ANALYSIS OF DESIGN ELEMENTS IN THE MACHINE-PLATFORM-CROWD TRANSFORMATION

by

Yipu Deng

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Krannert School of Management

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Karthik N. Kannan, Co-Chair

Krannert School of Management

Dr. Jinyang Zheng, Co-Chair

Krannert School of Management

Dr. Zaiyan Wei

Krannert School of Management

Dr. Warut Khern-Am-Nuai

McGill University

Approved by:

Dr. Yanjun Li

ACKNOWLEDGMENTS

First, I want to express my gratitude to Professor Karthik Kannan and Professor Jinyang Zheng, who served as my advisors. I'd like to offer my heartfelt gratitude to them who have served as, and continue to serve as, incredible mentors to me. Throughout the development of this dissertation, they provided me with unwavering support and important advice.

I'd like to express my gratitude to the other members of my dissertation committee, Professor Zaiyan Wei and Professor Warut Khern-Am-Nuai, for their invaluable assistance and advice.

I'm also grateful to other professors who took the time to talk with me about my research and provide constructive feedback on my dissertation papers.

I'm grateful to my fellow Purdue students and friends for sharing research and other aspects of the Ph.D. experience.

Finally, I'd like to convey my gratitude to my parents for their unwavering support throughout my life.

TABLE OF CONTENTS

ABSTRACT

Digital transformation greatly affects all segments of our society. There are three powerful trends unleashed by the digital revolution: machine, platform, and crowd. The first trend emphasizes that machine learning can either complements or supplements human capabilities, which leads to data-driven decision making. The second trend shows that value creation is moving from physical products to platforms (e.g., Uber and Airbnb) where network effects can have a great impact. The third trend is about the emergence of online crowds. Several good examples are crowdfunding platforms like Indiegogo and collaborative platforms such as Wikipedia. My research work studies these three trends from different aspects.

In the first project, we investigated how professional reviewers influence subsequent non-editor reviewers in their writing behaviors. Restaurants that receive editorial reviews are found to have reviewers who not only post more frequently, but also give lengthier and more neutral feedback. Further investigation of the mechanism finds that in terms of the topics, sentiment, and readability, following reviews of restaurants that receive editorial reviews become increasingly similar to their editorial reviews, indicating that a herding effect is the main driver of the shift in later reviews. In this study, we not only look at quantitative review characteristics such as rating and review length, but also extract qualitative review characteristics embedded in review text using Natural Language Processing (NLP) techniques (e.g., Topic modeling and Sentiment analysis).

In the second project, we studied how AI-based shelf monitoring can help manufacturers with their shelf management efforts. In general, we've discovered that AI-powered shelf monitoring boosts product sales. We further reveal that the positive effect shall be attributed to independent retailers rather than chained retailers. More broadly, the finding further suggests that AI-powered monitoring is more scalable, allowing manufacturers to cope more effectively with more heterogeneous objects. In this study, we analyzed shelf photos using deep learning (e.g., image recognition). Furthermore, we conducted a qualitative study (i.e., interviews) as a supplement attempt to uncover the underlying mechanism behind the interesting phenomenon found in our field experiment.

In the third project, we tried to understand the dynamic contribution patterns caused by backers' multiple roles and fundraisers' strategic behaviors. We show that projects described by more subjective content (i.e., title and introduction) significantly repel potential donors. We further show that fundraisers' contribution to their own projects might increase donor' intention to donate and has no significant impact on reward pledging of subsequent backers. Above that, we find a positive interplay between donation and reward pledge, suggesting a cross-channel peer influence that will facilitate the fundraising progress.

1. INTRODUCTION

Digital transformation refers to the adoption of digital technology that transforms and improves services or businesses. Such transformation has a great influence on our daily lives, companies and organizations, industries and various segments of our society. With recent advances in digital technologies, there are three powerful trends unleashed by the digital revolution: machine, platform, and crowd (McAfee and Brynjolfsson 2017). The first trend emphasizes that machine learning can either complements or supplements human capabilities. The second trend shows that value creation process is moving from physical products to online platforms such as Uber and Yelp. The third trend is about the emergence of online crowds with capitals, knowledge, or both who are able to complement or substitute core experts in companies.

Knowing that the machine-platform-crowd transformation is reshaping the business world, it is necessary to understand the dynamics of human behaviors in such transformation if we want to take advantage of digital technologies as well as make the difference between thriving and surviving. My dissertation seeks to contribute to this theme. Specifically, my dissertation studies human behaviors in the machine-platform-crowd transformation from different aspects.

Speaking of the platform where network effects can have a dramatic influence, we work with an online review platform and analyze strategies that a platform can use to shape users' review writing behaviors. In terms of machine, we collaborate with a fast-moving consumer goods manufacturer to study how people change their behaviors in response to Artificial Intelligence (AI) based intervention. In addition, we investigate the circumstance where the performance of machines can surpass human's in the specific context of shelf management. On the part of the crowd, we carefully look at people's contribution dynamics on a crowdfunding platform.

The three essays are summarized as following. The first essay investigates an editorial review program where a review platform supplements the user reviews with editorial ones written by professional writers. Specifically, we examine whether and how editorial reviews influence subsequent user reviews (i.e., reviews written by non-editor reviewers). In this research, we collaborate with a large restaurant review platform in Asia. We leverage a quasi-experiment research design in which our focal review platform launched the editorial review program while everything else remained the same. We first examine the impact of editorial reviews on basic characteristics of subsequent user-generated reviews including review volume, review length, and review valence. Subsequently, we incorporate natural-language processing (NLP) methods to examine further the textual content of the reviews and quantify the impact of editorial reviews on the content-related features of following user reviews. Thereafter, we leverage the theory of herding behavior to examine the mechanism of editorial reviews that drives the change in user review characteristics. In particular, we build similarity measures to compare the content differences, from multiple dimensions, between editorial reviews and user reviews before and after the existence of editorial reviews. We find an overall positive effect of editorial reviews on subsequent user reviews from the platform's perspective. For restaurants that receive editorial reviews, reviewers not only post more frequently, but also write longer and more neutral feedback. Further analysis of the mechanism reveals that the subsequent reviews of the restaurants that receive editorial reviews become more similar to their editorial reviews regarding the topics, sentiment, and readability, indicating a herding effect as the main driver of the change in the subsequent reviews. The findings suggest that review platforms could use an editorial review program to not only boost the review quantity, but also manage the content's quality. By supplementing high-quality editorial reviews with user reviews, the platform can improve the overall content quality of user reviews through a herding effect.

