss. It happens if $\lambda < \lambda_0$ and the derivative of the loss function at $\beta = 1$ is negative. Using the standard mean-variance decomposition, I can show that it is equivalent to the following inequality (for more details of $L'(\beta)$, see Appendix C.2):

$$\sigma_e^2 + \lambda m\left(m\mu_\gamma^2 + m\sigma_\gamma^2 + \rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\right) < m\sigma_\gamma\left(-\rho\sigma_\eta - 2m\sigma_\gamma\right)$$

The left-hand side positively correlates with (while is not equivalent to) the loss due to the noise and the spillover loss. The right-hand side must be positive for the inequality to hold, that is, $-\rho\sigma_\eta > 2m\sigma_\gamma \cdot 2m\sigma_\gamma$ increases when the stakes $s = 2m$ increases and the heterogeneity in gaming ability γ increases. In other words, it positively correlates with gaming behavior. $-\rho\sigma_\eta$ increases with the heterogeneity in the natural intensity $\text{Var}(\eta)$. Therefore, it positively correlates with the benefit of screening. It also increases as $\rho \to -1$, which indicates that there are more high-type natural and low-type gamer agents than the other two types in the market.

The low-type gamer may mimic the high-type natural in this situation. But sometimes, the difference in natural type η is so large that the low-type gamer expresses a lower level of x than the high-type natural even after considering gaming the system. In this vein, the firm finds it profitable to use $\beta > 1$ to optimally leverage the difference in x, making the difference in allocation y closer to the difference of their true types η. Then, an optimal choice of β_0 ensures the expected value of allocation y matches the expected value of η after correcting the bias of algorithmic noise $\mathbb{E}[e]$. An additional requirement is that such screening benefit is not fully eroded by the loss due to the variance of the noise and the spillover loss. This holds when, for instance, $\sigma_e^2 \to 0$ and $\lambda \to 0$. Nevertheless, Lemma 1 guarantees that even when $\sigma_e^2 \to 0$ and $\lambda \to 0$, using a very large β is unreasonable and the upper bound of β^* is 2.

The uniqueness of β^* indicates that the optimal linear allocation rule is unique and the market equilibrium is stable. The equilibrium is unique except the special case when $\lambda = \lambda_0^* \in (\lambda_0, \lambda_1)$. In this case, when β marginally increases from 0 to a positive value, the firm's loss increases because it produces a first-order loss due to the spillover effect. There is a second-order loss due to the noise and a second-order gain due to screening. But as β keeps increasing and reaches a local minimum point, the screening benefit happens to offset the loss due to the spillover effect and the loss due to the noise. Therefore, $\beta^* = 0$ and $\beta^* \in (0,1)$ are indifferent to the firm and two equilibria exist. The second part of Lemma 2 further guarantees that at most two equilibria exist. Figure 5.2 illustrates three cases of the equilibria. The numerical setting is $\mu_\gamma = \mu_\eta = 1, \sigma_\gamma = 1, \sigma_\eta = 0.1, \rho = 0.5$, and $m = 2$. As λ increases from 6.0×10^{-3} to 7.0×10^{-3}, the equilibria shift from $\beta^* > 0$ to $\beta^* = 0$. Notably, there exist two equilibria when $\lambda = \lambda_0^* = 6.45 \times 10^{-3}$ and $L(0) = L(\beta_+^*)$ where β_+^* is the positive minimum point of $L(\beta)$.

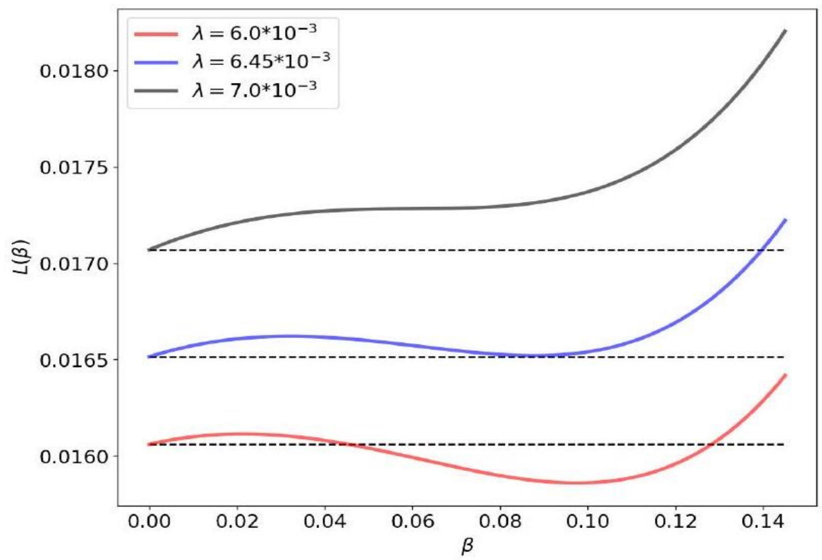


Figure 5.2 Firm's Loss Function $L(\beta)$

5.5.2 *Welfare Analysis*

Next, I analyze the impact of the emotion AI and its noise on the agent's, firm's, and social welfare. First, I focus on the agent's welfare. In Appendix C.3, I show that under the optimal β_0, the agent's welfare can be written as:

$$U_a = 2m\mu_\eta + 2m(\eta - \mu_\eta)\beta + m^2(\gamma - 2\mu_\gamma)\beta^2 + 2m\beta(e - \mu_e) \tag{7}$$

Equation 7 implies that when the firm implements an AI system (changing from $\beta = 0$ to $\beta > 0$), the impact on the agent's welfare is heterogeneous. Agents with above-average η (the high-type) obtain a first-order benefit and those with a relatively low gaming ability $(\gamma < 2\mu_\gamma)$ obtain a secondorder loss. If e is specific to each agent, agents would be affected by the algorithmic bias quantified by $2m\beta(e-\mu_e)$. Notably, the algorithmic bias is different from the bias of the estimator for the emotional expression μ_e. The algorithmic bias describes repeatable errors in a computer system that create "unfair" outcomes (Teodorescu et al., 2021). For example, for some agents, e may be consistently different from μ_e and the loss or gain from the algorithmic bias $2m\beta(e-\mu_e) \neq 0$. But in Appendix C.3 I show that, as one merit of the linear allocation policy, the firm would choose an optimal β_0^* and makes the expected algorithmic bias to be zero.

The agent's expected utility is quadratically decreasing in β and is independent of the noise e :

$$\mathbb{E}[U_a] = 2m\mu_\eta - m^2\beta^2\mu_\gamma \tag{8}$$

It follows that if the firm is indifferent between the equilibria of $\beta^* = 0$ and $\beta^* > 0$, the pooling equilibrium $\beta^* = 0$ is the Pareto dominant equilibrium.

Next, I will focus on the impact of implementing AI on social welfare. I define the expected social loss in each agent-firm interaction as:

$$L_s(\beta) \triangleq L(\beta) + m^2\beta^2\mu_\gamma$$

Denote $\beta_s^* \triangleq \arg\min_{\beta\geq 0} L_s(\beta)$ as the socially optimal point of β. In Appendix C.3, I show that the following proposition holds.

Proposition 2. β_s^* satisfies that (1) If $\beta^* = 0, \beta_s^* = 0$; (2) If $\beta^* > 0, 0 \leq \beta_s^* < \beta^*$. If $\lambda < \lambda_0$, $\beta_s^* > 0$.

The first part of Proposition 2 states that when the firm chooses a pooling equilibrium of not implementing AI, it is also the socially optimal equilibrium. This is because implementing AI reduces the agent's expected utility. The second part of Proposition 2 implies that if it is profitable for the firm to implement AI, the firm will choose a larger β than the socially optimal point β_s^*. In this vein, they can obtain more utility from screening, and the agent's expected utility strictly decreases when β increases. The firm's benefit from implementing AI is strictly smaller than the loss to the agents, which induces the deadweight loss. Intuitively, when the firm implements the AI, it incentivizes the agents to costly express negative emotions higher than their natural intensities, causing the "rat race" among agents. Figure 5.3 illustrate the losses for the firm, agent, and society as a function of β. The numerical setting is $\mu_\gamma = \mu_\eta = 1, \sigma_\gamma = 1, \sigma_\eta = 0.1, \rho = 0.5, m = 0.1, \lambda = 0.5$, and $\sigma_e = 0.1$. Notably, the figure shows $\beta^* > \beta_s^* > 0$.

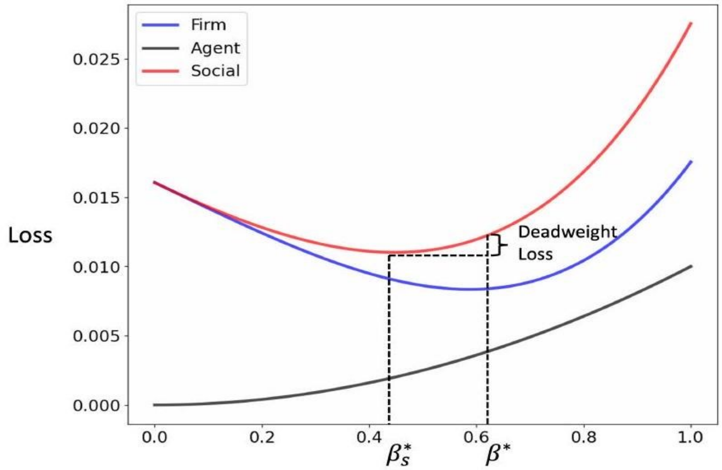


Figure 5.3 Loss Functions for the Firm, Agent, and Society

It implies that the public policymaker may help the market achieve a Pareto optimal point by first regulating the firm to underutilize the agent's emotional expression data and increase the agent's expected utility. Then, they can redistribute part of the agent's increased utility to the firm, for example, by taxing the agents and subsiding the firm. To achieve a Pareto optimal, the policy should make the firm at least indifferent between implementing a higher β and implementing β_s^* with receiving a subsidy.

It is worth noting that when $\lambda < \lambda_0$, meaning that the spillover loss to the firm is relatively less important than the misallocation loss, implementing AI is not only the firm's best strategy but the socially optimal choice. This is because the social loss is formed by adding a second-order agent's loss due to β, and the firm's benefit remains in the first order. Again, this condition holds regardless of the bias, variance, or distribution of the noise in the AI system.

However, the variance of the noise may have a marginal effect on the agent's, the firm's, and the social welfare. Intuitively, if the firm decides not to implement the emotion AI ($\beta^* = 0$) and it is also the socially optimal point, then the variance of the noise has no impact on the welfare. I consider the case when $\beta^* > 0$ and σ_e^2 increases marginally, that is, suppose that a slightly weaker AI system is implemented. I provide some intuitions about the welfare change before providing a formal proposition. First, as I show in Proposition 1, the firm's loss due to the variance of the noise is $\beta^2\sigma_e^2$. Intuitively, σ_e^2 serves as a regularization term for β and, therefore, marginally increasing σ_e leads to a smaller β^* (underutilizing data). Because the greater variance of the noise makes the firm commit to underutilizing the emotional expression data, it reduces the agents' gaming behavior. From the agent's perspective, the need for the cost signaling ("rat race") reduces, and the agent's expected utility increases. Again, this impact is heterogeneous among agents. A greater variance of the noise, and thus a smaller β^*, will benefit the lower type agents (those with smaller η and γ). If the error is individual-specific and algorithmic bias exists among different agents, the bias tends to be smaller because the algorithmic bias term is linear in β^*.

Second, the firm's welfare change is less obvious. Although the marginal increase in σ_e tends to increase the firm's loss through the term $\beta^2\sigma_e^2$ as I show Proposition 1, it enables the firm to credibly underutilize the data by implementing a smaller β^*. Further, such commitment reduces all agents' gaming behavior $m\beta\gamma$. As all agents are more likely to reveal true types, the firm's loss due to resource misallocation may be reduced. Moreover, the firm's loss due to the spillover effect of negative emotional expression will be reduced because all agents are expressing a lower level of emotional expression x. Therefore, one may expect the firm's loss to decrease by implementing a weaker AI. Indeed, in a different setting where the firm's allocation y is binary and the agent has to invest a fixed amount of cost (as the sunk cost) before entering the game, the noise may benefit the firm by reducing the agents' strategic behavior (Braverman & Garg, 2020). Counterintuitively, in my setting where the allocation is allowed to be continuous, and the agent's cost varies with her decision to express emotions, the firm's loss always marginally increases in the variance of the noise σ_e, despite the benefit of receiving more informative signals and receiving less negative spillover loss. This is partly because even though the signals are more informative, the firm has to implement a larger β to realize the benefit of screening. Further, because the firm has taken the spillover-misallocation tradeoff into account when implementing the optimal β^*, forcing the firm to decrease β^* can make the firm worse off.

Third, because the increased variance of the noise increases the firm's expected loss and decreases the agent's expected loss, it is intriguing to understand the impact on the expected social loss. Intuitively, one may expect the increased variance of the noise to decrease social welfare because it increases the information loss in agent-firm communication. If there is no gaming behavior, the market efficiency strictly decreases due to the increased σ_e^2. When gaming behavior exists, counterintuitively, I show that sometimes marginally increasing the variance of the noise can increase social welfare. This is because the gaming behavior causes communication inefficiency, and the increased noise unexpectedly regularizes such inefficiency. But the increased variance of the noise naturally has the negative side effect of undermining allocation efficiency. Therefore, the marginal impact of σ_e^2 on social welfare could be either positive or negative.

The following proposition formally verifies the above discussions. More importantly, it quantifies when the increased σ_e^2 positively or negatively impacts social welfare. I show that the direction depends on the curvature of the firm's loss function evaluated at the firm's optimal policy point β^* and a threshold that takes the simple form of $4m^2\mu_\gamma$ (or $s^2\mu_\gamma$).

Proposition 3. If $\beta^* > 0$ is the unique equilibrium, marginally increasing σ_e^2 leads to (1) decreased β^*, (2) decreased agent's expected loss, (3) increased firm's loss, and (4) either increased or decreased expected social loss, the direction of which depends on sign $\left(L''(\beta^*) - 4m^2\mu_\gamma\right)$.

Remark 9. Given $\beta^* > 0$ is the unique equilibrium, if $m \to 0, L''(\beta^*) - 4m^2\mu_\gamma > 0$; if $m \to \infty, L''(\beta^*) - 4m^2\mu_\gamma < 0$. If $\mu_\gamma \to \infty$, sign $\left(L''(\beta^*) - 4m^2\mu_\gamma\right) =$ sign $\left(\lambda\mu_\gamma - 2\right)$.

The proofs of Proposition 3 and the associated remark are detailed in Appendix C.3. Remark 9 indicates that, for a context with a small stake, a stronger AI is more beneficial to society, whereas, for a context with a large stake, a weaker AI is more beneficial. Notably, the proof is not obvious because $L''(\beta^*)$ is also a function of m and μ_γ. When m is sufficiently small, because the stakes of escalating emotional expression levels are small, the agents tend to reveal true types of η. Therefore, the benefit to the firm of implementing a stronger AI to screen the agents' types is greater than the agent's expected loss of being screened. When m is sufficiently large, the gaming behavior due to the high stakes is salient, which incurs a high signaling cost to all agents. Therefore, marginally increasing σ_e^2 reduces agents' "rat race" and increases their utility. The social welfare increases even though increased σ_e^2 reduces the firm's utility.

One might expect that if μ_γ is sufficiently large, I must have $L''(\beta^*) - 4m^2\mu_\gamma < 0$. However, Remark 9 indicates that this may not always be the case. It holds when $\lambda = 0$. This is because the firm chooses an optimal β_0^* to correct the effect of μ_γ in its allocation. Therefore, μ_γ can only affect the firm's loss due to the spillover effect and its coefficient λ. When β^* decreases, if $\lambda = 0$, the firm's increased loss would be irrelevant to μ_γ. However, the agent's loss will significantly decrease when μ_γ is large. When μ_γ is sufficiently large, the firm's loss will be smaller than the agent's expected gain, which results in an increase in social welfare. However, if $\lambda > 0$, social welfare can either increase or decrease. From $\beta^* > 0$ and Lemmas 1 and 2, I know

$\lambda < \max\{\lambda_0, \lambda_1\}$. As $\mu_\gamma \to \infty$, holding other parameters constant, I have $\max\{\lambda_0, \lambda_1\} \to 0_+$. This implies that the equilibrium holds only when λ is a small value. It can be shown that the upper bound of $\lambda\mu_\gamma$ depends on other parameters including μ_η, m, and σ_η^2. The upper bound may or may not exceed the threshold of 2.

Remark 10. If the firm is indifferent from $\beta^* = 0$ and a unique $\beta_+^* > 0$ but implements $\beta_+^* > 0$, a marginal increase in σ_e^2 leads to the equilibrium with $\beta^* = 0$. The firm's expected loss remains unchanged, whereas the agent's expected utility strictly increases. Therefore, social welfare strictly increases. Note that although the increase in σ_e^2 is marginal, the increase in the agent's and social welfare is not (equal to $m^2\beta_+^{*2}\mu_\gamma$).