The second essay studies whether the AI-powered shelf monitoring system could build up manufacturers' capability to improve shelf management in emerging markets. From the theoretical perspective, we further explore when and to which objects the AI-powered shelf monitoring system could generate incremental values comparing to the human delegates' monitoring. Our collaborative research with the leading fast-moving consumer goods manufacturer includes three stages. In Stage 1, we address the overall impact and economic value for launching AI-powered shelf monitoring of product sales by using a quasi-experiment scheme and analyzing observational data. In Stage 2, we conduct a randomized field experiment to more precisely quantify the causal effect of launching AI-powered shelf monitoring;

to do so, we include an additional treatment group that first is subject to AI-powered shelf monitoring and thereafter is terminated. This design is consistent with the literature on monitoring (e.g., Staats et al. 2016) and allows us to examine the persistence of AI-powered monitoring, so we may understand the long-term adoption effect of this program. In Stage 3, we attempt to disentangle the underlying mechanism, and this process consists of four parts. First, we assume that the incremental value derived from AI-powered monitoring comes from the higher scalability of AI in dealing with heterogeneous objects as compared to solely using humans, and this higher scalability allows delegates to expand their monitoring scope to more heterogeneous retailers. Therefore, we expect a more salient impact when using AI-powered monitoring, as the correspondingly monitored shelves will be more heterogeneous. We test this assumption by looking at a heterogeneous treatment effect of AI-powered monitoring considering differentiated store types by their relative degree of contract heterogeneity. Second, we assume that the observed increase in product sales results from retailers' better compliance with shelf display requirements, and we therefore verify this expectation by using shelf photos to analyze the change in retailers' compliance. Third, we tease out the effect of only taking shelf photos. Finally, we qualitatively examine our quantitative findings and our underlying mechanism by interviewing the manufacturer's delegates. We consistently find that retail stores that implement AI-powered shelf monitoring programs boost their product sales. However, the gains in incremental sales drop after the termination of the program. We further find that improved sales derived from AI-powered shelf monitoring are largely attributed to independent retail stores, which have more heterogeneous contracts and are more dominant in emerging markets. This finding supports the scalability of AI and shows how this scalability can be expanded to repetitive tasks with heterogeneous objects, which are typically more challenging for humans to complete. We also explores the causal chain between AI use and improved sales. Specifically, we hypothesize that the use of AI could lead to more effective monitoring by delegates, and that this improved monitoring could result in retailers' better compliance with shelf display requirements, which would yield improved sales. Finally, we find that the benefit of launching this AI-powered shelf monitoring system comes from AI rather than from taking photos. In other words, a placebo monitoring system that only takes shelf photos only will not boost sales.

The third essay scrutinizes the dynamic contribution patterns caused by backers' multiple roles and fundraisers' strategic behaviors in the context of crowdfunding. To be specific, we study the dynamics of donating, reward pledging, fundraiser's strategic behaviors, and their interplays on a crowdfunding platform that integrates reward pledge with donation and fundraiser's investment. In the first step of our analysis, we determine the particular project characteristics that backers value most and thus influence the aggregate demand of the projects. Beyond the directly observable quantitative characteristics provided on the project page (e.g., remaining budget, number of total backers, etc.), potential backers also tend to value project quality characteristics embedded in the project description, such as the subjectivity of project title and project introduction. We incorporate natural language processing (NLP) techniques to infer these textual features. In the second step, following Berry et al. (1995) we use demand estimation techniques (i.e., the random coefficient logit model) to quantify the economic influence and relative importance of project characteristics. Furthermore, we quantify the impact of peer's crowdfunding behaviors and fundraisers' strategic behaviors on contribution dynamics. Our analysis reveals several notable and interesting findings. First, we find evidence in support of the herding effect. As individuals observe others pledging/donating more money, the amount they are inclined to pledge/donate increases. Surprisingly, empirical evidence of positive cross-channel peer influence between donating and reward pledging is found. That is, more prior donating behaviors could lead to more reward pledging behaviors and vice versa. Second, projects with more subjective titles and introductions will repel potential donors. One possible explanation is that donors are more skeptical of subjective than objective narratives while evaluating projects (Susan and David 2010). Potential donors may perceive projects with the subjective description as less trustworthy because those projects do not convey much factual information that is useful. This finding yields a significant managerial insight: fundraisers could provide project narratives with more objective information to attract more donors and get more donations. Third, we find that if a fundraiser contributes to his/her own project, it will increase backers' intention to donate. One possible reason is that fundraisers' contributions can arouse donors' sympathy. At the same time, since reward buyers pay more attention to the creditworthiness of fundraisers and the quality of future products/services (Zhang and Liu 2012), fundraisers' contribution has no significant impact on the reward pledging of subsequent backers.

Putting together, the three studies in my dissertation have shown that all successful companies need to understand, predict, and shape human behaviors in the machine-platform-crowd transformation. All studies are proof-of-concept of a central theme: firms can apply appropriate interventions and thus guide those people with complicated behaviors driven by digital technologies. Overall, the findings from the three studies will provide valuable insights for platforms and social enterprises on how to utilize the trends of digital transformation. The rich data from collaboration allows me to test the underlying mechanism at work. In this way, my dissertation provides both managerial implication and theoretical contribution to the dynamics of human behaviors in the machine-platform-crowd transformation.

The rest of the dissertation is structured as follows. Chapters 2,3, and 4 discuss the first, second, and third essays of my dissertation as described above. Chapter 5 provides a summary of the main findings.

2. MORE THAN THE QUANTITY: THE VALUE OF EDITORIAL REVIEWS FOR A UGC PLATFORM

2.1 Introduction

Online user-generated content (UGC), such as consumer reviews, plays an important role in the decision-making process of shoppers and thus creates a direct economic impact [e.g., 1]. With the advent of technology, review platforms have benefited from a "wisdom of the crowd" effect, whereby peer consumers provide a significant amount of feedback. Not surprisingly, platform owners have employed various strategies to improve the quantity and quality of the review content, which has included offering extrinsic incentives [2], using intrinsic social norms [3], or creating social networks [4]. In addition, one of the techniques that has gained traction in recent years among online review platforms is to incorporate expert reviews. These reviews are generated by the third-party experts who are neither directly associated with the seller nor existing consumers, such as platform-hired professional editors/testers (e.g., Qunar.com), solicited guest testers from platform users who are recognized by their past expert-quality review generation (e.g., Amazon, Yelp, Dianping, FlyerTalk, Taobao, JD.com), or contracted experts/critics in the domain (e.g., Rotten-Tomatoes).

In this study, we focus on one specific type of such expert review program, namely, editorial review. As compared to other expert review programs (e.g., the invited critics on Rotten Tomatoes) that usually allow expert reviewers/critics to have a certain degree of freedom (i.e., how to review and which product to review), the editorial review program tends to be more restrictive, whereby expert reviewers serve as the editors affiliated with and managed by the platforms. It endows the platform with additional controls regarding how to organize the review content and to which product to provide such reviews. In practice, platforms that employ the editorial review program typically hire and train editors internally (e.g., Qunar's reviews by professional testers) or selectively invite high-quality reviewers as guest editors and solicit them to review targeted merchants/products/services through invitation-based and complimentary/paid testing programs (e.g., Amazon Vine Program, elite programs of Yelp and FlyerTalk, invited testing programs of Dianping, Taobao, and JD.com). Reviews generated under such an affiliation/management relationship are usually displayed with an explicit label, such as "editorial review" or "invited review," on the review page.

Theoretically, there exist competing tensions that may influence the implications of editorial reviews on the platform. On the one hand, these editorial reviews may augment the content of the review platform by providing discussion points and thus stimulate user participation. On the other hand, editorial reviews also may unintentionally discourage users by triggering substitution effects (i.e., deter users from contributing similar reviews) or by being perceived as marketer-generated content. In this research, we resolve these competing tensions by investigating empirically how the presence of editorial reviews affects subsequent consumer reviews in terms of content quantity and quality.

Our research agenda is important and relevant to research and practice. First, although the impact of existing user-generated reviews on subsequent reviews has been well studied [e.g., 5]–[8], these results may not be directly applicable to analyzing the implications of editorial reviews, as consumer reviews and editorial reviews are vastly different. For example, although editorial reviews, in general, are of high quality because experts have incentives to keep their reputation, this is usually not the case for consumer reviews [3]. In addition, existing research has demonstrated that review readers may react to these two types of reviews differently [9]. As a result, editorial reviews are expected to have a different impact than are consumer reviews on subsequent reviews, which is an issue that has not been studied in the literature.

To operationalize our research agenda, we collaborate with a large restaurant review platform in Asia to study the impact of editorial reviews. We leverage a quasi-experiment research design in which our focal review platform launched the editorial review program while everything else remained the same. We first examine the impact of editorial reviews on basic characteristics of subsequent user-generated reviews, including volume, length, and valence. Next, we incorporate natural language processing (NLP) methods to examine textual content of subsequent user reviews and to quantify the impact of editorial reviews on the content-related features of those reviews. We then leverage the theory of herding and differentiation behavior to explain and examine the mechanism of editorial reviews that drives the change in users' review-generating behavior. In a further examination, we develop similarity measures to compare the review-level content differences from multiple dimensions between editorial reviews and user reviews before and after the existence of editorial reviews. Finally, we develop a feasible identification strategy that leverages long-term users of the platform to isolate the change in their review-writing behavior that is driven by editorial reviews.

Our analyses reveal several notable findings. First, restaurants that receive editorial reviews enjoy an increase in review volume afterward, suggesting herding behavior from platform users toward editorial reviews in choosing which restaurants to review. In addition, in regard to the review content, user reviews posted after editorial reviews are longer and have a higher degree of content variety due to increased substantive content, such as discussions on food, desserts, and drinks. Notably, the topic distribution of user reviews gradually gravitates toward that of editorial reviews. Further, the existence of editorial reviews greatly affects the sentiment of subsequent user reviews as well. Specifically, user reviews become more neutral and are associated with lower rating valences after the editorial reviews are posted, also exhibiting gravitational effects toward the sentiment of editorial reviews. In this regard, our review-level mechanism analysis of similarity measures demonstrates quantitatively that the content of user reviews becomes similar to that of editorial reviews in all aspects (i.e., topic, sentiment/rating, length, and readability). In other words, with the editorial reviews present, review writing on the platform gradually tends toward the content of editorial reviews. We also conduct a reviewer-level analysis that leverages the long-term reviewers. Such an alternative identification strategy confirms that editorial reviews trigger the behavioral changes of the reviewers, suggesting the existence of a herding effect (i.e., users follow professional editors) in their review-writing behavior.

Our study offers several key contributions to the literature and practice. First, it contributes to the literature stream on herding or differentiation effects in the context of UGC. Although prior research has thoroughly examined herding and differentiation among users, we further explore the effects between users and platform editors, which helps to improve our understanding of review-generating behaviors. Given that it is particularly difficult for review platforms to manage the natural behavior of users, our work provides tangible practical insights on the herding behavior that occurs around the content generated by the platform. Second, this research contributes to the literature on third-party expert reviews by investigating how they affect word of mouth (WOM) in the format of consumer reviews. Because previous literature has demonstrated empirically a significant economic impact of WOM [e.g., 10], the insights on the impact of editorial reviews on WOM in multiple dimensions could be extrapolated to identify a potential indirect influence on the economic outcomes, which is increasingly important with the prevalence of digitized UGC and electronic WOM (eWOM). Third, our results yield important insights for managers of UGC platforms. Although numerous studies have examined the effectiveness of various managerial strategies to improve UGC, these strategies usually target at either content quantity [e.g., 11] or content quality [e.g., 12]. Some of these strategies may increase the quantity at the expense of lowered quality [e.g., 2], [13]. Our study considers editorial reviews as a viable strategy to manage UGC and shows its effectiveness in not only increasing the production of UGC but also improving the content quality. The insights of editorial reviews could help platform managers to derive content management policies that contribute to the platform's long-term viability and sustainability.

The remainder of the paper is organized as follows. In §2.2, we survey the literature related to our study. In §2.3, we describe the empirical context of this study, along with our dataset. In §4, we present the identification strategy and empirical models. In §5, we provide the results and findings. We conduct additional robustness checks, which are included in §6. Finally, in §7, we discuss the theoretical and practical contributions of this study and the limitations and future research opportunities as well as conclude the paper.

2.2 Literature Review

In this section, we survey the literature that is closely related to this study. In particular, we review previous works that examine herding and differentiation behavior. We then discuss studies that investigate third-party evaluation, followed by those that focus on how review platforms manage UGC.

2.2.1 Herding and Differentiation Behavior

Herding effects, known as an individual's tendency to follow actions taken by previous decision makers, have been studied empirically in several contexts, including microloan markets [14], restaurant dining [15], and online flash sales [16]. In those contexts, herd behavior is shown to be a consumer choice (i.e., whether to choose a product or which product to choose), which is then aggregated to changes in sales volume. Herding behavior is also known as the bandwagon effect [17], whereby individuals tend to choose the one chosen (more) by prior individuals. In the specific context of UGC, herding behavior also is used to explain how existing UGC affects subsequent content generation. For example, [8] finds that reviewers tend to decrease their product evaluations if they observe negative opinions expressed by others. Similarly, [6] show a bandwagon effect in reviewers' writing behavior: Reviewers tend to conform to the majority by adjusting their product evaluations upward in more positive review environments. Such herding effects are found to be especially prominent when reviewers share close social ties [5].

Conversely, differentiation effects are individuals' tendency to distinguish themselves from others. In the context of UGC, differentiation effects are used to explain the cases in which reviewers adjust their product evaluations such that they are different from others' evaluations. For instance, [7] provide empirical evidence that reviewers strive to differentiate their reviews from existing ones that consider rating dynamics. Likewise, [8] argues that self-presentation concerns lead reviewers to differentiate their reviews. In addition, differentiation effects influence consumer decisions on whether to contribute reviews by discouraging them from reviewing products with high review volume [6]. Such an effect occurs because new receivers perceive that their opinions could be inconsequential due to the number of existing reviews.

Although past research has thoroughly examined herding and differentiation in the context of UGC among end users, it is particularly difficult to derive practical insight into how to leverage these effects to improve review platforms, as review platforms cannot directly manage consumer reviews or intervene in consumers' review-writing process to take advantage of natural herding/differentiation behavior. In this regard, we extend this stream of literature by exploring users' herding or differentiation behavior with respect to platform editors, who are more subject to the platform's control. In particular, we examine whether the content of editorial reviews could influence the content-generating behavior of users and use the herding or differentiation effect to explain the underlying mechanisms. By doing so, we provide review platforms with practical implications for utilizing social influence theory, as platforms could centrally manage editorial reviews and use editorial reviews to stipulate users' content-generating behavior. Next, as editorial reviews are considered one of the third-party expert review programs, we review the literature on third-party evaluations.