5.6 Adopting Human or AI Service Systems

As I have discussed the differences between human and AI systems in Section 2, I formally highlight these differences in Table 5.2. First, human employees receive the true signal x. Second, due to employees' potential emotional responses to agents' emotional expressions, employees may modify the allocation rule set by the firm. I assume the policy parameter implemented by the employees to be a random variable $\beta_h \triangleq \beta + \alpha_h$ where $\alpha_h \sim \mathcal{G}_1(0, \sigma_\alpha^2)$ captures employees' heterogeneity in their behavioral responses to a given level of negative emotional expressions. The allocation y, in this case, is $(\beta + \alpha_h)x + \beta_0$. $\alpha_h > 0$ indicates over-allocation, and $\alpha_h < 0$ indicates under-allocation compared to the firm's policy. It is reasonable to assume $\mathbb{E}[\alpha_h] = 0$ because the firm would maintain a customer service team to implement the unbiased policy. If the team, on average, keeps over- or under-compensating the agents, the firm would earn a negative profit and exit the market in the long run. I assume that the variance of α_h is larger than zero ($\sigma_\alpha^2 > 0$) because emotional responses have psychological roots in human nature and cannot be completely removed (Y. Yu et al., 2020, 2023). Further, α_h is a characteristic of employees' personalities and is independent of agents' types (η, γ). Third, in addition to spillover costs to other agents (related to λ), negative emotions negatively affect employees' psychological well-being and incur a cost to the firm. I denote the heterogeneous spillover costs on employees in a single interaction as a positive random variable $\kappa_h \sim \mathcal{G}_2$ with mean μ_κ and variance σ_κ^2. For generalizability, I do not assume any functional form for $\mathcal{G}_1$ and $\mathcal{G}_2$.

Table 5.2 Comparing Human and AI Service Systems

Features	Human service	AI service
Signal received	x	$x+e, e \sim \mathcal{F}(\mu_e, \sigma_e^2)$
Allocation rule implemented	$\beta_h = \beta + \alpha_h, \alpha_h \sim \mathcal{G}_1(0, \sigma_\alpha^2)$	β
Spillover costs	$\lambda + \kappa_h, \kappa_h \sim \mathcal{G}_2(\mu_\kappa, \sigma_\kappa^2)$	λ

5.6.1 *Analysis of Human Service Systems*

Given the firm's policy $Y(x) = \beta x + \beta_0$, an employee's execution is $Y_h(x) = (\beta + \alpha_h)x + \beta_0$. The agent's problem of solving the optimal response x is:

$$x = \arg\max \mathbb{E}_{\alpha_h}[2mY_h(x) - (x-\eta)^2/\gamma].$$

From the first-order condition I have $x = \eta + m(\beta + \mathbb{E}[\alpha_h])\gamma = \eta + m\beta\gamma$, which is the same as the optimal x when the firm uses the AI systems (Equation 3). Intuitively, a rational agent will not change her decision because she knows that the average payoff for a given x is the same, i.e., $\mathbb{E}[\beta x] = \mathbb{E}[(\beta + \alpha_h)x]$. Given the agent's best response, the firm chooses an optimal pair of policy parameters (β, β_0) to minimize its loss denoted as L_h. I have the following proposition regarding decomposing the loss function (see proof in Appendix C.4).

Proposition 4. Under the optimal β_0, the firm's loss function using human service systems, L_h, can be written as a function of β :

$$L_h(\beta) = \underbrace{(\lambda + \mu_\kappa + \sigma_\alpha^2)\mathbb{E}[x^2]}_{\text{Spillover and execution loss}} + \underbrace{\mathbb{E}[(\mathbb{E}[\eta \mid x] - \eta)^2]}_{\text{Loss of estimating } \eta \text{ given the true } x} + \underbrace{\mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])^2]}_{\text{Misallocation loss given estimation}}, \quad (9)$$

where $Y_0(x) = \beta x + \mathbb{E}[\eta - \beta x]$.

Comparing to $L(\beta)$ in Proposition 1, $L_h(\beta)$ does not contain loss due to algorithmic noise but has two more loss terms. First, $\sigma_\alpha^2 \mathbb{E}[x^2]$ is the loss due to employees' behavioral responses to negative emotions. Second, $\mu_\kappa \mathbb{E}[x^2]$ is the negative emotions' spillover effect on employees' psychological well-being. Notably, although α_h and κ_h root in different psychological mechanisms, $\sigma_\alpha^2 = \mathbb{E}[\alpha_h^2]$ and $\mu_\kappa = \mathbb{E}[\kappa_h]$ have similar impacts on the firm's expected loss. I denote $\lambda_h \triangleq \mu_\kappa + \sigma_\alpha^2$ to capture the loss for a given level of x and specific to using the human service systems.

Denote $\lambda_{1h} \triangleq \frac{2m\text{Cov}(\eta,\gamma)-\text{Var}(\eta)}{m^2\mathbb{E}[\gamma^2]}$ and $\beta_h^* = \arg\min_{\beta\geq 0} L_h(\beta)$. Recall that $\lambda_0 = \frac{\text{Var}(\eta)}{m\mathbb{E}[\eta\gamma]}$. The following propositions hold, which characterize the equilibria of using human service systems:

Proposition 5. For $\rho \in (-1,0)$, (1) $\beta_h^* = 0$ if $\lambda + \lambda_h \geq \lambda_0$; (2) $\beta_h^* \in (0,2)$ and unique if $\lambda + \lambda_h < \lambda_0$.

Proposition 6. For $\rho \in [0,1)$, (1) When $\lambda_0 \geq \lambda_{1h}$, (1a) $\beta_h^* = 0$ if $\lambda + \lambda_h \geq \lambda_0$; (1b) $\beta_h^* \in (0,1)$ and unique if $\lambda + \lambda_h < \lambda_0$; (2) When $\lambda_0 < \lambda_{1h}$, there exists a unique $\lambda_{0h}^* \in (\lambda_0, \lambda_{1h})$ such that (2a) $\beta_h^* = 0$ if $\lambda + \lambda_h > \lambda_0^*$; (2b) $\beta_h^* \in (0,1)$ and unique if $\lambda + \lambda_h < \lambda_0^*$; (2c) there exist two equilibria, $\beta_h^* = 0$ and a unique $\beta_h^* \in (0,1)$, if $\lambda + \lambda_h = \lambda_0^*$.

Due to Propositions 1 and 4, Propositions 5 and 6 are the special cases of Lemmas 1 and 2 when Var $(e) = 0$ and the spillover parameter λ is replaced by $\lambda + \lambda_h$. Therefore, my discussion on Lemmas 1 and 2 is also applicable for the two propositions.

5.6.2 *Conditions of AI Outperforming Human Systems*

I am interested in the conditions under which the costs of using AI service systems are lower than or equal to the costs of using human systems. The sufficient and necessary condition is:

$$\Delta L_h = L_h(\beta_h^*) - L(\beta^*) \geq 0$$

Given L_h and L, one can numerically solve β^* and β_h^* and calculate ΔL_h. However, the close form of this condition is complicated and uninterpretable because β^* and β_h^* can both be a root of a cubic equation. To derive more managerial implications, I characterize sufficient conditions for $\Delta L_h \geq 0$. Let $I(\cdot)$ be the indicator function. Denote

$$\overline{\lambda_h} = \begin{cases} \lambda_{0h}^* \in (\lambda_0, \lambda_{1h}), & \text{if } \rho \geq 0 \text{ and } \lambda_0 < \lambda_{1h} \\ \lambda_0, & \text{otherwise} \end{cases}$$

Proposition 7. $\Delta L_h \geq 0$ holds if (1) $\lambda_h \geq \overline{\lambda_h} - \lambda$ or (2) $\frac{\sigma_e^2}{\lambda_h} \leq \mathbb{E}\left[\left(\frac{\eta}{1+I(\rho<0)} + m\gamma\right)^2\right]$.

The proof of Proposition 7 is provided in Appendix C.4. The first part of Proposition 7 implies that if λ_h is sufficiently large, regardless of the algorithmic noise e, the human systems can not outperform the AI systems. $\lambda_h \geq \overline{\lambda_h} - \lambda$ implies that $\beta_h^* = 0$. In this case, it can be either $\beta^* = 0$ or $\beta^* > 0$. Intuitively, when the stakes of employee-specific loss due to agents' negative emotional expressions (λ_h) are high, the negative emotional expressions are liabilities rather than signals to firms that implement human service systems. In this case, the AI system is preferred simply because negative emotions cannot affect machines.

The second part of Proposition 7 further implies that if it is profitable for the firm to ask employees to handle negative emotional expressions (λ_h is not too high so that $\beta_h^* > 0$), the firm needs to consider both the stakes of employee-specific loss (λ_h) and the variance of the algorithmic noise (σ_e^2). In particular, the proposition implies that the firm needs to compare the ratio between σ_e^2 and λ_h with an upper threshold determined by the exogenous market parameters (ρ, η, γ, m). Notably, the second part of Proposition 7 directly implies the following proposition (see proof in Appendix C.4), which indicates that AI systems are robust to extreme situations:

Proposition 8. Holding other parameters constant, AI systems are preferred if any one of $(m, \mu_\gamma, \sigma_\gamma, \mu_\eta, \sigma_\eta)$ is sufficiently large.

Intuitively, one would expect that human service systems are better when the allocation results are of high stakes (large m) to agents (e.g., customer service for high-involvement goods). For example, according to my interviews with several retailing managers from China, most managers are not confident in using AI to handle customers' negative emotions, especially when customer service stakes are high. However, many large companies in the U.S. delivering high-involvement services, including Bank of America and United Parcel Service, have adopted AI systems as the first interface to customer complaints. In addition to labor cost considerations, my result provides another rationale for using AI to respond to agents' negative emotions. More interestingly, even if the stakes for the agents are low, human systems may not necessarily outperform AI systems. In particular, I have a sufficient and necessary condition for preferring AI systems (see proof in Appendix C.4): Proposition 9. When $m \to 0, \Delta L_h \geq 0 \Leftrightarrow \lambda_h \geq \frac{\sigma_e^2}{(\sigma_\eta - \mu_\eta \sigma_e / \sigma_\eta)^2 + (\sigma_e + \mu_\eta)^2}$.

Remark 11. The above proposition implies that when $m \to 0$, the AI systems are preferred if (1) $\sigma_\eta \to 0$, (2)$\sigma_\eta \to \infty$, or (3)$\sqrt{\lambda_h} \geq \frac{\sigma_e}{\sigma_e + \mu_\eta}$.

Remark 12. When $m \to 0, \beta_h^* \to 1$ and $\beta^* \to \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_e^2} < 1$. Adopting AI systems increases the agent's expected utility compared to human service systems.

The impact of the heterogeneity in agents' natural intensities (σ_η) is nonlinear. If the heterogeneity is low, the AI systems can allocate resources well (e.g., simply setting the allocation near μ_η and the misallocation loss would be minimal) and prevent employees from exposing to negative emotions. If the heterogeneity is high, screening benefits dominate the misallocation loss. According to the above remark, both β^* and β_h^* are close to one, implying that the firm's optimal strategy is to almost fully utilize emotional expression data without discounting the part caused by gaming behavior. The loss due to algorithmic noise ($\beta^{*2}\sigma_e^2$) to the firm is smaller than the employee-specific loss ($\lambda_h \mathbb{E}[(\eta + m\beta_h^*\gamma)^2]$). Formally, $\lambda_h \mathbb{E}[(\eta + m\beta_h^*\gamma)^2] > \lambda_h \sigma_\eta^2$ and $\lambda_h \sigma_\eta^2$ will be larger than $\beta^* \sigma_e^2$ when σ_η is sufficiently large. Finally, when μ_η is sufficiently large, Condition (3) in Remark 11 holds. It implies that when agents express excessive negative emotions on average, AI systems are preferred because the employee-specific loss due to agents' negative emotions is unacceptably high.

5.7 CONCLUSION

I developed an analytic model to identify the conditions under which the firm should adopt emotion AI to automate emotion recognition and resource allocation processes for strategic individuals. I find that emotion AI is always valuable to the firm if the spillover effect of negative emotional expressions is negligible compared to the resource misallocation loss, regardless of algorithmic noise and gaming behavior. I provide optimal linear allocation policies for the firm after it adopts emotion AI. I also investigate the welfare impact of adopting emotion AI on agents, the firm, and society. Notably, I show that a stronger AI is not always socially desirable. Finally, I characterize a broad set of conditions under which it is profitable for the firm to use AI systems instead of hiring human employees to detect emotions and allocate resources.

5.7.1 *Theoretical Implications*

This work adds to the literature on emotion AI regarding its design, application, and regulation (McStay, 2018; Han et al., 2022; Y. Yu et al., 2023). Emotion AI is valuable under a broad set of market conditions despite its noise in emotion detection (Barrett et al., 2019). Emotion AI developers should pay attention to the variance of detection noise more than its biasedness because economic mechanisms can be designed to correct the bias at the resource allocation stage (Qiao & Huang, 2021; Chernozhukov et al., 2022). Underutilizing emotional expression data can not only protect individuals' privacy (Crawford et al., 2021) but also increase social welfare. The firm has an incentive to sometimes underutilize the data to mitigate agents' gaming behavior and its associated negative spillover effect. But the public policymaker should further regulate the data utilization to achieve social optimum.

This work adds to the literature on human-AI collaboration (Acemoglu, 2021). Notably, I do not argue that AI should replace employees in the entire process of emotional interaction between firms and agents. Instead, I highlight that emotion AI can automate the process of emotion recognition and data-driven decision-making before a human employee takes over the task under many conditions. The time and effort of employees (including customer service representatives and doctors) are considered one of the firm's most important and limited resources. The economic value of emotion AI lies in reducing the waste of limited resources, protecting employees from excessive negative emotions (Causon, 2021), and mitigating their psychological stress (Poddar & Madupalli, 2012), thereby increasing the firm's profits.

This work extends the framework of muddled information (Frankel & Kartik, 2019, 2022) where agents' private types have two-dimensional heterogeneity in a signaling game, and information about the dimension of interest is muddled by the other dimension. First, I incorporate noise in signaling into the framework. I highlight the unexpected impacts of noise in equilibria and welfare. This work is broadly related to the work that argues algorithmic noise may increase welfare in strategic classification (Braverman & Garg, 2020) but in a very different setting. Second, I consider the spillover effect of agents' actions and find it a major challenge for the economic value of emotion AI. Finally, I added employees' behavior as an additional layer to the framework structure. I consider the case where employees' behavior responses to negative emotions add variance in the allocation rule.

5.7.2 *Managerial Implications*

This work provides managerial implications for firms and managers. First, managers should be aware that a naive emotion AI system can suffer from gaming behavior and unexpectedly increase a firm's loss. Therefore, emotion AI adoption should be accompanied by optimal allocation policy design. Second, when agents' emotional expression has a negative spillover effect, for example, when the firm-agent interaction is through a public channel, the value of adopting emotion AI may be undermined. The firm can investigate the market parameters and leverage my analytical framework, particularly Lemmas 1 and 2 to determine whether emotion AI adoption is profitable. Third, if the spillover effect is negligible, adopting emotion AI is an optimal choice for both the firm and society. The firm needs to incorporate algorithmic noise and gaming behavior considerations into its allocation policy design. The most practical and interpretable mechanisms, linear allocation policies, work well in separating agents of different types because the slope of the optimal linear policy captures the key factor in data-driven policy design: the extent to which the firm should underutilize or overutilize the emotional expression data. The decision to underutilize or overutilize data depends on the market composition (i.e., the sign of the correlation between natural intensities and gaming abilities). The intercept of the optimal linear policy corrects the AI biasedness in emotion recognition. Further, other than one special case, the optimal linear policy would lead to a stable market equilibrium that is robust to gaming behavior. Fourth, although emotional expression data is more informative (contains less gaming behavior) when the variance of algorithmic noise is larger, from the firm's perspective, an emotion recognition algorithm with minimum variance in its detection error is preferred. Fifth, considering employees' behavioral responses in emotion recognition and resource allocation, human service systems may not outperform AI systems, regardless of scalability. When employee-specific loss, agents' stakes over allocation, natural intensities, or gaming behavior are of sufficiently high levels, AI systems are preferred. Further, even if agents' stakes and employee-specific loss are low, human systems do not outperform AI systems when the heterogeneity in natural intensities is either minimal or sufficiently large.

These results also provide implications for regulators and public policymakers. First, if the firm is unwilling to adopt emotion AI, promoting AI adoption cannot increase social welfare. Second, if the firm chooses to adopt emotion AI, it has the incentive to overutilize emotional expression data in its allocation policy compared to the socially optimal level. Therefore, the firm's data utilization should be regulated. For example, regulators may allow the firm only to access part of the individuals' emotional expression data. Third, the firm always seeks to minimize the variance of algorithmic noise, while such a tendency may or may not increase social welfare. Although there is a concern that a weak AI may produce problematic allocation results, a strong AI may also decrease social welfare. On average, adopting emotion AI transfers utility from agents to the firm. When agents are strategic and the signals are noisy, the game is not necessarily zero-sum. Whether a weak or strong AI is socially preferred depends on the firm's cost structure (curvature of the loss function evaluated at the minimum point), the agents' stakes over the firm's allocation, and the average gaming ability.

Chapter 6. CONCLUDING REMARKS

This dissertation aimed to understand the interaction between algorithms and user behavior in various business contexts, with a focus on algorithms that can process unstructured data, such as text, images, and videos. The main research question was: How do algorithms predict, influence, and in turn are influenced by, user behavior in different business settings?