2.2.2 Third-party Evaluations

Traditionally, third-party evaluations often refer to expert evaluations conducted by a third party that have no official association with the party being accessed [18]. Most studies of third-party evaluations in this form, such as certification and licensing, examine their impacts on traditional market outcomes. For example, [19] show that third-party evaluations can increase consumer demand and improve seller quality. Similarly, [20] finds that third-party evaluations can help to reduce information asymmetries and thus increase demand from buyers who otherwise may not have entered the market.

Our work connects to an emerging body of research that examines the role of third-party evaluations in an online review context. The literature has documented significant impacts of third-party reviews on consumer behavior. For instance, [21], [22], and [23] show that film critics can greatly affect the popularity of movies and can thus serve as a good predictor of movie box office revenue. Likewise, [24] find that newspaper critics have a significant impact on consumers' attitudes toward Broadway shows, with critics from popular newspapers such as *The New York Times* as yielding much more influence than those featured in relatively unpopular newspapers. [25] show that third-party evaluations can effectively reduce users' difficulties in evaluating, which can lead to a higher product adoption rate. Further, [26] show that, when the sentiment of third-party reviews affects consumers' product evaluations, the relevance between review topics and customer segments also plays a pivotal role in affecting consumer purchasing behavior. In this regard, [27] emphasize the importance of third-party reviews by showing that manufacturing firms should adapt their marketing strategies in response to third-party reviews.

Although prior studies have advanced our understanding of the effects of third-party evaluations on market outcomes (e.g., consumer demand), the implications and underlying mechanisms of third-party evaluations in the online review context remains understudied. Potentially, third-party reviews could have a salient influence on consumer-generated reviews. Indeed, the prevalence of online review platforms suggests an even stronger impact of third-party reviews if the review exhibits an effect on eWOM, such as UGC. Such an effect warrants investigation. In this regard, our research examines empirically whether editorial reviews, a form of third-party reviews, influence subsequent consumer reviews in terms of content quantity and quality. Below, as editorial reviews in the context of this research are under the control of a review platform, we survey the literature on the practice used by review platforms to manage UGC.

2.2.3 Review Platform's Management of UGC

Prior studies that have examined the review platform's management of UGC can be broadly classified into three substreams of research. The first substream concerns platform strategies to enhance the consumer's review-reading experience. For instance, many review platforms provide helpfulness information (i.e., helpful and unhelpful votes) alongside each review. This metric not only serves as a primary index for consumers to filter out low-quality reviews but also sorts and positions the most helpful reviews, both positive and negative, more prominently on the review pages [28]. In addition, because reviews written by opinion leaders are perceived as more influential than are others, many review platforms reward prolific reviewers with badges to signal their status to other users [29]. Further, to reduce fake reviews, some review platforms (e.g., Expedia.com) allow only those customers who transact on the platform to post user reviews [30], whereas other platforms (e.g., TripAdvisor) implement review filters to detect and eliminate suspicious content [31].

The second research substream focuses on feasible avenues for review platforms to increase UGC production. Given the consistent findings of the positive effects of review volume on economic outcomes [e.g., 7], [10], [32], it is important for review platforms to understand the antecedents of UGC production and to develop mechanisms that foster reviewer participation. Studies show that the motivational factors for UGC production vary widely based on reviewers' characteristics. The examples of these factors are audience group size [11], perceived identity verification [33], and community commitment [34]. Based on these factors, several platform interventions are designed to stimulate UGC production. For instance, [35] suggest the use of financial incentives to increase the likelihood of feedback. [36] show that the use of social norms (i.e., providing users with peers' contribution activity) can have a positive influence on the production of movie ratings. Online performance feedback interventions also are shown to potentially drive UGC contributions [37].

Third, prior works have investigated possible methods that review platforms can employ to improve the quality of review content. In particular, researchers have investigated the antecedents of UGC's qualitative characteristics on several dimensions, such as content length and linguistic features (e.g., sentiment, readability). For example, [4] find that allowing a subscription or "following" could be an effective strategy to improve review content because the contributor's popularity (e.g., number of subscribers) is shown to be correlated with the readability of the reviews. [38] support this strategy by showing that social ties (e.g., social network neighbors) can increase the similarities between the content that people create online. Further, review platforms could use the combination of financial incentives and social norms [3] or performance-contingent monetary incentives [39] to motivate reviews of greater length.

Our research complements this body of literature by exploring the strategic value of editorial reviews from the platform's perspective. To examine the effectiveness of this strategy, we investigate its effects not only in regard to improving user review production but also on the quality of subsequent review content. In particular, we study the impact of editorial reviews on subsequent user review volume, rating, length, topics, sentiment, and readability. Further, we reveal the underlying mechanisms for the change of reviewers' writing behaviors that are directly influenced by editorial reviews.

2.3 Research Context and Data

In this section, we present the empirical context of our study and discuss the descriptions of our data and variables of interest.

2.3.1 Research Context

In this research, we collaborate with a large restaurant review platform in Asia that provides services that are similar to those of other major review platforms in United States, such as Yelp. The design of its website with anonymized details is provided in Figure 1. Each restaurant on the platform has a dedicated page that contains detailed information about the restaurant (e.g., address, menus, opening hours), as well as reviews and ratings contributed by platform users, sorted by posting date (i.e., the most recent review is displayed on the top of the list). Users who register on the platform with a valid email address can submit reviews and ratings. A typical review generated by users consists of a restaurant rating (on a scale of 1 to 5), the textual content of the reviews, and any relevant photos. The platform also features a peer evaluation system that allows users to vote for other reviews. All reviews are stored and published by the review platform.

In addition to user-generated reviews, the platform started to provide reviews written by professional editors (i.e., editorial reviews) in late 2010. The editorial review program was launched to serve various purposes. First, the platform recognizes that user-generated reviews alone tend to suffer from reporting bias [40]. Therefore, consumers who read user-generated reviews may experience some disparities between the actual restaurant quality and the perceived quality based on the reviews. Hence, the platform intends for the editorial reviews to be objective and comprehensive evaluations of restaurant quality to balance the sentiment in user-generated reviews. Second, the comprehensiveness of editorial reviews could account for features, topics, or aspects of the restaurants that are missing or rarely discussed in user-generated reviews. In the same way, the editorial reviews could reduce noise that occurs in the user-generated reviews.

To ensure the quality of editorial reviews, the editors hired by the platform are professional editors/critics who have received sufficient training on the expectations for their

Figure 2.1. User Review and Editorial Review Examples

contributions. In addition, before posting editorial reviews, the platform cross-checks each editorial review to ensure that the content satisfies the platform's editorial review standard (i.e., that the review is reasonably objective and comprehensive). As we will demonstrate in the next subsection, the summary statistics of posted editorial reviews confirm the effectiveness of such practices. Specifically, editorial reviews are comprehensive in the covered topics, as they tend to substantially discuss food-related aspects (e.g., food courses, desserts, drinks). In addition, they are neutral in sentiment and of a substantial length (see Table 2.4 for details).

Due to the limited number of professional editors, the platform is unable provide editorial reviews for all restaurants. The platform managers are tasked to select restaurants to receive editorial reviews based on their business characteristics, including popularity, food category, price range, region, and so forth. Notably, the platform also favors restaurants that sign advertising contracts with the platform to receive editorial reviews.[1] Subsequently, the man-

[1]↑Among 5,158 restaurants with advertising contracts, 958 have received editorial reviews (about 1/5). In addition, among 139,983 restaurants without advertising contracts, 1,325 have received editorial reviews (about 1/100).

agers assign available professional editors to visit those restaurants as ordinary consumers and to write corresponding editorial reviews after the visit. In total, 2,283 out of 145,141 restaurants on the platform have received editorial reviews.