To answer this question, this dissertation adopted a mixed-methods approach, combining econometric models, deep learning algorithms, and game-theoretical models. The dissertation consisted of three empirical essays, each addressing a specific sub-question and context related to the main research question.

The first essay (Chapter 3) investigated how emotion artificial intelligence (AI) can detect discrete emotions in online reviews and how these emotions affect product sales and consumer decisions. The second essay (Chapter 4) examined how outstream video advertising can drive user attention and clicking behaviors, and how video content features influence consumer learning and product quality perception. The third essay (Chapter 5) explored how emotion AI may shape user behaviors, especially strategic behaviors, when organizations adopt it to allocate limited resources based on individuals' negative emotions.

The key findings of this dissertation are as follows:

- Emotion AI can automatically construct domain-specific emotion lexicons and detect multi-dimensional emotions in texts with high accuracy. Discrete emotions in online reviews have significant predictive power on product sales and consumer decisions, mediated by perceived processing fluency and review helpfulness.
- Outstream video ads attract consumer attention and are more effective when products are less differentiated from each other in a market. Video content features that facilitate efficient consumer learning or signal product quality significantly increase consumers' likelihood of clicking a product.
- Emotion AI may induce gaming behavior among strategic users who escalate emotional intensities to obtain more resources. Adopting emotion AI is always valuable to the organization if the spillover effect of negative emotions is negligible

compared to resource misallocation loss. However, a stronger AI is not always socially desirable and regulation on data-driven allocation is needed.

This dissertation contributes to the burgeoning field of artificial intelligence (AI) and its applications in various business contexts. It provides novel insights into the complex relationship between advanced algorithms and user behavior, as well as the economic value and social welfare implications of AI adoption. It also demonstrates the feasibility and usefulness of combining econometric models, deep learning algorithms, and game-theoretical models to analyze multi-modal data and user behaviors.

In conclusion, this dissertation has shed light on the intriguing interaction between algorithms and user behavior in different business settings. It has shown that algorithms can not only predict and understand user behavior, but also influence and be influenced by it. It has also highlighted the challenges and opportunities of adopting AI and unstructured-data-driven algorithms in business practice and policy making. This dissertation hopes to inspire more research on this important topic and to contribute to the advancement of AI and its applications in business and society.

BIBLIOGRAPHY

Akpinar, E., & Berger, J. (2017). Valuable Virality. *Journal of Marketing Research*, *54*(2), 318–330. https://doi.org/10.1509/jmr.13.0350

Amirshahi, S. A., Hayn-Leichsenring, G. U., Denzler, J., & Redies, C. (2014). Evaluating the rule of thirds in photographs and paintings. *Art & Perception*, *2*(1-2), 163–182.

Anderson, E. T., & Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, *51*(3), 249–269.

Arora, S., Liang, Y., & Ma, T. (2017). A simple but tough-to-beat baseline for sentence embeddings. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*.

Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. *Proceedings - 2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010*. https://doi.org/10.1109/WI-IAT.2010.63

Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research. Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, *8*(1), 10–14. https://doi.org/10.1111/1467-8721.00003

Benway, J. P. (1998). Banner blindness: The irony of attention grabbing on the World Wide Web. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(5), 463–467.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(1), 281–305.

Biswas, A., Bhutani, M., & Sanyal, S. (2017). Mrnet-product2vec: A multi-task recurrent neural network for product embeddings. *Joint European Conference on Machine Learning and*

Knowledge Discovery in Databases, 153–165.

Blei, D. M., Edu, B. B., Ng, A. Y., Edu, A. S., Jordan, M. I., & Edu, J. B. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993

Blunch, N. J. (1984). Position bias in multiple-choice questions. *Journal of Marketing Research*, *21*(2), 216–220.

Bordwell, D., Thompson, K., & Smith, J. (1993). *Film art: An introduction* (Vol. 7). McGraw-Hill New York.

Braverman, M., & Garg, S. (2020). The role of randomness and noise in strategic classification. *Leibniz International Proceedings in Informatics, LIPIcs*, *156*. https://doi.org/10.4230/LIPIcs.FORC.2020.9

Campagna, R. L., Mislin, A. A., Kong, D. T., & Bottom, W. P. (2016). Strategic consequences of emotional misrepresentation in negotiation: The blowback effect. *Journal of Applied Psychology*, *101*(5), 605–624. https://doi.org/10.1037/apl0000072

Campbell, C., Mattison Thompson, F., Grimm, P. E., & Robson, K. (2017). Understanding Why Consumers Don’t Skip Pre-Roll Video Ads. *Journal of Advertising*, *46*(3), 411–423. https://doi.org/10.1080/00913367.2017.1334249

Cao, L. (2010). Domain-driven data mining: Challenges and prospects. *IEEE Transactions on Knowledge and Data Engineering*, *22*(6), 755–769. https://doi.org/10.1109/TKDE.2010.32

Causon, J. (2021). *Open letter calling for action to end abuse of public-facing workers*. The Institute of Customer Service. https://www.instituteofcustomerservice.com/media/open-letter-dec-21/

Cetin, U., & Bingol, H. O. (2014). Attention competition with advertisement. *Physical Review E*, *90*(3), 32801.

Chen, F., Liu, X., Proserpio, D., Troncoso, I., & Xiong, F. (2020). Studying product competition using representation learning. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1261–1268.

Colon, G. (2016). *Disruptive Marketing: What Growth Hackers, Data Punks, and Other Hybrid Thinkers Can Teach Us About Navigating the New Normal*. AMACOM.

Comanor, W. S., & Wilson, T. A. (1979). The effect of advertising on competition: A survey. *Journal of Economic Literature*, *17*(2), 453–476.

Crawford, K. (2021). Time to regulate AI that interprets human emotions. In *Nature* (Vol. 592, Issue 7853, p. 167). https://doi.org/10.1038/d41586-021-00868-5

Deke, J. (2014). *Using the linear probability model to estimate impacts on binary outcomes in randomized controlled trials*.

der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11).

Desimone, J. A., Harms, P. D., & Desimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*(2), 171–181. https://doi.org/10.1002/job.1962

Desmet, P. (2010). Three levels of product emotion. *Proceedings of Kansei Engineering and Emotion Research International Conference*.

East, R., Romaniuk, J., Chawdhary, R., & Uncles, M. (2017). The impact of word of mouth on intention to purchase currently used and other brands. *International Journal of Market Research*, *59*(3), 321–334. https://doi.org/10.2501/IJMR-2017-026

Eid, M., & Diener, E. (2001). Norms for experiencing emotions in different cultures: Inter- and intranational differences. *Journal of Personality and Social Psychology*, *81*(5), 869–885. https://doi.org/10.1037/0022-3514.81.5.869

Ekelund, R. B., Mixon, F. G., & Ressler, R. W. (1995). Advertising and information: an empirical study of search, experience and credence goods. *Journal of Economic Studies*.

Felbermayr, A., & Nanopoulos, A. (2016). The Role of Emotions for the Perceived Usefulness in Online Customer Reviews. *Journal of Interactive Marketing*, *36*, 60–76. https://doi.org/10.1016/j.intmar.2016.05.004

Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information and Management*, *55*(8), 956–970. https://doi.org/10.1016/j.im.2018.04.010

Fong, N. M. (2017). How targeting affects customer search: A field experiment. *Management Science*, *63*(7), 2353–2364.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural model with unobserved variables and measurement errors. *Journal of Marketing Research*, *18*(1), 39–50.

Fortin, D. R., & Dholakia, R. R. (2005). Interactivity and vividness effects on social presence and involvement with a web-based advertisement. *Journal of Business Research*, *58*(3), 387–396.

Frankel, A., & Kartik, N. (2019). Muddled information. *Journal of Political Economy*, *127*(4), 1739–1776. https://doi.org/10.1086/701604

Frankel, A., & Kartik, N. (2022). Improving Information from Manipulable Data. *Journal of the European Economic Association*, *20*(1), 79–115. https://doi.org/10.1093/jeea/jvab017

Fulmer, I. S., Barry, B., & Long, D. A. (2009). Lying and smiling: Informational and emotional deception in negotiation. *Journal of Business Ethics*, *88*(4), 691–709. https://doi.org/10.1007/s10551-008-9975-x

Geva, T., Oestreicher-Singer, G., Efron, N., & Shimshoni, Y. (2017). Using Forums and Search for Sales Prediction of High-Involvement Products. *MIS Quarterly*, *41*(1).

Goodrich, K., Schiller, S. Z., & Galletta, D. (2015). Consumer reactions to intrusiveness of online-video advertisements do length, informativeness, and humor help (or hinder) marketing outcomes? *Journal of Advertising Research*, *55*(1), 37–50. https://doi.org/10.2501/JAR-55-1-037-050

Grbovic, M., & Cheng, H. (2018). Real-time personalization using embeddings for search ranking at airbnb. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 311–320.

Han, E., Yin, D., & Zhang, H. (2022). Bots with Feelings: Should AI Agents Express Positive Emotion in Customer Service? *Information Systems Research*. https://doi.org/10.1287/isre.2022.1179

Hardt, M., Megiddo, N., Papadimitriou, C., & Wootters, M. (2016). Strategic classification. *ITCS 2016 - Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, 111–122. https://doi.org/10.1145/2840728.2840730

Havlena, W. J., & Holbrook, M. B. (1986). The Varieties of Consumption Experience: Comparing Two Typologies of Emotion in Consumer Behavior. *Journal of Consumer Research*, *13*(3), 394–394. https://doi.org/10.1086/209078

Heckler, S. E., & Childers, T. L. (1992). The Role of Expectancy and Relevancy in Memory for Verbal and Visual Information: What is Incongruency? *Journal of Consumer Research*, *18*(4), 475. https://doi.org/10.1086/209275

Hennig-Thurau, T., Wiertz, C., & Feldhaus, F. (2015). Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. *Journal of the*

Academy of Marketing Science, *43*(3), 375–394. https://doi.org/10.1007/s11747-014-0388-3

Hickey, B. A., Chalmers, T., Newton, P., Lin, C. T., Sibbritt, D., McLachlan, C. S., Clifton-Bligh, R., Morley, J., & Lal, S. (2021). Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review. In *Sensors* (Vol. 21, Issue 10). https://doi.org/10.3390/s21103461

Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, *355*(6324). https://doi.org/10.1126/science.aal3856

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior*, *48*, 17–27. https://doi.org/10.1016/j.chb.2015.01.010

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, *21*(2), 155–172.

James, W. (2004). The principles of psychology, Vol I. In *The principles of psychology, Vol I.* https://doi.org/10.1037/10538-000

Jiang, L., Yin, D., & Liu, D. (2019). Can joy buy you money? The impact of the strength, duration, and phases of an entrepreneur's peak displayed joy on funding performance. *Academy of Management Journal*, *62*(6), 1848–1871. https://doi.org/10.5465/amj.2017.1423

Joa, C. Y., Kim, K., & Ha, L. (2018). What Makes People Watch Online In-Stream Video Advertisements? *Journal of Interactive Advertising*, *18*(1), 1–14. https://doi.org/10.1080/15252019.2018.1437853

Kalamas, M., Laroche, M., & Makdessian, L. (2008). Reaching the boiling point: Consumers' negative affective reactions to firm-attributed service failures. *Journal of Business Research*, *61*(8), 813–824. https://doi.org/10.1016/j.jbusres.2007.09.008

Kartik, N., Ottaviani, M., & Squintani, F. (2007). Credulity, lies, and costly talk. *Journal of Economic Theory*, *134*(1), 93–116. https://doi.org/10.1016/j.jet.2006.04.003

Kenter, T., Borisov, A., & De Rijke, M. (2016). Siamese CBOW: Optimizing word embeddings for sentence representations. *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, *2*, 941–951. https://doi.org/10.18653/v1/p16-1089

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

Kim, J., Guo, P. J., Cai, C. J., Li, S.-W., Gajos, K. Z., & Miller, R. C. (2014). Data-driven interaction techniques for improving navigation of educational videos. *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, 563–572.

Kim, J., & Gupta, P. (2012). Emotional expressions in online user reviews: How they influence consumers' product evaluations. *Journal of Business Research*, *65*(7), 985–992. https://doi.org/10.1016/j.jbusres.2011.04.013

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, *27*(4), 435–454.

Klauer, K. C. (1997). Affective Priming. *European Review of Social Psychology*, *8*(1), 67–103. https://doi.org/10.1080/14792779643000083

Kong, S., Huang, Z., Scott, N., Zhang, Z., & Shen, Z. (2019). Web advertisement effectiveness evaluation: Attention and memory. *Journal of Vacation Marketing*, *25*(1), 130–146.

Krages, B. (2012). *Photography: the art of composition*. Simon and Schuster.

Krishnamoorthy, S. (2015). Linguistic features for review helpfulness prediction. *Expert Systems with Applications*, *42*(7), 3751–3759. https://doi.org/10.1016/j.eswa.2014.12.044

Kumar Madupalli, R., & Poddar, A. (2014). Problematic customers and customer service employee retaliation. *Journal of Services Marketing*, *28*(3), 244–255. https://doi.org/10.1108/JSM-02-2013-0040

Kusumasondjaja, S., Shanka, T., & Marchegiani, C. (2012). Credibility of online reviews and initial trust: The roles of reviewer's identity and review valence. *Journal of Vacation Marketing*, *18*(3), 185–195.

Lazarus, R. S. (1991). Cognition and Motivation in Emotion. *American Psychologist*, *46*(4), 352–352. https://doi.org/10.1037/0003-066X.46.4.352

Lee, K., Choi, J., & Li, Y. J. (2014). Regulatory focus as a predictor of attitudes toward partitioned and combined pricing. *Journal of Consumer Psychology*, *24*(3), 355–362. https://doi.org/10.1016/j.jcps.2014.01.001

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science*, *15*(5), 337–341. https://doi.org/10.1111/j.0956-7976.2004.00679.x

Levinson, J. (2011). *The Best of Guerrilla Marketing: Guerrilla Marketing Remix*. Entrepreneur Press.

Li, Y., & Xie, Y. (2020). Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement. *Journal of Marketing Research*, *57*(1), 1–19. https://doi.org/10.1177/0022243719881113

Lin, Y., Yao, D., & Chen, X. (2021). Happiness Begets Money: Emotion and Engagement in Live Streaming. *Journal of Marketing Research*, *58*(3), 417–438. https://doi.org/10.1177/00222437211002477

Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, *70*(3), 74–89. https://doi.org/10.1509/jmkg.70.3.74

Loewenstein, G. (2000). Emotions in Economic Theory and Economic Behavior. *American Economic Review*, *90*(2), 426–432. https://doi.org/10.5815/ijigsp.2012.01.06

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

Malik, M. S. I., & Hussain, A. (2017). Helpfulness of product reviews as a function of discrete positive and negative emotions. *Computers in Human Behavior*, *73*(1), 290–302. https://doi.org/10.1016/j.chb.2017.03.053

McStay, A. (2018). *Emotional AI: The rise of empathic media*. Sage.

McStay, A. (2020). Emotional AI: The Rise of Empathic Media. In *Emotional AI: The Rise of Empathic Media*. https://doi.org/10.4135/9781526451293

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 1–12.

Miniard, P. W., Bhatla, S., Lord, K. R., Dickson, P. R., & Unnava, H. R. (1991). Picture-based persuasion processes and the moderating role of involvement. *Journal of Consumer Research*, *18*(1), 92–107.

Minka, T. (2000). Automatic choice of dimensionality for PCA. *Advances in Neural Information Processing Systems*, *13*, 598–604.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. *Management Information Systems*, *34*(1), 185–200. https://doi.org/10.2307/20721420

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, *78*(2), 311–329.

Ngo-Ye, T. L., & Sinha, A. P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, *61*(1), 47–58. https://doi.org/10.1016/j.dss.2014.01.011

Nguyen, H., Calantone, R., & Krishnan, R. (2020). Influence of social media emotional word of mouth on institutional investors' decisions and firm value. *Management Science*, *66*(2), 887–910. https://doi.org/10.1287/mnsc.2018.3226

O'Donnell, K., & Cramer, H. (2015). People's perceptions of personalized ads. *Proceedings of the 24th International Conference on World Wide Web*, 1293–1298.

Oliveira, N., Cortez, P., & Areal, N. (2016). Stock market sentiment lexicon acquisition using microblogging data and statistical measures. *Decision Support Systems*, *85*(1), 62–73. https://doi.org/10.1016/j.dss.2016.02.013

Ott, M., Cardie, C., & Hancock, J. (2012). Estimating the prevalence of deception in online review communities. *Proceedings of the 21st International Conference on World Wide Web*, 201–210.

Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *ACL-05 - 43rd Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*. https://doi.org/10.3115/1219840.1219855

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1–24). Springer.

Picard, R. W. (2003). Affective computing: Challenges. *International Journal of Human Computer Studies*, *59*(1), 55–64. https://doi.org/10.1016/S1071-5819(03)00052-1

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

Plutchik, R. (1984). Emotions: A General Psychoevolutionary Theory. In *Approaches to Emotion*.

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*.

Plutchik, R., & Kellerman, H. (1982). Emotions are more than subjective experiences. *Contemporary Psychology*, *27*(7), 581–581. https://doi.org/10.1037/021362

Poddar, A., & Madupalli, R. (2012). Problematic customers and turnover intentions of customer service employees. *Journal of Services Marketing*, *26*(7), 551–559. https://doi.org/10.1108/08876041211266512

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. *Information Processing and Cognition: The Loyola Symposium.*, *Informatio*, 55–85.