The format of editorial reviews is generally consistent with that of user-generated reviews, as shown in Figure 2.1. Specifically, both types of reviews offer some generic information in regard to reviewers (e.g., screen name, profile picture), while the detailed descriptions of the restaurants are provided as textual content. Nevertheless, there are some differences between the two. First, editorial reviews are clearly marked as "Editorial." Second, no numeric ratings are associated with the editorial reviews, as the platform aims to provide reviews with objective evaluations. Third, editorial reviews, if they exist, are placed on the top of the review pages (i.e., above all user-generated reviews). Note that one restaurant may receive more than one editorial review, which may or may not be written by the same editor. In that case, only the most recent one appears at the top, whereas the former ones will be pooled with user reviews and sorted chronologically as of the posting time.

2.3.2 Data

In this subsection, we present the dataset used in this study. We begin by describing the data and variables of interest. Then, we explain NLP techniques that we apply to textual content in our dataset.

Data Description and Variable Definitions

Through the collaboration, we obtain access to the complete review-level data of 145,141 restaurants, including all user-generated reviews from July 1, 2010, to May 30, 2017, and all editorial reviews during this period. Restaurants in the dataset started to receive editorial reviews in late 2010. We check that none of the restaurants has an editorial review as its first review (i.e., before any user reviews). The dataset contains the information that pertains to online reviews, including the review's textual content, date, rating, and so forth. Further, we have access to the generic information for all restaurants on the platform, including whether a restaurant has signed an advertisement contract with the platform. Using this dataset, we conduct two sets of analyses in our study.

First, we focus on simple review characteristics that are not embedded in the textual content. Following [10], we examine three basic review characteristics: (1) *ReviewVolume*, (2) *Rating*, and (3) *Length*. We construct our dataset as panel data at the restaurant level such that each observation is a restaurant and each time period is a month. Therefore, $ReviewVolume_{\mathrm{it}}$ denotes the total number of user reviews that are posted for the restaurant i in month t. $Rating_{\mathrm{it}}$ represents the average rating of reviews for restaurant i in month t. $Length_{\mathrm{it}}$ is the average number of review characters for restaurant i in month t.

Second, to better capture characteristics of the content of reviews, we extend our study beyond the surface of the aggregated review characteristics by focusing on the textual content. Here, the unit of analysis becomes the review level. We follow [41] to extract three major qualitative metrics from the textual content of each review (editorial and user-generated reviews) by using NLP techniques. The qualitative metrics extracted are: (1) topics covered in reviews, (2) review sentiment (i.e., the fraction of negative, positive, or neutral emotion in the review text), and (3) readability (e.g., Simple Measure of Goobledygook (SMOG) index). We provide more details on the construction of these variables in §2.3.2.

The independent variables and control variables are tailored to the two aforementioned levels of analyses. Our main independent variable of interest is the treatment indicator, $Editorial\ Review_{\mathrm{i}(k)t}$, which indicates whether an editorial review has been provided for restaurant i by (the time when posting the k^{th} user review in) month t. Similarly, our control variables include:

- $Ad_{\mathrm{i}(k)t}$: whether restaurant i is in an advertisement contract with the review platform by (the time when posting the k^{th} user review in) month t
- $PriorVol_{\mathrm{i}(k)t}$: the cumulative user review volume for restaurant i by (the time of posting the k^{th} user review in) month t
- $PriorERVol_{\mathrm{i}(k)t}$: the cumulative editorial review volume for restaurant i by (the time of posting the k^{th} user review in) month t

- $PriorRate_{\mathrm{i}(k)t}$: the cumulative rating valence for restaurant i by (the time of posting the k^{th} user review in) month t

We also include the reviewer experience of the k^{th} user review in month t, $Experience_{\mathrm{i}kt}$, measured as the duration between t and her first posting time (in month), as a control variable in the mechanism analysis. We further collect various static restaurant characteristics, including product tags, number of available payment methods, access to free parking, number of seats static, food category, city, and price range, to augment our identification strategy, which we explain in §2.4.

Natural Language Processing (NLP) for Review Text

Because the review text may contain additional information that is missing from aggregated variables such as review volume and valence [42], we apply the following NLP techniques to better extract the qualitative characteristics from the text content of editorial reviews and user reviews. Recall that all analyses in this subsection are performed at the review level.

Topics First, we utilize topic modeling to discover hidden semantic structures in the textual content of the reviews in our dataset. For each review, we find the abstract topics through the following steps. First, we apply the data pre-processing techniques of tokenizing to delete meaningless stop words (e.g., "a," "and," "the") and stemming to convert our raw review data into a clean dataset. Second, making use of term frequency–inverse document frequency (TF-IDF), we vectorize each review based on the extracted keywords in the context of restaurant reviews. Third, we use latent Dirichlet allocation (LDA) [43] as our topic model to identify the latent topics mentioned in our review data and to obtain the proportion of each latent topic in each review. Note that, in the topic modeling process, we pool editorial reviews and user reviewers together for the following reasons. First, by pooling both types of reviews together, LDA results provide us with a unified Euclidean space that makes any editorial reviews and user reviews comparable. This structure is particularly important for our mechanism analysis, for which we measure the similarity between an editorial review and a user review. Second, pooling provides additional variations (i.e., the variations between editorial reviews and user reviews) for LDA to uncover more topics with which user reviews are distinct from editorial reviews in addition to the topics that differentiate one review from the others within user reviews or editorial reviews.

To determine the number of topics, we develop a performance metric based on the coherence of the words in each topic and the separation between each topic, following [44], [45], and [46]. A smaller value of the metric suggests better performance of the model. We then vary the number of LDA topics from two to ten and calculate the metrics for each value. The performance metrics identify five topics as the optimal choice, as reported in Table 2.1. Using this model, we list the major contributing words in each topic and manually label the latent topics, as shown in Table 2.2. The five topics are: (1) *Food Course*, (2) *Dessert & Drink*, (3) *Emoji*, (4) *Service*, and (5) *Atmosphere*. In addition, topic modeling allows us to represent the topics of each review by using a topic vector $T_{\mathrm{i}kt} = [t_1, t_2, ..., t_5]_{\mathrm{i}kt}$ where $\sum_{\mathrm{j}=1}^{5} t_{\mathrm{j}} = 1$. Specifically, t_{j} (where j = 1, 2, ... , 5) in this vector represents the percentage of topic j in the k^{th} review of restaurant i in month t. For example, if a topic vector of a certain review is {0.7, 0, 0.2, 0.1, 0}, this indicates that the topic in regard to *Food Course* dominates in this review, while this review also includes some *Emoji* and a few descriptions of *Service*.

Table 2.1. Determination of the Number of Topics

Topics	2	3	4	5	6	7	8	9	10
Performance Metrics	0.445	0.434	0.382	0.346	0.374	0.415	0.443	0.548	0.571

Table 2.2. Most Contributing Words for the Five Latent Topics in Reviews

Topic	Most contributing words
1 (Food Course)	Pork, shrimp, fish, chicken, rice...
2 (Dessert & Drink)	Cream, sweet, ice, tea, milk...
3 (Emoji)	U+1F60x, U+1F92x, U+1F9Ex...
4 (Service)	Wait, service, staff, time, people...
5 (Atmosphere)	Atmosphere, nice, air, friend, environment...