Quan, C., & Ren, F. (2010). A blog emotion corpus for emotional expression analysis in Chinese. *Computer Speech and Language*, *24*(4), 726–749. https://doi.org/10.1016/j.csl.2010.02.002

Rawal, P. (2013). AIDA Marketing Communication Model: Stimulating a purchase decision in the minds of the consumers through a linear progression of steps. *International Journal of Multidisciplinary Research in Social & Management Sciences*, *1*, 37–44.

Reber, R., Fazendeiro, T. A., & Winkielman, P. (2002). Processing fluency as the source of experiences at the fringe of consciousness. *Psyche*, *8*(10), 1–21.

Ripollone, J. E., Huybrechts, K. F., Rothman, K. J., Ferguson, R. E., & Franklin, J. M. (2018). Implications of the propensity score matching paradox in pharmacoepidemiology. *American Journal of Epidemiology*, *187*(9), 1951–1961.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? the effect of tweets on movie sales. *Decision Support Systems*, *55*(4), 863–870. https://doi.org/10.1016/j.dss.2012.12.022

Russell, J. A., & Mehrabian, A. (1974). Distinguishing anger and anxiety in terms of emotional response factors. *Journal of Consulting and Clinical Psychology*, *42*(1), 79–83. https://doi.org/10.1037/h0035915

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., & Nummenmaa, L. (2016). Discrete Neural Signatures of Basic Emotions.

Cerebral Cortex, *26*(6), 2563–2573. https://doi.org/10.1093/cercor/bhv086

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of "humanizing" customer service chatbots. *Information Systems Research*, *32*(3), 736–751. https://doi.org/10.1287/ISRE.2021.1015

Scherer, K. R. (2001). Appraisal Considered as a Process of Multilevel Sequential Checking. In *Appraisal processes in emotion: Theory, Methods, Research* (pp. 92–120).

Schindler, R. M., & Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour*, *11*(3), 234–243. https://doi.org/10.1002/cb.1372

Schwarz, N. (1990). Feelings as information: Informational and motivational functions of affective states. *Handbook of Motivation and Cognition*.

Sen, S., & Lerman, D. (2007). Why are you telling me this? An examination into negative consumer reviews on the web. *Journal of Interactive Marketing*, *21*(4), 76–94. https://doi.org/10.1002/dir.20090

Shi, W. (2017). *Big Data Mining of Chinese Weibo Texts: Sentiment Analysis Perspective*. China Social Sciences Press.

Shimp, T. A., & Stuart, E. W. (2004). The role of disgust as an emotional mediator of advertising effects. *Journal of Advertising*, *33*(1), 43–53. https://doi.org/10.1080/00913367.2004.10639150

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K.-C. (2020). Enhancing social media analysis with visual data analytics: A deep learning approach. *MIS Quarterly*, *44*(4), 1459–1492.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., & others. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354–359.

Simes, R. J. (1986). An improved bonferroni procedure for multiple tests of significance. *Biometrika*, *73*(3), 751–754. https://doi.org/10.1093/biomet/73.3.751

Smith, A. N., Fischer, E., & Yongjian, C. (2012). How does brand-related user-generated content differ across YouTube, Facebook, and Twitter? *Journal of Interactive Marketing*, *26*(2), 102–113.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of Cognitive Appraisal in Emotion. *Journal of*

Personality and Social Psychology, *48*(4), 813–838. https://doi.org/10.1037/0022-3514.48.4.813

Song, T., Huang, J., Tan, Y., & Yu, Y. (2019). Using User- and Marketer-Generated Content for Box Office Revenue Prediction: Differences Between Microblogging and Third-Party Platforms. *Information Systems Research*, *30*(1), 191–203. https://doi.org/10.1287/isre.2018.0797

Song, Y., Shi, S., Li, J., & Zhang, H. (2018). Directional skip-gram: Explicitly distinguishing left and right context forword embeddings. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *2*, 175–180. https://doi.org/10.18653/v1/n18-2028

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374.

Storme, M., Myszkowski, N., Davila, A., & Bournois, F. (2015). How subjective processing fluency predicts attitudes toward visual advertisements and purchase intention. *Journal of Consumer Marketing*, *32*(6), 432–440. https://doi.org/10.1108/JCM-10-2014-1187

Sture Holm. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

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

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

Tellis, G. J., MacInnis, D. J., Tirunillai, S., & Zhang, Y. (2019). What Drives Virality (Sharing) of Online Digital Content? The Critical Role of Information, Emotion, and Brand Prominence. *Journal of Marketing*, *83*(4), 1–20. https://doi.org/10.1177/0022242919841034

Telser, L. G. (1964). Advertising and competition. *Journal of Political Economy*, *72*(6), 537–562.

Teng, S., Khong, K. W., Chong, A. Y. L., & Lin, B. (2017). Examining the impacts of electronic word-of-mouth message on consumers' attitude. *Journal of Computer Information Systems*, *57*(3), 238–251. https://doi.org/10.1080/08874417.2016.1184012

Teodorescu, M. H. M., Morse, L., Awwad, Y., & Kane, G. C. (2021). Failures of Fairness in

Automation Require a Deeper Understanding of Human-ML Augmentation. *MIS Quarterly*, *45*(3).

Thielmann, I., & Hilbig, B. E. (2019). No gain without pain: The psychological costs of dishonesty. *Journal of Economic Psychology*, *71*, 126–137. https://doi.org/10.1016/j.joep.2018.06.001

Thompson, R., & Bowen, C. (2009). *Grammar of the Shot*. Taylor & Francis.

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

Todri, V., Ghose, A., & Singh, P. V. (2020). Trade-Offs in Online Advertising: Advertising Effectiveness and Annoyance Dynamics Across the Purchase Funnel. *Information Systems Research*, *31*(1), 102–125. https://doi.org/10.1287/isre.2019.0877

Tomkins, S. S. (1962). Affect, Imagery, Consciousness: Vol. 1. In *America*.

Van Kleef, G. A. (2009). How emotions regulate social life: The emotions as social information (EASI) model. *Current Directions in Psychological Science*, *18*(3), 184–188. https://doi.org/10.1111/j.1467-8721.2009.01633.x

Vehlen, A., Spenthof, I., Tönsing, D., Heinrichs, M., & Domes, G. (2021). Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-81987-x

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. *Journal of Cognitive Neuroscience*, *22*(12), 2864–2885. https://doi.org/10.1162/jocn.2009.21366

Wang, C., Tang, L., Lu, Y., Bian, S., Fujita, H., Zhang, D., Zhang, Z., & Wu, Y. (2019). ProductNet: a Collection of High-Quality Datasets for Product Representation Learning. *Companion Proceedings of The 2019 World Wide Web Conference*, 411–414.

Wang, J., Huang, P., Zhao, H., Zhang, Z., Zhao, B., & Lee, D. L. (2018). Billion-scale commodity embedding for e-commerce recommendation in alibaba. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 839–848.

Wang, Q., Huang, Y., Jasin, S., & Singh, P. V. (2022). Algorithmic Transparency with Strategic Users. *Management Science*, *69*(4), 2297–2317. https://doi.org/doi.org/10.1287/mnsc.2022.4475

Wang, X., Yu, Y., Huang, J., & Tan, Y. (2022). Picture-Evoked Affect and Review Helpfulness:

An Affect-as-Information Perspective. *SSRN Electronic Journal*. https://doi.org/dx.doi.org/10.2139/ssrn.3998058

Watson, L., & Spence, M. T. (2007). Causes and consequences of emotions on consumer behaviour: A review and integrative cognitive appraisal theory. *European Journal of Marketing*, *41*(5-6), 487–511. https://doi.org/10.1108/03090560710737570

Watts, S. A., & Zhang, W. (2008). Capitalizing on content: Information adoption in two online communities. *Journal of the Association for Information Systems*, *9*(2), 3.

Wood, A., Rychlowska, M., Korb, S., & Niedenthal, P. (2016). Fashioning the Face: Sensorimotor Simulation Contributes to Facial Expression Recognition. In *Trends in Cognitive Sciences* (Vol. 20, Issue 3, pp. 227–240). https://doi.org/10.1016/j.tics.2015.12.010

Xiao, Y., Zhang, H., & Cervone, D. (2018). Social Functions of Anger: A Competitive Mediation Model of New Product Reviews. *Journal of Product Innovation Management*, *35*(3), 367–388. https://doi.org/10.1111/jpim.12425

Xie, G. X., & Lee, M. J. (2008). Anticipated violence, arousal, and enjoyment of movies: viewers' reactions to violent previews based on arousal-seeking tendency. *Journal of Social Psychology*, *148*(3), 277–292. https://doi.org/10.3200/SOCP.148.3.277-292

Xue, B., Fu, C., & Shaobin, Z. (2014). A study on sentiment computing and classification of sina weibo with Word2vec. *Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014*, *2013*, 358–363. https://doi.org/10.1109/BigData.Congress.2014.59

Yang, X., Zhang, Z., Zhang, Z., Mo, Y., Li, L., Yu, L., & Zhu, P. (2016). Automatic Construction and Global Optimization of a Multisentiment Lexicon. *Computational Intelligence and Neuroscience*, *2016*. https://doi.org/10.1155/2016/2093406

Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, *38*(2), 539–560.

Yin, D., Bond, S. D., & Zhang, H. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research*, *54*(3), 447–463. https://doi.org/10.1509/jmr.13.0379

Yin, D., Bond, S. D., & Zhang, H. (2021). Anger in consumer reviews: Unhelpful but persuasive? *MIS Quarterly: Management Information Systems*, *45*(3), 1059–1086. https://doi.org/10.25300/MISQ/2021/15363

Yin, D., Mitra, S., & Zhang, H. (2016). Research Note—When Do Consumers Value Positive vs. Negative Reviews? An Empirical Investigation of Confirmation Bias in Online Word of Mouth. *Information Systems Research*, *27*(1), 131–144. https://doi.org/10.1287/isre.2015.0617

Yu, X., Liu, Y., Huang, X., & An, A. (2012). Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain. In *IEEE Transactions on Knowledge and Data Engineering* (Vol. 24, Issue 4, pp. 720–734). https://doi.org/10.1109/TKDE.2010.269

Yu, Y., Huang, S., Liu, Y., & Tan, Y. (2020). Emotions in Online Content Diffusion. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3724011

Yu, Y., Yang, Y., Huang, J., & Tan, Y. (2023). Unifying Algorithmic and Theoretical Perspectives: Emotions in Online Reviews and Sales. *MIS Quarterly*, *47*(1), 127–160.

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2022). What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features. *Management Science*, *68*(8), 5644–5666. https://doi.org/10.1287/mnsc.2021.4175

Zhou, M., Chen, G. H., Ferreira, P., & Smith, M. D. (2021). Consumer Behavior in the Online Classroom: Using Video Analytics and Machine Learning to Understand the Consumption of Video Courseware. *Journal of Marketing Research*, *58*(6), 1079–1100. https://doi.org/10.1177/00222437211042013

APPENDIX A

A.1 TRAINING WORD2VEC MODEL

I trained a Word2Vec model by using 3.26 million Chinese movie-related microblogging messages and 5 million microblogging messages from the NLPIR microblog corpus. First, the Chinese word segmentation module in Python segments all of the training text into words. Second, the Word2Vec model constructs a dictionary for all unique words. Every word is represented by a one-hot vector. The vector has the same dimensions as the dictionary, with only one non-zero element (equal to 1), representing the corresponding word. Third, every one-hot vector is input into a multi-layer neural network, whose output is the probabilistic distribution of the words adjacent to the input word. The initial distribution is randomly assigned. The model learns from the corpus and optimizes its weights so that the predicted probabilistic distribution ultimately approximates the observed distribution. Finally, for every word in the dictionary, I can extract its corresponding weight vector in the embedding layer (i.e., the first layer of the neural network). The weight vector, called a *word vector*, contains the semantic information of the word (Mikolov et al., 2013).

A.2 RANDOM HYPERPARAMETER SEARCH

In Table A.1, I present the results of the random hyperparameter search detailed in Algorithm 3. I find that the optimal parameter is K = 5 and α = 0.15, and the corresponding validation MAE is 0.07155. I note that all validation MAEs are between 0.071 to 0.077, indicating that the performance of Algorithm f is not very sensitive to hyperparameters.

Table A.1. Hyperparameters and Validation MAEs

K	α	Validation Error	K	α	Validation Error
5	**0.15**	**0.07155**	3	0.20	0.07274
5	0.06	0.07156	4	0.20	0.07293
4	0.12	0.07156	2	0.10	0.07325
3	0.10	0.07186	2	0.08	0.07328
5	0.18	0.07192	2	0.06	0.07337
8	0.16	0.07193	2	0.00	0.07345
3	0.08	0.07202	8	0.20	0.07355
3	0.18	0.07219	9	0.20	0.07376
4	0.02	0.07220	10	0.20	0.07397
9	0.16	0.07244	1	0.14	0.07679
7	0.18	0.07248	1	0.04	0.07694
5	0.20	0.07249	1	0.00	0.07696
7	0.00	0.07251			

A.3 COMPARISON OF ALTERNATIVE IMPLEMENTATIONS OF ALGORITHM f

Alternatively, Algorithm f in Algorithm 2 can be constructed by using the mean model and more sophisticated state-of-the-art models, i.e., RF and XGB. For the mean model, the prediction of a new word is produced by using the average emotion intensities of all words in the training set, i.e., $f_{mean}(w; L_0^{tr}) = \sum_i \frac{v_i}{|L_0^{tr}|}$, where $v_i \in L_0^{tr}$.

Algorithm 4: Mapping Word Vectors to Emotional Intensities and Testing	**Remarks**
Input a basic lexicon $L_0 = \{(w_i, v_i)\}_i^N$;	L_0 is the same as in Algorithm 1.
Randomly split L_0 into training (80%), validation (10%), and test (10%) sets: L_0^{tr}, L_0^{va}, and L_0^{te};	
Input a pre-trained Word2Vec model;	
FOR each emotion dimension k ($k \in \{1,2,\ldots,8\}$):	
Initiate a machine learning model g_k;	In this case, g_k can be RF or XGB. I use the default setting suggested by *sklearn*.
FOR each word $w_i \in L_0^{tr}$:	
y_i ←retrieve the true value of the kth emotional intensity e_k^0 from v_i;	
x_i ←retrieve the word vector of w_i;	x_i is a 200-dimension vector.
END FOR	
Train g_k with data $D_k = \{(x_i, y_i)\}_{i=1}^{\|L_0^{tr}\|}$;	
END FOR	
FOR each word $w_i \in L_0^{te}$:	Calculate test errors with the test set.
Retrieve the true value of the emotional intensities $v_i = (e_1^0, e_2^0, \ldots, e_8^0)$;	
x_i ←retrieve the word vector of w_i;	
FOR $k \in \{1,2,\ldots,8\}$:	
$e_k \leftarrow g_k(x_i)$;	
END FOR	
Calculate MAE $\epsilon_i = \sum_{i=1}^{8} \frac{\|e_i - e_i^0\|}{8}$;	
END FOR	
Calculate average MAE $\epsilon_g = \sum_i \frac{\epsilon_i}{\|L_0^{te}\|}$;	
RETURN ϵ_g;	

The mean model serves as a benchmark for other models. In addition, I note that Algorithm 2 is based on using the semantic similarity information embedded in the cosine distance between word vectors. Alternatively, one may directly utilize the 200-dimension word vectors as features to predict the emotional intensities of the word. As detailed in Algorithm 4, I use two state-of-the-art models, RF and XGB, to implement this idea. I find that the out-of-sample MAEs of the mean model, RF, XGB, and my model (Algorithm 2) are 0.117, 0.094, 0.091, and 0.073, respectively. The results show that, although RF and XGB can outperform the mean model, my algorithm achieves the best performance.

A.4 BALANCE CHECKS

In Table A.2, I provide the balance checks on all of the observable characteristics, i.e., gender, age, frequency of watching movies, and the extent to which they consider online reviews before they go to the movies (considering reviews, for short). Then, a Pearson's chi-square test is applied to test whether the assignment is balanced (not significant indicates that I cannot reject the null hypothesis of the balanced assignment). The Pearson's chi-square test is commonly used for testing statistical independence of categorical variables (e.g., group assignment and gender). I find that all characteristics are balanced ($p > 0.05$) except the frequency of watching movies ($p < 0.01$). To get unconfounded and robust treatment effects, I use a regression approach to explicitly control for all of the four factors.

Table A.2. Balance Checks

	Gender			**Age**			**Frequency of Watching Movies**			**Considering Reviews**		
Group	**Female**	**Male**	**Other**	**21–39 years old**	**40+ years old**	**Up to 20 years old**	**Less than once a month**	**Once a month to once a week**	**More than once a week**	**Yes**	**Maybe**	**No**
Positive	31	42	0	52	21	0	25	32	16	55	14	4
Love	19	40	0	34	25	0	43	16	0	47	12	0
Joy	28	32	0	30	29	1	37	22	1	46	12	2
Surprise	17	52	0	34	35	0	42	26	1	54	11	4
Anticipation	35	33	0	37	31	0	59	9	0	48	19	1
Negative	17	34	0	31	20	0	10	31	10	38	11	2
Anger	29	32	1	29	33	0	45	14	3	42	17	3
Anxiety	23	43	0	38	28	0	13	39	14	53	12	1
Disgust	31	34	0	23	42	0	44	17	4	48	16	1
Sadness	23	38	0	29	32	0	41	17	3	46	12	3
p-value of chi-square test (n = 634)	0.18			0.07			0.00			0.81		

APPENDIX B

B.1 Object-Level Complexity

To measure the object-level complexity of a video ad, I dynamically calculate the Shannon diversity index of objects for every 5 seconds in the video and average all of the indices of the video. Specifically, based on computer vision and deep learning techniques, Amazon Rekognition develops a scalable and accurate method for object detection in videos. For each millisecond, Rekognition detects the objects in the video content. Thousands of commonly seen objects in video ads can be detected. See https://aws.amazon.com/rekognition/video-features/ for more details. I am not able to disclose the concrete number of objects that can be detected due to confidentiality. But I argue that the number is large enough for detecting commonly seen objects in advertising. I note that Shin et al. (2020) used a model that can detect 1,700 objects in image content, and the model that I use can detect a larger number of objects.

These labels can be detected with a confidence score for each object valued from 0 to 1. For each time interval, I obtain the confidence score vector $q = (q_1, q_2, \ldots, q_d) \in [0,1]^d$, where d is the total number of objects that can be detected. Then, I normalize the vector such that $\sum_i^d q_i = 1$. Then, the object-level complexity for the focal time interval is calculated as $-\sum_{i=1}^{d} q_i \log(q_i)$, which is maximized when q is uniformly distributed. Intuitively, the value of the object-level complexity measure is higher when a video ad contains a higher number of unique objects in the time interval. Next, I average object-level complexity values of all time intervals for each video ad, denoted as $ObjCmp_{jt}$. I chose the time interval as 5 seconds, the minimum length of videos in my sample. If I choose a value larger than 5 seconds, longer videos will have systematically higher values of the complexity measure. I do not choose a shorter time interval because objects in video ads tend to last for a few seconds so that video viewers will recognize them.

B.2 Product Relevance

I detail the procedure of measuring ad-product relevance in this section. First, I use Amazon Rekognition to detect all object labels and texts in video ads, which summarize the main content of the video. Second, to measure the semantic relevance of the content and the product, I resort to the Word2Vec model, a commonly used neural network language model pre-trained on about

100 billion words and containing 300-dimensional vectors for 3 million words and phrases (known as word vectors) (Mikolov et al., 2013). For each word in the video content (including object labels and texts), I use the model to obtain its corresponding vector. I average the vectors of all words in the video content as the semantic representation for the video ad. It is empirically shown that the average word vector is a successful and efficient way of representing the semantic information of a short document (Arora et al., 2017). Similarly, I obtain the average word vector of the product name, which summarizes the basic information of the product. The cosine similarity of two average word vectors is used to measure the product relevance of the video ad. The higher the value, the more relevant the video ad and the product.

B.3 Color Complexity

To construct color complexity, I first sample 3 frames per second of the videos. The colors of each frame are mapped into their closest color in a standard 16-color palette. In other words, I calculate the proportion of pixels that belong to each of the 16 categories of standard colors, following Shin et al. (2020). Thus, for each frame i of a certain video, I have a color distribution vector $\Gamma_i \in [0,1]^{16}$ and $\sum_{k=1}^{16} \Gamma_{ik} = 1$. Then, I construct color complexity by the entropy formula: $ColorCmp_i = -\sum_{k=1}^{16} \Gamma_{ik} \log(\Gamma_{ik})$. Finally, I average all $ColorCmp_i$ for a given video as a measure of video colorfulness. The larger the variable, the more colorful the video.

B.4 Correlations between Video Features

Figure B.1 shows the correlation matrix of video features. The analysis is conducted on videos with $HasPerson_{jt} = 1$ because otherwise $TwoShot_{jt}$ and $HasCelebrity_{jt}$ are not defined. The correlations are shown in percentages (ranging from -100 to 100). The figure shows that the video features in my analyses are relatively orthogonal to each other. Due to confidentiality, I cannot show the summary statistics of these variables. However, I note that the continuous variables are approximately normally or uniformly distributed, without any outliers observed.

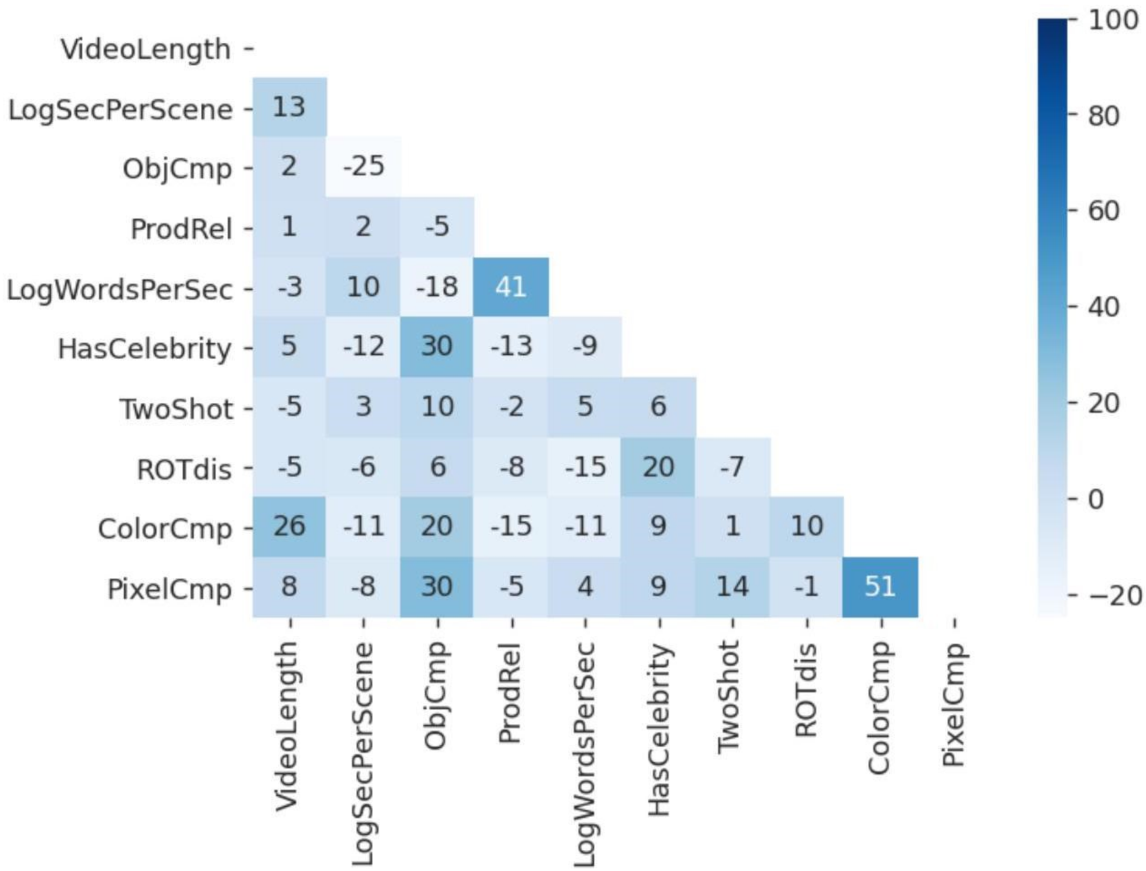


Figure B.1. Correlations in Percentages (Ranging from -100 to 100)

B.5 VALIDATING PRODUCT EMBEDDING VECTORS

As the 16-dimension vectors summarize the information of the 144-dimension embedding vectors, validity of the latter would be guaranteed by that of the former. Thus, it is sufficient to illustrate the validity of the 16-dimension vectors. I illustrate the validity of the 16-dimension vectors $ProdEmb_{jt}$ as follows. If the 16-dimension vectors summarize the product characteristics, they should be able to differentiate products of different categories. Such a practice is a common way to validate embedding vectors in the literature (Grbovic & Cheng, 2018; J. Wang et al., 2018). Based on the search-experience-credence (SEC) good classification (Ekelund et al., 1995), I choose one representative type of products from each of the three categories, namely, automotive (AUTO) in search goods, drinking (DRINKING) in experience goods, and health-related products (HEALTH) in credence goods. Then the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm (der Maaten & Hinton, 2008) is leveraged to visualize the product embedding vectors in a three-dimensional space. In a t-SNE plot, each point represents an embedding vector. If two points are close to each other in the plot, their corresponding products are similar. Figure B.2 shows that the three types of points (products) form three highly separated clusters, indicating the validity of the vectors in terms of differentiating products with distinct characteristics.

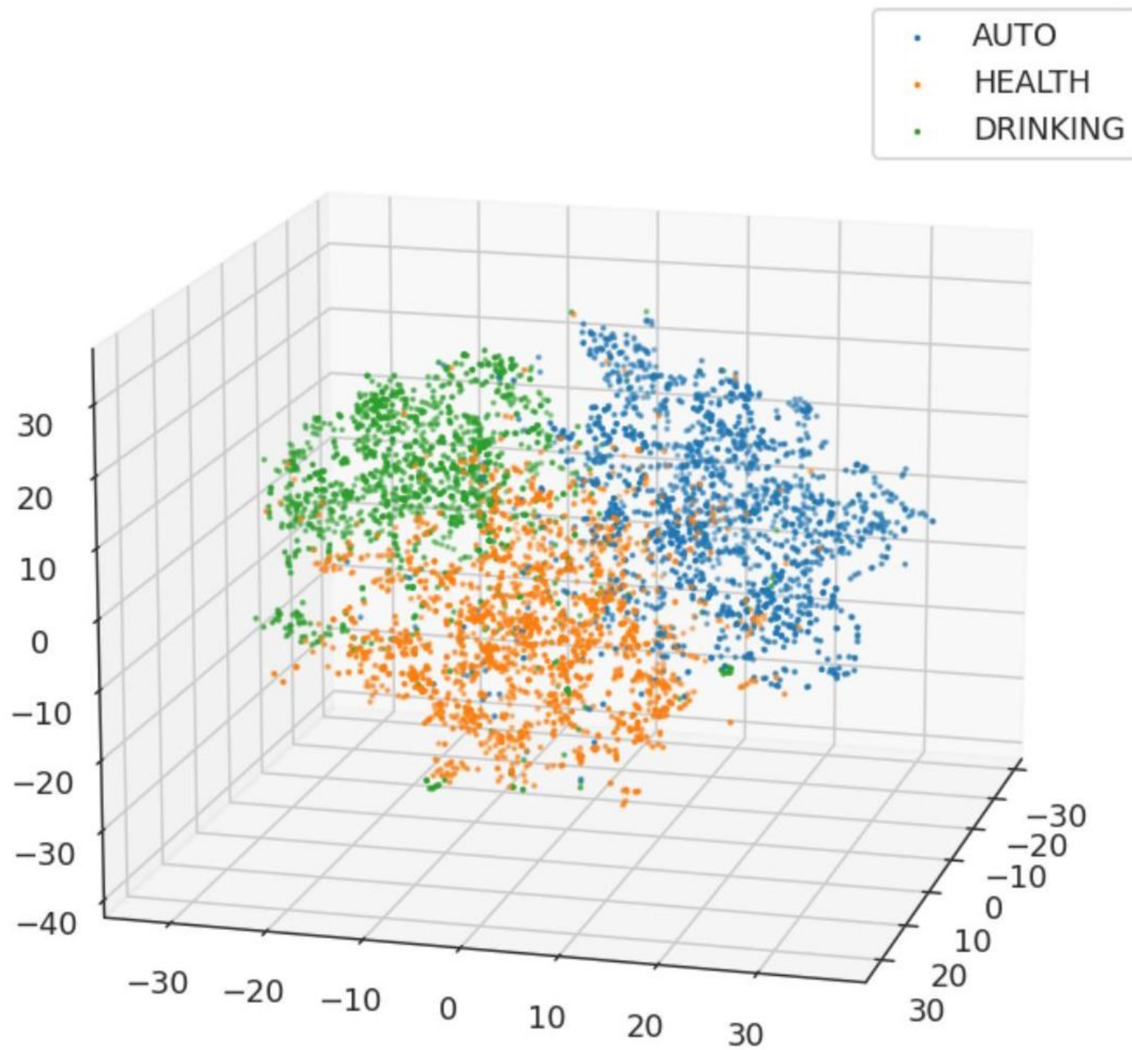


Figure B.2. 3-D Illustration of Product Vectors Generated by t-SNE

B.6 MAHALANOBIS DISTANCE MATCHING RESULTS

Despite its prevalence in empirical studies, propensity score matching (PSM) has been recently accused of yielding fragile and non-robust estimates, because its parametric form of estimating propensity score may not always be correctly specified (King & Nielsen, 2019). Mahalanobis

Distance Matching (MDM), in contrast, is free of parametric assumptions and is empirically shown to work better than PSM (Ripollone et al., 2018). Thus, I adopt MDM in my analysis. The Mahalanobis distance is a multi-dimensional generalization of the idea of measuring how many standard deviations a point is away from the mean of a distribution. MDM measures the distance between the two observations X_i and X_j with the Mahalanobis distance $M(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$, where S is the sample covariance matrix of X. In Table B.3, I present the results of MDM. Percentage biases in the table are derived by percentage differences of variables of the queries without a video ad and those of the queries with a video ad. As all biases are significantly reduced and less than 0.23%, indicating that MDM eliminates query-level imbalance after matching. Due to confidentiality, mean values are not disclosed.

Table B.3. Percentage Bias between Queries with and without a Video Ad

Matching Variable	Before matching	After matching
Log-transformed average price	−2.81%	−0.07%
Log-transformed average number of reviews	−13.58%	−0.23%
Average ratings	−3.97%	0.05%
Product differentiation measure ProdDif	3.61%	0.04%
Proportion of queries containing experience goods	−8.89%	−0.11%
Proportion of queries containing search goods	−13.16%	0.00%
Proportion of queries at mobile end	3.42%	0.00%

B.7 Estimation of the Two-stage Model

I use maximum likelihood estimation for parameters of interest $\theta = (\beta^Q, \gamma^Q, \sigma_Q), Q \in \{A, C\}$ in the two-stage model. The likelihood function L_{jt} can be written as follows:

$$\begin{aligned} L_{jt} &= \mathcal{P}(y_{jt} \mid Z_{jt}, X_{jt}, M_j) \\ &= y_{jt}[\mathcal{P}(y_{jt} = 1 \mid A_{jt} = 1, X_{jt}, M_j)\mathcal{P}(A_{jt} = 1 \mid Z_{jt}, M_j)] \\ &\quad +(1 - y_{jt})[\mathcal{P}(y_{jt} = 0 \mid A_{jt} = 1, X_{jt}, M_j)\mathcal{P}(A_{jt} = 1 \mid Z_{jt}, M_j) + \mathcal{P}(A_{jt} = 0 \mid Z_{jt}, M_j)], \end{aligned}$$

where the first part in the above equation is the likelihood of observing clicking $(y_{jt}=1)$, where attention has to be received $(A_{jt}=1)$; the second part in the above equation is the likelihood of observing not clicking, where either attention is received but the interest level is too low to induce clicking, or attention is not received $(A_{jt}=0)$. Further,

$$\begin{aligned}\mathcal{P}(y_{jt} \mid A_{jt}=1, X_{jt}, M_j) &= \int \mathcal{P}(y_{jt} \mid A_{jt}=1, X_{jt}, \alpha_j^C) dF(\alpha_j^C \mid M_j) \\ &= \int \frac{y_{jt}\exp\left(X_{jt}\beta^C + M_j\gamma^C + \epsilon_j^C\right) + 1 - y_{jt}}{1+\exp\left(X_{jt}\beta^C + M_j\gamma^C + \epsilon_j^C\right)} \phi^C(\epsilon_j^C) d\epsilon_j^C,\end{aligned}$$

Similarly, $\mathcal{P}(A_{jt} \mid Z_{jt}, M_j) = \int \frac{A_{jt}\exp\left(X_{jt}\beta^A + M_j\gamma^A + \epsilon_j^A\right) + 1 - A_{jt}}{1+\exp\left(X_{jt}\beta^A + M_j\gamma^A + \epsilon_j^A\right)} \phi^A(\epsilon_j^A) d\epsilon_j^A$,

where $F(\cdot)$ denotes the cumulative density function of $\alpha_j^C \mid M_j$, and $\phi^Q(\cdot)$ is the probability function of $\mathcal{N}(0, \sigma_Q^2)$, for $Q \in \{A, C\}$. To deal with integrals, I use the Monte Carlo approximation in numerical estimation. Then, I estimate $\hat{\theta}$ by maximizing the log-likelihood function $L = \sum_{jt} \log L_{jt}$. Due to the efficiency of maximum likelihood estimators, the variance of $\hat{\theta}$ is obtained by the diagonal values of the inverse Hessian matrix.