Sentiment Sentiment analysis is an NLP technique that extracts review sentiment in a quantitative manner. The primary objective of our sentiment analysis is to determine the perception of a reviewer with respect to the restaurant while writing reviews. Specifically, we conduct the sentiment analysis to classify the polarity of each review text by quantifying whether the expressed opinion in the review text is positive, negative, or neutral [47]. In terms of sentiment analysis algorithms, we implement a rule-based approach that is based on the lexicon VADER, which is a sentiment analysis package developed primarily for social media text [48]. This analysis enables us to label each review with three sentiment scores (i.e., positive score S^{pos}, negative score S^{neg}, and neutral score S^{neu}), where the summation of three scores equals 1 (i.e., $S^{pos} + S^{neg} + S^{neu} = 1$), and each score is non-negative. S^{neu} rates the neutral assertions that exist in a review. Similarly, S^{pos} and S^{neg} represent the proportion of positive and negative assertions, respectively. When a particular score is greater than the other two, this indicates that the tone of this review is driven toward a certain polarity. For instance, if the sentiment scores $\{S^{pos}, S^{neg}, S^{neu}\}$ of a certain review are $\{0.8, 0, 0.2\}$, respectively, we can infer that this review is positive.

Readability The concept of readability analysis refers to the ease with which a reader can understand the written textual content. The readability score is calculated by the complexity of vocabulary and syntax. In the field of NLP, various readability specifications are available for an overall readability score evaluation. We use the SMOG formula to determine the years of education needed to understand the review text, following prior literature [e.g., 41]. By using the number of polysyllabic words, we calculate the SMOG score for each review. A higher SMOG score indicates that the review is more difficult to understand. For example, if the SMOG score of a certain review equals 6, it indicates that six years of education are required to fully understand this review.

2.3.3 Model-free Evidence

In this subsection, we visualize the changes in user reviews after the restaurants receive editorial reviews as model-free evidence, as shown in Figure 2.2.

We find that, as compared to user reviews before editorial reviews, those written after editorial reviews have left-shifted ratings and positive sentiment scores as well as right-shifted lengths, neutral sentiment scores, and topic density on *Food Course* and *Dessert & Drink*,

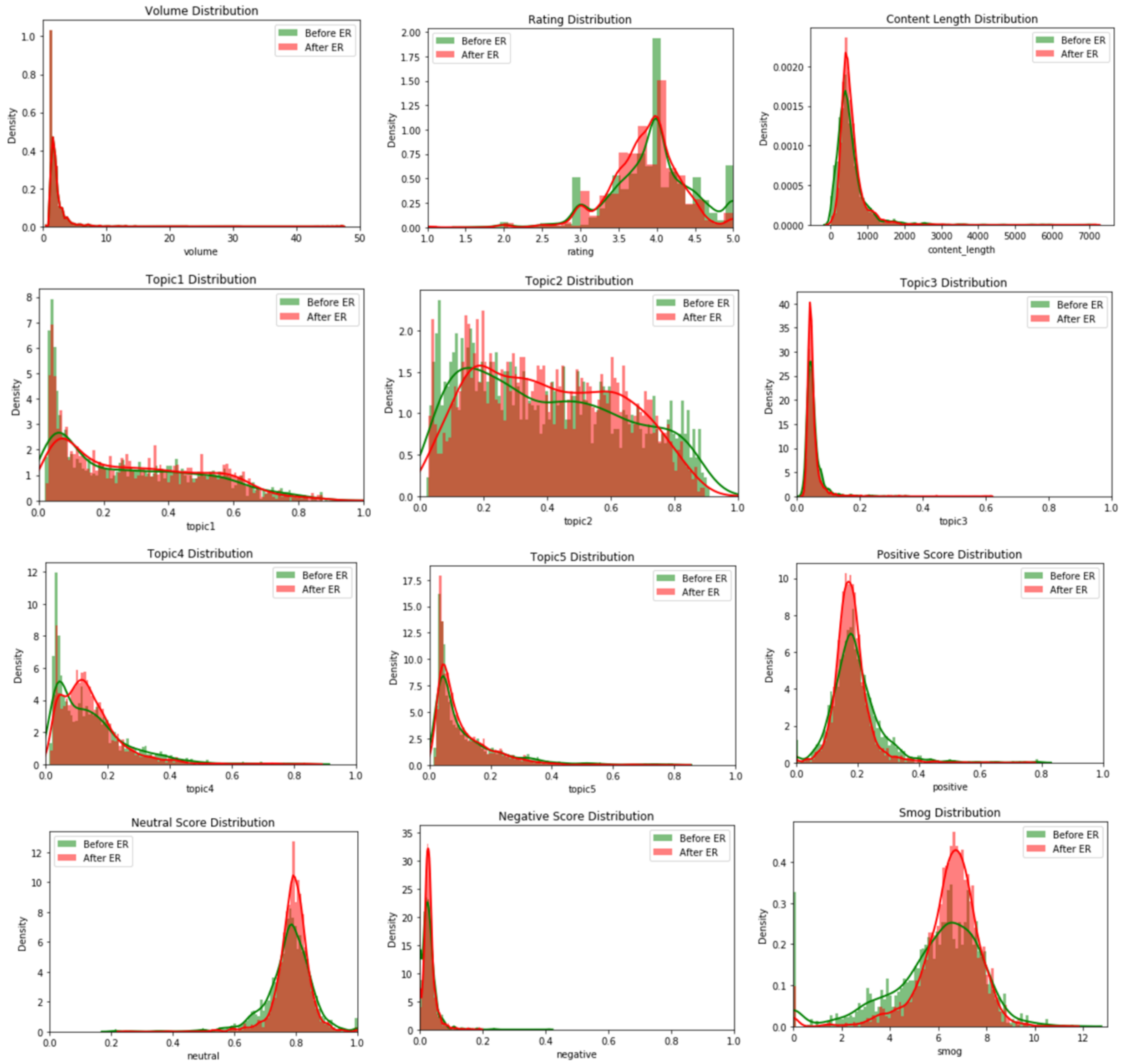


Figure 2.2. Review Characteristics Before and After Editorial Reviews

whereas there are unclear changes for other characteristics. This suggests that reviews after the editorial reviews tend to have lower ratings, less positive sentiments, more neutral sentiments, increased length, and more discussions on food course, desserts, and drinks. Moreover, almost all review characteristics (except review volume) exhibit thinner distributions after editorial reviews, suggesting a homogenizing tendency.[2] These findings, however,

[2]↑In Appendix Figure 2.16, we also show that the user review differences before and after signing advertising contracts, visualized with the same approach, are not as salient, indicating that the advertising contracts may not confound dramatically the impact of editorial reviews on user reviews.

reveal only the before-and-after differences in the user reviews and may be confounded by several factors. We next present our econometric analyses on the impact of editorial reviews.

2.4 Identification Strategy

In this section, we describe our identification strategy, which depends primarily on the combination of propensity score matching (PSM) and difference-in-differences (DID) regression specification.

2.4.1 Main Specification

When identifying the potential effects of editorial reviews, it is important to note that restaurants that receive editorial reviews may be significantly different from restaurants that do not receive them. Therefore, we utilize the matching technique to create a group of restaurants that receive editorial reviews ("treatment group") and a group of restaurants that are similar to the ones in the treatment group but do not receive editorial reviews ("control group"). Using matching, we structure our data to mimic a controlled experiment, which allows us to tease out the causal effects of editorial reviews. Following that, we use the DID technique as the main regression specification. This method is commonly used to estimate the effect of an exogenous treatment by comparing the treatment and control groups' outcomes over time [e.g., 2], [49]. DID is suitable in this research context because we observe a fundamentally quasi-experiment design, whereby the treatments (i.e., editorial reviews) are applied to different businesses at different time points that have user reviews both before and after, whereas many other restaurants do not receive this treatment. In particular, for treatment restaurants, user reviews will be affected by editorial reviews if they are posted after editorial reviews. For control restaurants, there are no such effects.

Matching

As previously noted, the platform selects restaurants to receive editorial reviews based on multiple business characteristics. Thus, restaurants that receive editorial reviews may be different from those that do not receive editorial reviews. Such a selection bias could confound our results. Therefore, to reduce potential differences across the treatment and control restaurants, we rely on four matching specifications to derive a sample of control restaurants that is similar to the treatment restaurants in terms of observable business characteristics. Essentially, each restaurant that receives editorial reviews is paired with a control restaurant that is not "treated" but is similar to the treatment group counterpart in terms of its probability of being treated, allowing a fair comparison of these two groups of restaurants.

1. First, for the static and categorical characteristics, i.e., food category, city, and price range, we use exact matching to pair restaurants across the treatment group, and the control group restaurants in a given food category, city, and price range will be matched only with control group restaurants in the same food category, city, and price range.

2. Second, we drop all (145) restaurants with prior editorial reviews, i.e., restaurants with more than one editorial review by the end of our observation, as no control restaurants have prior editorial reviews. To ensure that there is no systematic difference of heterogeneity of the "prior editorial review environment" between the treatment and control groups, an exact matching must drop all treatment group restaurants with prior editorial reviews.

3. Third, we further exclude all restaurants with advertising contracts for the main analysis and use them for separate sub-sample and pooled sensitivity checks (see §2.6.4). This is equivalent to adding one more exact matching criterion where only samples without advertising contracts exist. Such a restriction helps to ensure that there is no heterogeneity of advertising contracts and, thus, no systematic difference of that heterogeneity between the treatment and control groups to prevent confounding bias introduced by the correlation between advertising contracts and editorial reviews.

4. Fourth, in addition to the first three instances exact matching, we apply PSM on the numerical characteristics of the matched samples from the prior three steps to enhance the validity of the matching process. Recall, however, that our dataset has a unique aspect, whereby each treatment restaurant has a different treatment time (i.e., restau-

rants receive editorial reviews at different times), while none of the control restaurants has the treatment, which is not compatible with traditional PSM approaches. As such, we adopt a technique called two-stage PSM, which is used by earlier studies that faced the same challenge [39], [50]. This technique also has an advantage over the conventional static PSM, as it allows dynamic matching in regard to time-varying characteristics, such as rating and review volume. In two-stage PSM, the first stage approximates restaurant i's overall propensity score as a logistic function of a vector of all static numerical characteristics (i.e., product tags, number of available payment methods, access to free parking, and number of seats) and utilizes one-to-one nearest neighbor matching without replacements to derive the closest matched controlled-and-treatment restaurants pairs. This results in 1,180 pairs of restaurants. In each pair, the treatment time of the treated restaurant is taken as a "hypothetical" treatment time for the untreated counterpart. The second stage extracts the value of the time-varying characteristics, including number of reviews and rating at the "hypothetical" treatment time, and calculates a logistic function to approximate restaurant i's overall propensity score. Again, the process utilizes one-to-one nearest neighbor matching to derive matched pairs, which results in 882 pairs of restaurants. After this matching, the difference in regard to prior review environment characteristics between the treatment group restaurants and the control ones is minimized.

To evaluate the success of the PSM analysis, we compare the overall distributions of propensity scores of the matched and unmatched samples [51]. Figure 2.3 shows the propensity score distribution addition for the treatment and control groups. It is clear that the discrepancy of the propensity score between the treatment and control units before the matching is largely erased and that the distributions of the propensity score after the matching are almost identical across the treatment units and the control ones.[3] We also conduct a balance test to compare all the numerical covariates used in the matching process, as shown in Table 2.3. The results suggest that none of the variables is significantly different between

[3]↑Few treatment restaurants are dropped in the PSM stage. Therefore, there is a minimal distribution addition for the treatment group in Figure 2.3.

the treatment and control groups, indicating that the control group is comparable to the treatment group before the intervention.[4]

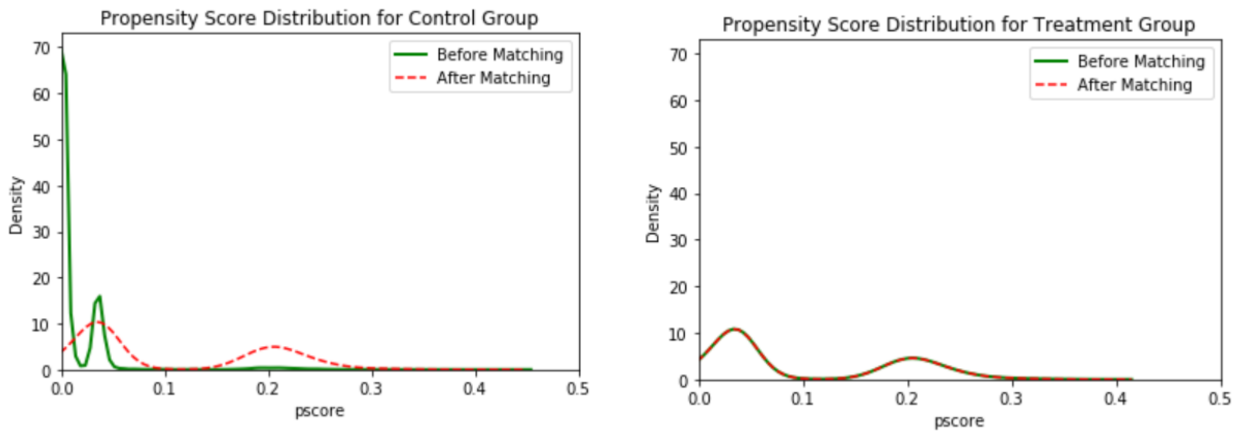


Figure 2.3. Propensity Scores Distribution Addition of Matching for the Treatment and Control Groups

Table 2.3. Balance Tests on Numerical Covariates (Pre-treatment) after Matching

	Treatment				Control				Difference
Variables	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max	p-value
Prior Vol	14.490	24.503	0	310	19.448	24.584	0	247	0.425
Prior Rate	3.929	0.559	1	5	3.893	0.502	1	5	0.195
# of Product Tags	3.561	3.815	0	31	3.527	3.996	0	35	0.790
# of Payment Methods	2.416	0.493	1	3	2.416	0.493	1	3	0.972
Free Parking (dummy)	0.219	0.414	0	1	0.222	0.415	0	1	0.869
# of Seats	22.648	14.889	2	400	22.773	17.866	2	400	0.822

In summary, our final dataset consists of 1,764 restaurants with 86,185 reviews. We provide the complete summary statistics of all relevant variables in Table 2.4. Note that the statistics of all variables are calculated across pre- and post-treatment and across treatment and control restaurants. We also report the summary statistics of all variables for treatment restaurants and control restaurants separately in Tables 2.12 and 2.13 in the appendix.