APPENDIX C

C.1 PROOF OF PROPOSITION 1

Denote the loss function $L_0 = \mathbb{E}[(Y_0(x) - \eta)^2 + \lambda x^2]$ when $e \equiv 0$. Plug in the linear allocation rule $Y_0(x) = \beta x + \beta_0$, I have:

$$L_0 = \mathbb{E}[(\beta x + \beta_0 - \eta)^2 + \lambda x^2]$$

To derive the optimal β_0 for Y_0 based on the first-order condition, I have:

$$\partial L_0 / \partial \beta_0 = \mathbb{E}[2(\beta x + \beta_0 - \eta)] = 0$$

Therefore, the optimal β_0 for Y_0 is $\mathbb{E}[\eta - \beta x]$, and L_0 as a function of β can be written as:

$$L_0(\beta) = \mathbb{E}[(\beta x + \mathbb{E}[\eta - \beta x] - \eta)^2 + \lambda x^2]$$

Now, I consider the case when e is a random variable with the mean and variance as μ_e and σ_e^2, respectively. The loss function L can be written as:

$$L = \mathbb{E}[(Y(\hat{x}) - \eta)^2 + \lambda x^2] = \mathbb{E}[(\beta \hat{x} + \beta_0 - \eta)^2 + \lambda x^2]$$

Similarly, from the first-order condition, I have the optimal β_0 for Y as $\mathbb{E}[\eta - \beta \hat{x}] = \mathbb{E}[\eta - \beta x] - \beta \mu_e$. Plug this optimal β_0 into L, and I have:

$$\begin{aligned}
L(\beta) &= \mathbb{E}[(\beta x + \beta e + \mathbb{E}[\eta - \beta x] - \beta \mu_e - \eta)^2 + \lambda x^2] \\
&= \mathbb{E}[(\beta x - \eta + \mathbb{E}[\eta - \beta x])^2] + \lambda \mathbb{E}[x^2] + 2\beta \mathbb{E}[(e - \mu_e)(\beta x - \eta + \mathbb{E}[\eta - \beta x])] + \beta^2 \mathbb{E}[(e - \mu_e)^2] \\
&= L_0(\beta) + 2\beta \mathbb{E}[(e - \mu_e)(\beta x - \eta + \mathbb{E}[\eta - \beta x])] + \beta^2 \mathbb{E}[(e - \mu_e)^2] \\
&= L_0(\beta) + \beta^2 \sigma_e^2.
\end{aligned}$$

The last step holds because

$$\mathbb{E}[(e - \mu_e)(\beta x - \eta + \mathbb{E}[\eta - \beta x])] = \mathbb{E}[(e - \mu_e)]\mathbb{E}[(\beta x - \eta + \mathbb{E}[\eta - \beta x])] = 0,$$

due to the independence between e and (x, η) and because $\mathbb{E}[(e - \mu_e)^2] = \sigma_e^2$ by definition.

Next, note that I have:

$$\begin{aligned}
\mathbb{E}[(Y_0(x) - \eta)^2] &= \mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x] + \mathbb{E}[\eta \mid x] - \eta)^2] \\
&= \mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])^2] + \mathbb{E}[(\mathbb{E}[\eta \mid x] - \eta)^2] + 2\mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])(\mathbb{E}[\eta \mid x] - \eta)] \\
&= \mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])^2] + \mathbb{E}[(\mathbb{E}[\eta \mid x] - \eta)^2]
\end{aligned}$$

The last step holds due to the law of iterated expectation:

$$\begin{aligned}
\mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])(\mathbb{E}[\eta \mid x] - \eta)] &= \mathbb{E}\big[\mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])(\mathbb{E}[\eta \mid x] - \eta) \mid x]\big] \\
&= \mathbb{E}[Y_0(x) - \mathbb{E}[\eta \mid x]]\mathbb{E}[(\mathbb{E}[\eta \mid x] - \eta) \mid x] \\
&= \mathbb{E}[Y_0(x) - \mathbb{E}[\eta \mid x]](\mathbb{E}[\eta \mid x] - \mathbb{E}[\eta \mid x]) = 0
\end{aligned}$$

In sum, I have:

$$\begin{aligned} L(\beta) &= L_0(\beta) + \beta^2\sigma_e^2 \\ &= \mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])^2] + \mathbb{E}[(\mathbb{E}[\eta \mid x] - \eta)^2] + \lambda\mathbb{E}[x^2] + \beta^2\sigma_e^2. \end{aligned}$$

C.2 Proof of Lemmas 1 and 2

From Appendix C.1, I have:

$$L(\beta) = \mathbb{E}[(\beta x - \eta + \mathbb{E}[\eta - \beta x])^2] + \lambda\mathbb{E}[x^2] + \beta^2\sigma_e^2.$$

Plug $x = \eta + m\beta\gamma$ in the above equation, I get:

$$L(\beta) = \mathbb{E}\left[\left(m\beta^2(\gamma - \mu_\gamma) - (1-\beta)(\eta - \mu_\eta)\right)^2\right] + \lambda\mathbb{E}[(\eta + m\beta\gamma)^2] + \beta^2\sigma_e^2. \tag{EC.1}$$

Use standard mean-variance decomposition:

$$\begin{aligned} L(\beta) &= m^2\sigma_\gamma^2\beta^4 + 2m\rho\sigma_\eta\sigma_\gamma\beta^3 + \left[\sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \lambda m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\right]\beta^2 \\ &\quad + \left[-2\sigma_\eta^2 + 2m\lambda\left(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\right)\right]\beta + (1+\lambda)\sigma_\eta^2 + \lambda\mu_\eta^2 \end{aligned} \tag{EC.2}$$

and the first-order and second-order derivatives:

$$\begin{aligned} L'(\beta) &= 4m^2\sigma_\gamma^2\beta^3 + 6m\rho\sigma_\eta\sigma_\gamma\beta^2 && \text{(EC.3)} \\ &\quad + 2\left[\sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \lambda m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\right]\beta + \left[-2\sigma_\eta^2 + 2m\lambda\left(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\right)\right], && \text{(EC.4)} \\ L''(\beta) &= 12m^2\sigma_\gamma^2\beta^2 + 12m\rho\sigma_\eta\sigma_\gamma\beta + 2\left[\sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \lambda m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\right]. && \text{(EC.5)} \end{aligned}$$

Proof of Lemma 1: I note that when $\rho \in (-1,0)$,

$$\begin{aligned} L''(0) &= 2\left[\sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \lambda m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\right] > 0 \\ L'(2) &= 2\left(4m\sigma_\gamma + \rho\sigma_\eta\right)^2 + 2(1-\rho^2)\sigma_\eta^2 + 4\sigma_e^2 + 2m\lambda(\mathbb{E}[\eta\gamma] + m\mathbb{E}[\gamma^2]) > 0 \end{aligned}$$

First, I show when $\lambda \geq \frac{(3\rho^2-2)\sigma_\eta^2 + 4m\rho\sigma_\eta\sigma_\gamma - 2\sigma_e^2}{2m^2(\mu_\gamma^2 + \sigma_\gamma^2)}$, the conclusion holds. Note that this condition guarantees that $L''(\beta) \geq 0$ because:

$$\begin{aligned} L''(\beta) &= 12m^2\sigma_\gamma^2\beta^2 + 12m\rho\sigma_\eta\sigma_\gamma\beta + 2\left[\sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \lambda m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\right] \\ &= 12\left(m\sigma_\gamma\beta + \rho\sigma_\eta/2\right)^2 - \left[(3\rho^2 - 2)\sigma_\eta^2 + 4m\rho\sigma_\eta\sigma_\gamma - 2\sigma_e^2\right] + 2m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\lambda \geq 0 \end{aligned}$$

It indicates the quartic function $L(\beta)$ is convex at $[0, \infty)$ and its minimum point β^* is unique. If $\lambda \geq \lambda_0$, I have $L'(0) \geq 0$. Due to the convexity, the minimum point $\beta^* = 0$ is unique. If $\lambda < \lambda_0$, which implies that $L'(0) < 0$, then $\beta^* > 0$. Because I have shown that $L'(2) > 0$ and due to the continuity of $L'(\beta)$, there exists $\beta^* \in (0,2)$ such that $L'(\beta^*) = 0$ which minimizes the convex function $L(\beta)$.

Second, I show that when $\lambda < \frac{(3\rho^2-2)\sigma_\eta^2+4m\rho\sigma_\eta\sigma_\gamma-2\sigma_e^2}{2m^2(\mu_\gamma^2+\sigma_\gamma^2)}$, the conclusion holds. The proof is less straightforward because $L(\beta)$, in this case, is not necessarily convex. Denote $L_1(\beta) \triangleq \mathbb{E}\left[\left(m\beta^2(\gamma-\mu_\gamma)-(1-\beta)(\eta-\mu_\eta)\right)^2\right]$ and $L_2(\beta) \triangleq \lambda\mathbb{E}[(\eta+m\beta\gamma)^2]+\beta^2\sigma_e^2$. I have $L(\beta) = L_1(\beta)+L_2(\beta)$ according to Equation EC.1.

Step 1. I show:

$$\forall\beta_1^* \in \arg\min_{\beta\geq 0} L_1(\beta),\ \beta_1^* \in (0,2).$$

Use the standard mean-variance decomposition, I have:

$$L_1(\beta) = (1-\beta)^2\sigma_\eta^2 + m^2\beta^4\sigma_\gamma^2 - 2(1-\beta)m\beta^2\rho\sigma_\eta\sigma_\gamma$$

I slightly abuse the notation L_1 and consider it as a function of both β and ρ, that is, $L_1(\beta) = L_1(\beta,\rho)$. It is helpful to include the case when $\rho = -1$.

Because $\left.\frac{\partial L_1}{\partial\beta}\right|_{\beta=0} = -2\sigma_\eta^2 < 0, \beta_1^* > 0$. Next, I prove that $\beta_1^* < 2$. Denote $\underline{\beta}$ as the smallest β that minimizes $L_1(\beta,\rho=-1)$, that is,

$$\underline{\beta} = \min\left\{\arg\min_{\beta\geq 0} L_1(\beta,\rho=-1)\right\}.$$

Note that $L_1(\beta,\rho=-1)$ can be written as the square of a quadratic function L_{11} of β, that is,

$$L_1(\beta,\rho=-1) = \left[m\sigma_\gamma\beta^2+(1-\beta)\sigma_\eta\right]^2 \triangleq [L_{11}(\beta)]^2.$$

Because $L_{11}(\beta) = m\sigma_\gamma\left(\beta-\frac{\sigma_\eta}{2m\sigma_\gamma}\right)^2 - \sigma_\eta\left(\frac{\sigma_\eta}{4m\sigma_\gamma}-1\right)$, if $\frac{\sigma_\eta}{4m\sigma_\gamma} \leq 1, L_{11}(\beta) \geq 0, \forall\beta \in [0,\infty)$. Then, the minimum point $\underline{\beta} = \frac{\sigma_\eta}{2m\sigma_\gamma} \leq 2$. Otherwise, $L_1(\beta,\rho=-1)$ is minimized when $L_{11}(\beta) = 0$. By definition, $\underline{\beta}$ is the smaller root of the equation $L_{11}(\beta) = 0$. That is, $\underline{\beta} = \frac{\sigma_\eta-\sqrt{\sigma_\eta^2-4m\sigma_\eta\sigma_\gamma}}{2m\sigma_\gamma}$. Note that $\frac{\sigma_\eta-\sqrt{\sigma_\eta^2-4m\sigma_\eta\sigma_\gamma}}{2m\sigma_\gamma} < 2$ holds $\forall m,\sigma_\gamma,\sigma_\eta \in (0,\infty)$ and $\frac{\sigma_\eta}{4m\sigma_\gamma} > 1$. In sum, I have $\underline{\beta} \leq 2$.

Step 1.1. I show that $\beta_1^* \leq \underline{\beta} \leq 2$ when $\underline{\beta} > 2/3$. Suppose $\beta_1^* > \underline{\beta}$. Note that $\forall\beta$ such that $\beta > \underline{\beta} > 2/3$, I have:

$$\frac{\partial^2 L_1}{\partial\beta\,\partial\rho} = 2m\sigma_\eta\sigma_\gamma\beta(3\beta-2) > 0$$

For $\rho \in (-1,0)$, by the definitions of β_1^* and $\underline{\beta}$, I have $L_1(\beta_1^*, \rho) \leq L_1(\underline{\beta}, \rho)$ and $L_1(\beta_1^*, -1) \geq L_1(\underline{\beta}, -1)$. Therefore,

$$\begin{aligned}
0 &\geq \left[L_1(\beta_1^*, \rho) - L_1(\underline{\beta}, \rho)\right] - \left[L_1(\beta_1^*, -1) - L_1(\underline{\beta}, -1)\right] \\
&= \int_{\underline{\beta}}^{\beta_1^*} \frac{\partial L_1(\beta, \rho)}{\partial \beta} d\beta - \int_{\underline{\beta}}^{\beta_1^*} \frac{\partial L_1(\beta, -1)}{\partial \beta} d\beta \\
&= \int_{\underline{\beta}}^{\beta_1^*} \int_{-1}^{\rho} \frac{\partial^2 L_1(\beta, t)}{\partial \beta\, \partial t} d\beta dt > 0.
\end{aligned}$$

In other words, I have the contradiction of $0 > 0$. Hence, it must hold that $\beta_1^* \leq \underline{\beta}$ when $\underline{\beta} > 2/3$. Because I have shown that $\underline{\beta} \leq 2$, I have $\beta_1^* \leq \underline{\beta} \leq 2$.

Step 1.2. I show that $\beta_1^* = \underline{\beta}$ cannot hold when $\underline{\beta} > 2/3$. Therefore, I must have $\beta_1^* < \underline{\beta} \leq 2$ when $\underline{\beta} > 2/3$. Suppose $\beta_1^* = \underline{\beta} > 2/3$, I have:

$$\left.\frac{\partial L_1(\beta, \rho)}{\partial \beta}\right|_{\beta=\underline{\beta}} = \left.\frac{\partial L_1(\beta, -1)}{\partial \beta}\right|_{\beta=\underline{\beta}} = 0$$
$$\Rightarrow 4m^2\sigma_\gamma^2\underline{\beta}^3 + 6m\rho\sigma_\eta\sigma_\gamma\underline{\beta}^2 + 2\left(\sigma_\eta^2 - 2m\rho\sigma_\eta\sigma_\gamma\right)\underline{\beta} - 2\sigma_\eta^2 = 0$$

And $4m^2\sigma_\gamma^2\underline{\beta}^3 - 6m\rho\sigma_\eta\sigma_\gamma\underline{\beta}^2 + 2\left(\sigma_\eta^2 + 2m\sigma_\eta\sigma_\gamma\right)\underline{\beta} - 2\sigma_\eta^2 = 0$

Take the difference between the above two equations and I have:

$$2m\sigma_\eta\sigma_\gamma(1+\rho)(3\underline{\beta} - 2)\underline{\beta} = 0,$$

which contradicts the conditions $\rho > -1, \underline{\beta} > 2/3$, and $m\sigma_\eta\sigma_\gamma > 0$.

Step 1.3. I show that $\beta_1^* < 2$ when $\underline{\beta} \leq 2/3$. To simplify my notation, denote $k \triangleq m\sigma_\gamma/\sigma_\eta > 0$. I have already shown that $\underline{\beta} = \frac{1}{2k}$ if $k \in [1/4, \infty)$ and $\underline{\beta} = \frac{1-\sqrt{1-4k}}{2k}$ if $k \in (0, 1/4)$.

Suppose $\underline{\beta} = \frac{1-\sqrt{1-4k}}{2k} \leq 2/3$, I will have $0 < 1 - 4k/3 \leq \sqrt{1-4k}$ which implies $16k/9 < -4/3$ and contradicts with $k > 0$. Therefore, it must be $\underline{\beta} = \frac{1}{2k}$. Further, from $\underline{\beta} \leq 2/3$, I have $k \in [3/4, \infty)$. Now, suppose $\beta_1^* \geq 2$,

$$\begin{aligned}
\left.\frac{\partial L_1(\beta, \rho)}{\partial \beta}\right|_{\beta=\beta_1^*} = 0 \quad &\Rightarrow \sigma_\eta^2\left(4k\beta_1^{*3} + 6k\rho\beta_1^{*2} + 2(1-2k\rho)4k\beta_1^* - 2\right) = 0 \\
&\Rightarrow 4k\beta_1^{*3} + 6k\rho\beta_1^{*2} + 2(1-2k\rho)4k\beta_1^* - 2 = 0 \\
&\Rightarrow 2k\beta_1^*[2k\beta_1^{*2} + (3\beta_1^* - 2)\rho)] + 2(\beta_1^* - 1) = 0.
\end{aligned}$$

Note that the left-hand side is increasing in ρ given $\beta_1^* \geq 2$ and, therefore, I have:

$$\begin{aligned} 0 &= 2k\beta_1^*[2k\beta_1^{*2} + (3\beta_1^* - 2)\rho)] + 2(\beta_1^* - 1) \\ &> 2k\beta_1^*\left[2k\beta_1^{*2} - (3\beta_1^* - 2)\right)\big] + 2(\beta_1^* - 1) \\ &= 2k\beta_1^*[(2k\beta_1^* - 3)\beta_1^* + 2] + 2(\beta_1^* - 1) > 0 \end{aligned}$$

The last step is due to $k \geq 3/4, \beta_1^* \geq 2$, and $2k\beta_1^* \geq 3$. But I have a contradiction of $0 > 0$. Hence, it must be $\beta_1^* < 2$ when $\underline{\beta} \leq 2/3$. Step 2. I show $\beta^* \leq \beta_1^*$, and then from Step 1 I have $\beta^* < 2$. Note that

$$L_2(\beta) = \lambda\mathbb{E}[(\eta + m\beta\gamma)^2] + \beta^2\sigma_e^2 = \lambda\mathbb{E}[\eta^2] + 2m\lambda\mathbb{E}[\eta\gamma]\beta + (\lambda m\mathbb{E}[\gamma^2] + \sigma_e^2)\beta^2$$

It implies that L_2 is strictly increasing in $\beta \in [0, \infty)$. Thus $\forall\beta > \beta_1^*$, I have:

$$L(\beta) = L_1(\beta) + L_2(\beta) > L_1(\beta_1^*) + L_2(\beta_1^*) > L(\beta_1^*)$$

which implies $0 \leq \beta^* \leq \beta_1^* < 2$.

Step 3. I show that $L(\beta)$ has at most one minimum point at $\beta \in (0, \infty)$. Given $0 \leq \lambda < \frac{(3\rho^2-2)\sigma_\eta^2 + 4m\rho\sigma_\eta\sigma_\gamma - 2\sigma_e^2}{2m^2(\mu_\gamma^2 + \sigma_\gamma^2)}$, I have:

$$\begin{aligned} &(3\rho^2 - 2)\sigma_\eta^2 + 4m\rho\sigma_\eta\sigma_\gamma > 0 \text{ and } \rho \in (-1,0) \\ &\Rightarrow -\rho\sigma_\eta/(m\sigma_\gamma)[-\rho\sigma_\eta/(m\sigma_\gamma) - 4/(3 - 2/\rho^2)] > 0 \\ &\Rightarrow -\rho\sigma_\eta/(m\sigma_\gamma) > 4/(3 - 2/\rho^2) > 4 \end{aligned}$$

Therefore, the minimum point of the quadratic function $L''(\beta)$ satisfies $\beta = -\rho\sigma_\eta/(2m\sigma_\gamma) > 2$. I denote the two roots of $L''(\beta) = 0$ as $\beta_{21} < \beta_{22}$ and it follows that $\beta_{22} > -\rho\sigma_\eta/(2m\sigma_\gamma) > 2$. In addition, I have $\beta_{21} \geq 0$ because I have already shown that $L''(0) \geq 0$.

I note that $L'(\beta)$ is a cubic function with a positive coefficient of the third-order term $(4m^2\sigma_\gamma^2 > 0)$. If $\lambda \geq \lambda_0$, which implies that $L'(0) \geq 0, L'(\beta) = 0$ has at most two positive roots. If $L'(\beta) = 0$ does not have a positive root, which implies that $L'(\beta) > 0$ in $(0, \infty)$, then $\beta^* = 0$ is the unique minimum point for $L(\beta)$.

If $L'(\beta) = 0$ has only one positive root, denoted as β_{11}, due to the property of the cubic function, I have $L'(\beta) > 0$ for $\beta \in (\beta_{11}, \infty)$. Suppose that $L'(\beta) < 0$ for $\beta \in (0, \beta_{11})$. Note that β_{22} is a positive minimum point of $L'(\beta)$, I have $L'(\beta_{22}) < 0$ and $\beta_{22} < \beta_{11}$. Therefore, for $\beta \in (0, \beta_{22})$, $L'(\beta) < 0$. It contradicts my conclusion that $2 \in (0, \beta_{22})$ and $L'(2) > 0$. In other words, I must have $L'(\beta) > 0$ for $\beta \in (0, \beta_{11})$. Thus, $L(\beta)$ is strictly increasing in $(0, \infty)$ because the only positive root (if it exists) of $L'(\beta) = 0$ is a stationary point for $L(\beta)$ and $L(\beta)$ keeps increasing before and after this point (i.e., $L'(\beta) > 0$ for $\beta \in (0, \beta_{11}) \cup (\beta_{11}, \infty)$). Therefore, I have $\beta^* = 0$ and it is the unique minimum point for $L(\beta)$.

Otherwise, suppose that $L'(\beta) = 0$ has two positive roots $0 < \beta_{11} < \beta_{12}$. In this case, $L'(\beta) > 0$ for $\beta \in (0, \beta_{11}) \cup (\beta_{12}, \infty)$. Because β_{22} is the inflection point of $L'(\beta)$, I have $2 < \beta_{22} < \beta_{12}$. Therefore, from $L'(2) > 0$ I have $2 \in (0, \beta_{11})$. It implies $L'(\beta) > 0$ for $\beta \in (0,2)$. Therefore, $L(\beta)$ is strictly increasing in $(0,2)$. Because in Step 2 I have already shown that $\beta^* \in [0,2)$, I conclude that the unique minimum point is $\beta^* = 0$. In sum, when $\lambda \geq \lambda_0$, the unique minimum point is $\beta^* = 0$. Next, I focus on the case when $\lambda < \lambda_0$. I show in this case $\beta^* \in (0,2)$ and is unique. First, $\lambda < \lambda_0$ implies $L'(0) < 0$ and thus I must have $\beta^* > 0$. Second, note that $L'(2) > 0$, which implies that there exists at least one $\beta^* \in (0,2)$ such that $L'(\beta^*) = 0$.

I show that this β^* is unique. Suppose otherwise, I have another local minimum point $\beta_{12} > 0$ which satisfies that given β is in the neighborhood of $\beta_{12}, L'(\beta) < 0$ if $\beta < \beta_{12}$ and $L'(\beta) > 0$ if $\beta > \beta_{12}$. Due to the cubic function's property, β_{12} is the largest root of $L'(\beta) = 0$. Because $\beta_{22} > 2$ is the last inflection point the cubic function $L'(\beta)$, I conclude that $\beta_{12} > \beta_{22} > 2$. From Step 2, β_{12} is not a global minimum point of $L(\beta)$. Therefore, I have a unique $\beta^* \in (0,2)$.

Proof of Lemma 2: For $\rho \in [0,1)$, I first note that $L''(\beta)$ is a quadratic function strictly increasing in $\beta \in [0, \infty)$. It implies that for $\beta \to \infty, L''(\beta) > 0$. Further, I have:

$$L''(1) = 12m^2\sigma_\gamma^2 + 8m\rho\sigma_\gamma\sigma_\eta + 2\sigma_\eta^2 + 2\sigma_e^2 + 2\lambda m^2(\mu_\gamma^2 + \sigma_\gamma^2) > 0$$

$$L''(0) = 2m^2(\mu_\gamma^2 + \sigma_\gamma^2)\left(\lambda - \frac{2m\rho\sigma_\gamma\sigma_\eta - \sigma_\eta^2 - \sigma_e^2}{m^2(\mu_\gamma^2 + \sigma_\gamma^2)}\right) = 2m^2(\mu_\gamma^2 + \sigma_\gamma^2)(\lambda - \lambda_1)$$

$$L'(1) = 4m^2\sigma_\gamma^2 + 2m\rho\sigma_\gamma\sigma_\eta + 2\sigma_e^2 + 2\lambda m^2(\mu_\gamma^2 + \sigma_\gamma^2) + 2\lambda m(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma) > 0$$

$$L'(0) = 2m(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma)\left(\lambda - \frac{\sigma_\eta^2}{m(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma)}\right) = 2m(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma)(\lambda - \lambda_0).$$

Step 1. When $\lambda \in (0, \lambda_0)$, I show that $\beta^* \in (0,1)$ and is unique.

$\lambda \in (0, \lambda_0)$ implies $L'(0) < 0$ and $\beta^* > 0$. If $\lambda < \lambda_1, L''(0) < 0$. Because $L''(\beta)$ is monotone in $[0, \infty), L''(\beta)$ is first negative then positive as β increases. It implies that $L(\beta)$ is first concave and then convex in $(0, \infty)$. Therefore, $L(\beta)$ has a unique minimum point in $(0, \infty)$. Otherwise if $\lambda \geq \lambda_1$ which implies that $L''(0) \geq 0$, then $L''(\beta) > L''(0) \geq 0$ for $\beta > 0$. It implies that $L(\beta)$ is convex in $(0, \infty)$. Note that due to continuity, $L'(\beta) < 0$ as $\beta \to 0_+$. Therefore in this case $L(\beta)$ also has only one minimum point in $(0, \infty)$. From $L'(0) < 0$ and $L'(1) > 0$, I have $\beta^* \in (0,1)$.

Step 2. When $\lambda = \lambda_0$, I show that $\beta^* \in (0,1)$ is unique if $\lambda < \lambda_1$ and that $\beta^* = 0$ if $\lambda \geq \lambda_1$.

$\lambda = \lambda_0$ implies that $L'(0) = 0$. If $\lambda < \lambda_1 (\Leftrightarrow L''(0) < 0), L(\beta)$ is first concave and then convex. Therefore there exists only one minimum point, and $\beta = 0$ is not the minimum point. From $L'(1) > 0$ and $L''(1) > 0$, I conclude that the (unique) inflection point and the minimum point are both less than 1 . In other words, there exist a unique $\beta^* \in (0,1)$.

If $\lambda \geq \lambda_1 (\Leftrightarrow L''(0) \geq 0)$, then $L''(\beta) > L''(0) \geq 0$ for $\beta \in (0, \infty)$. Therefore, $L'(\beta)$ is strictly increasing and $L'(\beta) > L'(0) \geq 0$ for $\beta \in (0, \infty)$. It follows that $L(\beta)$ is also strictly increasing for $\beta \in (0, \infty)$. Therefore, $\beta^* = 0$ is the unique minimum point due to continuity.

Step 3. When $\lambda > \lambda_0$ and $\lambda \geq \lambda_1$, I show that $\beta^* = 0$ is unique. Similar as the second part of Step 2, I have $L''(\beta) > L''(0) \geq 0$ for $\beta \in (0, \infty)$ and $L'(\beta) > L'(0) > 0$ for $\beta \in (0, \infty)$. Again, $L(\beta)$ is strictly increasing for $\beta \in (0, \infty)$ and $\beta^* = 0$ is the unique minimum point due to continuity.

Step 4. When $\lambda_0 < \lambda < \lambda_1$, I show that there exists a unique $\lambda_0^* \in (\lambda_0, \lambda_1)$ such that (a) $\beta^* = 0$ if $\lambda > \lambda_0^*$; (b) $\beta^* \in (0,1)$ and unique if $\lambda < \lambda_0^*$; (c) there exist two equilibria, $\beta^* = 0$ and a unique $\beta^* \in (0,1)$, if $\lambda = \lambda_0^*$.

$\lambda_0 < \lambda < \lambda_1$ implies $L'(0) > 0$ and $L''(0) < 0. L(\beta)$ is first increasing and concave near 0 . Then because $L''(\beta)$ is increasing and eventually becomes positive, $L(\beta)$ becomes convex after an inflection point. Because $L(\beta)$ is first concave and then convex, it at most has a local minimum point for $\beta \in (0, \infty)$. The local minimum, if exists, must be in the convex part of $L(\beta)$ where $L''(\beta) > 0$. Note that $L'(1) > 0$ and $L''(1) > 0$. Because $L''(\beta)$ is increasing, for $\beta > 1$, I have $L''(\beta) > L''(1) > 0$. It implies that $L'(\beta)$ is increasing when $\beta > 1$ and, therefore, $L'(\beta) > L'(1) > 0$. It implies that $L(\beta)$ is also increasing in $[1, \infty)$ and $\beta^* < 1$.

Therefore, $L(\beta)$ has at most two local minimum points, $\beta = 0$ or $\beta = \beta_+^* \in (0,1)$. Due to continuity, as $\lambda \to \lambda_0, L'(0) \to 0$, from the first part of Step 2, $\beta^* = \beta_+^*$ and $L(\beta_+^*) < L(0)$. As $\lambda \to \lambda_1$, from Step 3, $\beta^* = 0$ and $L(\beta_+^*) > L(0)$. Due to continuity, there must exist $\lambda_0^* \in (\lambda_0, \lambda_1)$ such that $L(\beta_+^*) = L(0)$. Because $L(\beta_+^*) - L(0)$ is increasing in λ, such λ_0^* is unique. Combining the above four steps, Lemma 2 is proved.

Solving $\boldsymbol{\lambda_0^*}$: I note that λ_0^* in Step 4 can be numerically solved because λ_0^* and $L(\beta_+^*)$ satisfy that $L(\beta_+^*) = L(0), L'(\beta_+^*) = 0$, and $L''(\beta_+^*) > 0$. The first condition implies that when $\beta = \beta_+^* > 0$ and $\lambda = \lambda_0^*$:

$$m^2\sigma_\gamma^2\beta^3 + 2m\rho\sigma_\eta\sigma_\gamma\beta^2 + \left[\sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \lambda m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\right]\beta + \left[-2\sigma_\eta^2 + 2m\lambda\left(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\right)\right] = 0.$$

The second condition implies that:

$$4m^2\sigma_\gamma^2\beta^3 + 6m\rho\sigma_\eta\sigma_\gamma\beta^2 + 2\left[\sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \lambda m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\right]\beta + \left[-2\sigma_\eta^2 + 2m\lambda\left(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\right)\right] = 0$$

Take the difference of the above equation, I have:

$$\beta\left(3m^2\sigma_\gamma^2\beta^2 + 3m\rho\sigma_\eta\sigma_\gamma\beta + \sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \lambda m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\right) = 0$$

$$\Rightarrow \beta_+^* = \frac{-2\rho\sigma_\eta + \sqrt{4\rho^2\sigma_\eta^2 - 3\left[\sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \lambda_0^* m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right)\right]}}{3m\sigma_\gamma}$$

Plug back to the equation implied by the first or the second condition, I can numerically solve a polynomial equation with non-integer powers and represent λ_0^* by the parameters $(m, \mu_\eta, \mu_\gamma, \sigma_\eta, \sigma_\gamma, \rho, \sigma_e)$.

C.3 Proofs of Welfare Analysis

Agent Utility: The agent's utility under the linear allocation rule is:

$$U_a = 2m(\beta\hat{x} + \beta_0) - (x-\eta)^2/\gamma = 2m(\beta(x+e) + \beta_0) - (x-\eta)^2/\gamma$$

I have shown that the agent's best response function is $x = \eta + m\beta\gamma$. Note that the firm's optimal choice of β_0 given β has been shown as $\beta_0^* = \mathbb{E}[\eta - \beta x] - \beta\mu_e$ in Appendix C.1. Therefore, I have:

$$\beta_0^* = (1-\beta)\mu_\eta - m\beta^2\mu_\gamma - \beta\mu_e.$$

Plug the best response function and the optimal β_0 to the equation of U_a I have:

$$U_a = 2m\mu_\eta + 2m\left(\eta - \mu_\eta\right)\beta + m^2\left(\gamma - 2\mu_\gamma\right)\beta^2 + 2m\beta(e - \mu_e).$$

Proof of Proposition 2: Note that $L_s(\beta) = L(\beta) + m^2\mu_\gamma\beta^2$. For $\beta \geq 0$, if $\beta^* = 0$ is the minimum point of L, because it is also the minimum point of $m^2\mu_\gamma\beta^2$, I have $\beta_s^* = 0$.

If $\beta^* > 0$, note that $L_s'(\beta) = L'(\beta) + 2m^2\mu_\gamma\beta$. Therefore, $L_s'(\beta^*) = L'(\beta^*) + 2m^2\mu_\gamma\beta^* > 0$, which implies $\beta_s^* \neq \beta^*$. In addition, by definition of β^*, I have $\forall\beta > \beta^*, L_s(\beta) = L(\beta) + m^2\mu_\gamma\beta^2 > L(\beta^*) + m^2\mu_\gamma\beta^{*2} = L_s(\beta^*)$. This implies $\beta_s^* \leq \beta^*$. In sum, I have $0 \leq \beta_s^* < \beta^*$.

From $L_s'(\beta) = L'(\beta) + 2m^2\mu_\gamma\beta$, I have $L_s'(0) = L'(0)$. If $\lambda < \lambda_0$, from Appendix C.2 I know that $L'(0) < 0$, which implies that $L_s'(0) < 0$. Therefore, $\beta_s^* > 0$.

Proof of Proposition 3: Suppose $\sigma_{e1}^2 > \sigma_e^2$. The firm's loss function of implementing AI with σ_{e1}^2 can be written as follows according to Proposition 1:

$$\begin{aligned} L_e(\beta) &= \mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])^2] + \mathbb{E}[(\mathbb{E}[\eta \mid x] - \eta)^2] + \lambda\mathbb{E}[x^2] + \beta^2\sigma_{e1}^2 \\ &= L(\beta) + \beta^2(\sigma_{e1}^2 - \sigma_e^2) \end{aligned}$$

Denote that $\beta_e^* \triangleq \arg\min_{\beta\geq 0} L_e(\beta)$ and note that I have ready defined $\beta^* = \arg\min_{\beta\geq 0} L(\beta)$. $\forall\beta > \beta^*$, I have:

$$\begin{aligned} L_e(\beta) &= L(\beta) + \beta^2(\sigma_{e1}^2 - \sigma_e^2) > L(\beta^*) + \beta^{*2}(\sigma_{e1}^2 - \sigma_e^2) = L_e(\beta^*), \\ L_e'(\beta^*) &= L'(\beta^*) + 2\beta^*(\sigma_{e1}^2 - \sigma_e^2) > 0. \end{aligned}$$

The first inequality implies that $\forall\beta > \beta^*, \beta_e^* \leq \beta^*$. The second inequality further implies $\beta_e^* < \beta^*$. Therefore, the first part of the proposition is proved. Because in Equation 8 I have shown that the agent's expected utility is decreasing in β, a smaller β^* decreases the agent's expected loss, which proves the second part of the proposition. When the variance of the noise increases from σ_e^2 to σ_{e1}^2, the change in the firm's loss function is:

$$\Delta L_e = L_e(\beta_e^*) - L(\beta^*) = L(\beta_e^*) - L(\beta^*) + \beta_e^{*2}(\sigma_{e1}^2 - \sigma_e^2) \geq L(\beta_e^*) - L(\beta^*) > 0.$$

The last step holds because when $\beta^* > 0$, such β^* is unique according to Lemmas 1 and 2 and because I have already shown $\beta^* \neq \beta_e^*$. Finally, I show the last part of the proposition as follows.

The expected social loss with σ_e^2 is $L_s = L(\beta^*) + m^2\mu_\gamma\beta^{*2}$. When the variance of the noise increases from σ_e^2 to σ_{e1}^2, the expected social loss can be written as follows:

$$L_{s1} = L_e(\beta_e^*) + m^2\mu_\gamma\beta_e^{*2} = L(\beta_e^*) + \beta_e^{*2}(\sigma_{e1}^2 - \sigma_e^2) + m^2\mu_\gamma\beta_e^{*2}.$$

Therefore, the change in the expected social loss is:

$$\Delta L_s = L(\beta_e^*) - L(\beta^*) + \beta_e^{*2}(\sigma_{e1}^2 - \sigma_e^2) + m^2\mu_\gamma(\beta_e^{*2} - \beta^{*2})$$

Note that by definition I have $L'(\beta^*) = 0$ and $L_e'(\beta_e^*) = 0$. Because $L_e'(\beta_e^*) = L'(\beta_e^*) + 2\beta_e^*(\sigma_{e1}^2 - \sigma_e^2)$, I have $L'(\beta_e^*) = -2\beta_e^*(\sigma_{e1}^2 - \sigma_e^2)$.

As $\sigma_{e1}^2 \to \sigma_e^2$, due to continuity, I have $\beta_e^* \to \beta^*$. From Taylor's Theorem,

$$
\begin{aligned}
L(\beta_e^*) - L(\beta^*) &= L'(\beta^*)(\beta_e^* - \beta^*) + \frac{1}{2}L''(\beta^*)(\beta_e^* - \beta^*)^2 + op((\beta_e^* - \beta^*)^2) \\
&= \frac{1}{2}L''(\beta^*)(\beta_e^* - \beta^*)^2 + op((\beta_e^* - \beta^*)^2) \\
L'(\beta_e^*) &= L'(\beta_e^*) - L'(\beta^*) = L''(\beta^*)(\beta_e^* - \beta^*) + op(\beta_e^* - \beta^*)
\end{aligned}
$$

Plug the expansion of $L(\beta_e^*) - L(\beta^*)$ in the above equation of ΔL_s, I have

$$
\Delta L_s = \frac{1}{2}L''(\beta^*)(\beta_e^* - \beta^*)^2 + op((\beta_e^* - \beta^*)^2) + \beta_e^{*2}(\sigma_{e1}^2 - \sigma_e^2) + m^2\mu_\gamma(\beta_e^{*2} - \beta^{*2})
$$

Plug the expansion of $L'(\beta_e^*)$ in $L'(\beta_e^*) = -2\beta_e^*(\sigma_{e1}^2 - \sigma_e^2)$, I have

$$
\beta_e^*(\sigma_{e1}^2 - \sigma_e^2) = -\frac{1}{2}L''(\beta^*)(\beta_e^* - \beta^*) + op(\beta_e^* - \beta^*)
$$

Therefore,

$$
\begin{aligned}
\Delta L_s \quad &= \frac{1}{2}L''(\beta^*)(\beta_e^* - \beta^*)^2 + op((\beta_e^* - \beta^*)^2) - \frac{1}{2}\beta_1^* L''(\beta^*)(\beta_e^* - \beta^*) + op(\beta_e^* - \beta^*) + m^2\mu_\gamma(\beta_e^{*2} - \beta^{*2}) \\
&= (\beta_e^* - \beta^*)\left[\frac{1}{2}L''(\beta^*)(\beta_e^* - \beta^*) + op(\beta_e^* - \beta^*) - \frac{1}{2}\beta_e^* L''(\beta^*) + op(1) + m^2\mu_\gamma(\beta_e^* + \beta^*)\right] \\
&= (\beta_e^* - \beta^*)\left[-\frac{1}{2}\beta_e^* L''(\beta^*) + m^2\mu_\gamma(\beta_e^* + \beta^*) + op(1)\right] \\
&= \frac{1}{2}(\beta_e^* - \beta^*)\beta_e^*\left[-L''(\beta^*) + 2m^2\mu_\gamma(\beta_e^* + \beta^*)/\beta_e^* + op(1)\right]
\end{aligned}
$$

Note that $\beta_e^* - \beta^* < 0$ and $\beta_e^* \to \beta^* > 0$. Therefore, $\text{sign}\,(\Delta L_s) = \text{sign}\,\big(L''(\beta^*) - 2m^2\mu_\gamma(\beta_e^* + \beta^*)/\beta_e^* + op(1)\big) = \text{sign}\,\big(L''(\beta^*) - 4m^2\mu_\gamma\big)$.

Proof of Remark 9: If $m \to 0_+$, from Equation EC.5, given $\beta^* > 0$, I have

$$
L''(\beta^*) - 4m^2\mu_\gamma \to 2\big(\sigma_\eta^2 + \sigma_e^2\big) > 0
$$

If $m \to \infty$, re-write $L'(\beta)$ as a function of m, I have:

$$
\begin{aligned}
L'(\beta) \quad &= \big[4\sigma_\gamma^2\beta^3 + 2\lambda\beta\big(\mu_\gamma^2 + \sigma_\gamma^2\big)\big]m^2 \\
&+ \big[6\rho\sigma_\eta\sigma_\gamma\beta^2 + 4\rho\sigma_\eta\sigma_\gamma\beta + 2\lambda\big(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\big)\big]m \\
&+ 2\big(\sigma_\eta^2 + \sigma_e^2\big)\beta - 2\sigma_\eta^2.
\end{aligned}
$$

If $\beta^* > 0$ and $L'(\beta^*) = 0$ holds, I must have $\beta^* \to 0_+$and $\lambda \to \frac{\sigma_\eta^2}{m(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma)}$. In this case, from Equation EC.5, I have

$$
L''(\beta^*) - 4m^2\mu_\gamma \to 2\left(\sigma_\eta^2 + \sigma_e^2 - 2m\rho\sigma_\eta\sigma_\gamma + \frac{\sigma_\eta^2}{\big(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\big)}\big(\mu_\gamma^2 + \sigma_\gamma^2\big)m - 2\mu_\gamma m^2\right).
$$

Note that the coefficient of m^2 is $-4\mu_\gamma < 0$, as $m \to \infty$, I have $L''(\beta^*) - 4m^2\mu_\gamma < 0$.

Finally, from Equation EC.5, I know that $L''(\beta^*) - 4m^2\mu_\gamma$ can be re-written as a function of μ_γ:

$$L''(\beta^*) - 4m^2\mu_\gamma = C + 2m^2\mu_\gamma\big(\lambda\mu_\gamma - 2\big)$$

where C does not depend on μ_γ and is finite. As $\mu_\gamma \to \infty$, the sign of $L''(\beta^*) - 4m^2\mu_\gamma$ is equal to the sign of $\lambda\mu_\gamma - 2$.

C.4 Proofs of Human-AI Comparison

Proof of Proposition 4: The firm's expected loss:

$$\begin{aligned}
L_h &= \mathbb{E}[(Y_h(x) - \eta)^2 + (\lambda + \kappa_h)x^2] \\
&= \mathbb{E}[(\beta x + \alpha_h x + \beta_0 - \eta)^2 + (\lambda + \kappa_h)x^2] \\
&= \mathbb{E}[(\beta x + \beta_0 - \eta)^2 + \alpha_h^2 x^2 + 2\alpha_h x(\beta x + \beta_0 - \eta) + (\lambda + \kappa_h)x^2] \\
&= \mathbb{E}[(\beta x + \beta_0 - \eta)^2 + (\lambda + \kappa_h + \alpha_h^2)x^2]
\end{aligned}$$

The last step holds because due to the independence of α_h and (η, γ), I have:

$$\begin{aligned}
\mathbb{E}[\alpha_h x(\beta x + \beta_0 - \eta)] &= \mathbb{E}\big[\mathbb{E}[\alpha_h x(\beta x + \beta_0 - \eta) \mid \eta, \gamma]\big] \\
&= \mathbb{E}[\mathbb{E}[\alpha_h \mid \eta, \gamma] x(\beta x + \beta_0 - \eta)] \\
&= \mathbb{E}[\alpha_h]\mathbb{E}[x(\beta x + \beta_0 - \eta)] = 0
\end{aligned}$$

Further, from the first-order condition, I can solve the optimal β_0 :

$$\frac{\partial L_h}{\partial \beta_0} = 0 \Longrightarrow \beta_0^* = \mathbb{E}[\eta - \beta x].$$

In sum, I have:

$$\begin{aligned}
L_h &= \mathbb{E}[(\beta x + \beta_0 - \eta)^2 + (\lambda + \kappa_h + \alpha_h^2)x^2] \\
&= \mathbb{E}[(\beta x + \mathbb{E}[\eta - \beta x] - \eta)^2 + (\lambda + \kappa_h + \alpha_h^2)x^2] \\
&= \mathbb{E}[(\beta x + \mathbb{E}[\eta - \beta x] - \eta)^2] + (\lambda + \mu_\kappa + \sigma_\alpha^2)\mathbb{E}[x^2] \\
&= \mathbb{E}[(Y_0(x) - \mathbb{E}[\eta \mid x])^2] + \mathbb{E}[(\mathbb{E}[\eta \mid x] - \eta)^2] + (\lambda + \mu_\kappa + \sigma_\alpha^2)\mathbb{E}[x^2]
\end{aligned}$$

Note that in Appendix C.1, I have already shown that the last step holds.

Proof of Proposition 7: From Equation 4 and the agent's best response function $x = \eta + m\beta\gamma$, I have:

$$L_h(\beta) = \mathbb{E}\left[\Big(m\beta^2(\gamma - \mu_\gamma) - (1 - \beta)(\eta - \mu_\eta)\Big)^2\right] + (\lambda + \lambda_h)\mathbb{E}[(\eta + m\beta\gamma)^2]$$

Use standard mean-variance decomposition:

$$
\begin{aligned}
L_h(\beta) \quad &= m^2\sigma_\gamma^2\beta^4 + 2m\rho\sigma_\eta\sigma_\gamma\beta^3 + \left[\sigma_\eta^2 - 2m\rho\sigma_\eta\sigma_\gamma + (\lambda+\lambda_h)m^2\left(\mu_\gamma^2+\sigma_\gamma^2\right)\right]\beta^2 \\
&+\left[-2\sigma_\eta^2 + 2m(\lambda+\lambda_h)\left(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\right)\right]\beta + (1+\lambda+\lambda_h)\sigma_\eta^2 + (\lambda+\lambda_h)\mu_\eta^2,
\end{aligned} \tag{EC.6}
$$

From Propositions 5 and 6, if $\lambda_h \geq \overline{\lambda_h} - \lambda, \beta_h^* = 0$. Then, the first part of the proposition holds because:

$$\Delta L_h = L_h(0) - L(\beta^*) = L_h(0) - L(0) + L(0) - L(\beta^*) \geq L_h(0) - L(0) = \lambda_h\left(\sigma_\eta^2 + \mu_\eta^2\right) > 0.$$

Note that $L(0) - L(\beta^*) \geq 0$ due to the definition of β^*. The last step holds due to Equations EC.2 and EC.6.

Next, I prove the second part of the proposition. Because I have shown that if $\beta_h^* = 0$, then $\Delta L_h > 0$, I only need to focus on the case when $\beta_h^* > 0$. From Propositions 5 and 6, let $I(\cdot)$ be the indicator function, $\beta_h^* \in (0, 1 + I(\rho < 0))$ and is unique.

By the definition of β^* :

$$\Delta L_h = L_h(\beta_h^*) - L(\beta^*) = \Delta L_{h1}(\beta_h^*) + L(\beta_h^*) - L(\beta^*) \geq \Delta L_{h1}(\beta_h^*)$$

where $\Delta L_{h1}(\beta) = L_h(\beta) - L(\beta)$. $\Delta L_{h1}(\beta_h^*)$ is a lower bound of ΔL_h. Due to Equations EC.2 and EC.6, I have:

$$\Delta L_{h1}(\beta) = \lambda_h\left\{\left[m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right) - \sigma_e^2/\lambda_h\right]\beta^2 + 2m\left(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\right)\beta + \sigma_\eta^2 + \mu_\eta^2\right\}$$

Note that $2m\left(\rho\sigma_\eta\sigma_\gamma + \mu_\eta\mu_\gamma\right) = 2m\mathbb{E}[\eta\gamma] > 0$. If $m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right) - \sigma_e^2/\lambda_h \geq 0, \forall\beta, \Delta L_h \geq \Delta L_{h1}(\beta_h) > 0$. Therefore, the second part of the proposition holds in this case. If $m^2\left(\mu_\gamma^2 + \sigma_\gamma^2\right) - \sigma_e^2/\lambda_h < 0, \Delta L_{h1}(\beta)$ is a quadratic function. Note that $\Delta L_{h1}(0) = \lambda_h\left(\sigma_\eta^2 + \mu_\eta^2\right) > 0$, due to the property of the quadratic function, $\Delta L_{h1}(\beta) = 0$ has one and only one positive root denoted as $\overline{\beta_h}$ such that for $\beta \in \left(0, \overline{\beta_h}\right]$, $\Delta L_{h1}(\beta) \geq 0$ and for $\beta \in \left(\overline{\beta_h}, \infty\right), \Delta L_{h1}(\beta) < 0$. Therefore, a sufficient condition for $\Delta L_{h1}(\beta_h^*) \geq 0$ is

$$
\begin{aligned}
&\Delta L_{h1}(1 + I(\rho < 0)) \geq 0 \Leftrightarrow \\
&\frac{\sigma_e^2}{\lambda_h} \leq \frac{\mathbb{E}[\eta^2]}{[1 + I(\rho < 0)]^2} + \frac{2m\mathbb{E}[\eta\gamma]}{1 + I(\rho < 0)} + m^2\mathbb{E}[\gamma^2] = \mathbb{E}\left[\left(\frac{\eta}{1 + I(\rho < 0)} + m\gamma\right)^2\right]
\end{aligned}
$$

The above inequality directly implies that holding other parameters constant, $\Delta L_{h1}(1 + I(\rho < 0)) \geq 0$ (AI systems are preferred) if any one of $\left(m, \mu_\gamma, \sigma_\gamma, \mu_\eta, \sigma_\eta\right)$ is sufficiently large.

Proof of Proposition 9: When $m \to 0$, I evaluate the first-order conditions $L_h'(\beta_h^*) = 0$ and $L'(\beta^*) = 0$. From Equations EC. 6 and EC.2, I have $\beta_h^* \to 1$ and $\beta^* \to \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_e^2} < 1$. Notably, this implies that the agent's expected utility is lower when the firm uses human service systems than AI systems. Further, plug β_h^* and β^* into the equation of $L_h(\beta)$ and $L(\beta)$, I have:

$$L_h(\beta_h^*) \to (\lambda_h + \lambda)\left(\sigma_\eta^2 + \mu_\eta^2\right)$$

$$L(\beta^*) \to \frac{\sigma_\eta^2 \sigma_e^2}{\sigma_\eta^2 + \sigma_e} + \lambda\left(\sigma_\eta^2 + \mu_\eta^2\right)$$

$$\Delta L_h = L_h(\beta_h^*) - L(\beta^*) \to \left(\sigma_\eta^2 + \mu_\eta^2\right)\left[\lambda_h - \frac{\sigma_e^2}{\left(\sigma_\eta - \mu_\eta \sigma_e/\sigma_\eta\right)^2 + \left(\sigma_e + \mu_\eta\right)^2}\right].$$

Therefore, $\Delta L_h \geq 0 \Leftrightarrow \lambda_h \geq \frac{\sigma_e^2}{\left(\sigma_\eta - \mu_\eta \sigma_e/\sigma_\eta\right)^2 + \left(\sigma_e + \mu_\eta\right)^2}$.

VITA

Yifan Yu is a researcher in the intersection of Economics, AI, and Data Analytics. His research mainly focuses on the connection between algorithms and customer behavior. He designs and implements algorithms to derive economic and managerial implications from online user- and marketer-generated unstructured data, including text, images, videos, and networks. He is especially interested in the interaction between customers and Emotion AI, the algorithm that recognizes, interprets, and responds to various human emotions. His research methodologies include machine/deep learning, affective computing, causal inference, lab/field experiments, econometric modeling, and game theories. He received his Ph.D. degree at the Michael G. Foster School of Business, University of Washington (UW). Prior to joining UW, he completed his undergraduate and graduate study at School of Economics and Management, Tsinghua University. He will be joining the Department of Information, Risk, and Operations Management at McCombs School of Business, University of Texas at Austin (UT Austin) as an Assistant Professor as of September 2023.