[4]↑For the categorical covariates included, we are unable to aggregate, present, and compare their values in a concise manner because the number of categories is particularly large. Note, however, that we apply exact matching so that the systematic difference between treatment and control groups is zero.

Table 2.4. Summary Statistics

Variables	Mean	Std. Dev.	Min	Max
<u>User Review Characteristics</u>				
Monthly Review Volume†	1.457	1.885	1.000	532.000
Monthly Rating†	3.833	0.920	1.000	5.000
Rating	3.863	0.876	1.000	5.000
Monthly Length†	401.975	511.078	0.000	41,600.000
Length	502.866	605.890	0.000	42,748.000
Food Course	0.227	0.272	0.000	0.998
Dessert & Drink	0.305	0.299	0.000	0.997
Emoji	0.127	0.177	0.000	0.998
Service	0.175	0.226	0.000	0.998
Atmosphere	0.166	0.205	0.000	0.998
Positive Score	0.183	0.124	0.000	1.000
Negative Score	0.031	0.047	0.000	1.000
Neutral Score	0.787	0.123	0.000	1.000
SMOG Score	6.260	2.905	0.000	17.500
Ad	0.000	0.000	0.000	0.000
Prior Vol*	43.818	55.679	1.000	427.000
Prior ER Vol*	0.000	0.000	0.000	0.000
Prior Rate*	3.888	0.422	1.000	5.000
Experience†*	10.295	12.884	0.000	79.000
<u>Editorial Review Characteristics</u>				
Editorial Review (Treatment)	0.484	0.634	0.000	1.000
Length	2,137.417	1,102.662	70.000	6,978.000
Food Course	0.317	0.260	0.014	0.923
Dessert & Drink	0.558	0.271	0.019	0.939
Emoji	0.034	0.022	0.013	0.347
Service	0.037	0.058	0.016	0.752
Atmosphere	0.054	0.078	0.013	0.796
Positive Score	0.134	0.047	0.000	0.575
Negative Score	0.017	0.015	0.000	0.150
Neutral Score	0.849	0.048	0.425	1.000
SMOG Score	7.788	1.092	0.000	13.400

Note: †: a monthly-level variable; *: cumulative variable; otherwise, review level or non-cumulative.

Difference-in-Differences (DID)

After we constructed a comparable treatment group and control group in regard to prior review environment characteristics, advertising contracts, prior editorial review characteristics, and other business characteristics, we utilize the DID regression to uncover the causal effects of editorial reviews. In particular, our regression specification for the first part of our analysis that focuses on aggregated review characteristics is:

$$Y_{\mathrm{it}} = \beta EditorialReview_{\mathrm{it}} + \alpha_{\mathrm{i}} + \gamma_t + \zeta_{\mathrm{it}} + \epsilon_{\mathrm{it}}, \tag{2.1}$$

where Y includes *ReviewVolume*, *Rating*, and *Length* of restaurant i in month t.[5] The main independent variable is $EditorialReview_{\mathrm{it}}$, which shows whether an editorial review is provided for restaurant i by month t.[6] To account for time-invariant unobserved restaurant-specific characteristics that may influence user reviews (e.g., geographic location of the restaurant), we incorporate a restaurant-fixed effects term α_{i}. To account for the time-variant unobservables that affect all businesses, we include a time-fixed effects term γ_t, which captures seasonality in the change of user review characteristics. In addition, we note that review characteristics tend to change over different phases in the review life cycle, and the phases for different restaurants vary at a fixed time. We further include a relative-time fixed effects term ζ_{it} to control for review-phase-specific unobserved heterogeneity, where the relative time refers to the time difference (by month) between the posted time of a focal review and that of the first review for the same restaurant. Finally, ϵ_{it} represents the error term. This specification allows us to estimate the effect of editorial reviews on subsequent user reviews by observing the key coefficient β. To reduce heteroscedasticity concerns, we leverage robust standard errors clustered at the restaurant level [52].

For the second part of the analysis, in which we focus on the textual content of the review, the unit of our analysis is review level. Therefore, our regression specification becomes:

$$Y_{\mathrm{i}kt} = \beta EditorialReview_{\mathrm{i}kt} + \alpha_{\mathrm{i}} + \gamma_t + \zeta_{\mathrm{it}} + \epsilon_{\mathrm{i}kt}, \tag{2.2}$$

where subscript ikt represents the k^{th} user review for restaurant i in month t. Y here includes the five topics' proportions (i.e., *Food Course*, *Dessert & Drink*, *Emoji*, *Service*,

[5]↑Throughout this paper, because the characteristic of each dependent variable is different, each model is estimated separately.

[6]↑It is possible that editorial reviews for a restaurant are provided in the middle of time T. According to our current specification, the full month T is taken as the treatment, which might result in an underestimated effect of the treatment. As such, we show the robustness of its qualitative interpretation in §2.6.2.

and *Atmosphere*), *Positive/Negative/Neutral Score*, and *SMOG Score*. The independent variable of interest in this specification becomes $EditorialReview_{ikt}$, which indicates whether an editorial review is provided for restaurant i as of the posted time of the k^{th} review in month t. In the same way, restaurant-fixed, time-fixed effects, and relative-time fixed effects are included, and ϵ_{ikt} represents the error term.

2.4.2 Relative Time Model

The validity of our primary identification strategy that leverages the matching and DID model relies critically on the pre-treatment parallel trend assumption (i.e., that there is no significant difference between treatment and control restaurants before the treatment). To test this assumption, we utilize the relative time model with the leads and lags periods [53]. With this model, we add a series of time dummies that indicate the relative chronological distance between observation time and the time one editorial review is written for restaurant i. For the first part of the analysis, in which we focus on the aggregated review characteristics, the specification of our relative time model is as follows:

$$Y_{it} = \sum_{j} \tau_j Pre_{it}(j) + \sum_{l} \omega_l Post_{it}(l) + \alpha_i + \gamma_t + \zeta_{it} + \epsilon_{it}, \tag{2.3}$$

where Y_{it} represents one of review characteristics (i.e., review volume, content length, and rating) for restaurant i in month t. α_i, γ_t, and ζ_{it} denote restaurant-fixed effects and time-fixed effects, and relative-time-fixed effect respectively. The newly added term $Pre_{it}(j)$ is an indicator function that equals 1 if month t is j month(s) prior to the treatment. Here, we consider the posted time of each restaurant's editorial review as the treatment time. For example, if one user review was posted 0-30 day(s) earlier than the editorial review, then the indicator $Pre_{it}(1) = 1$. Similarly, the term $Post_{it}(l)$ is an indicator function that equals 1 if month t is l month(s) after the written of editorial review. For simplicity, we set the indicator $Post_{it}(0) = 1$ if the user review was posted 0-30 day(s) after the editorial review. To estimate all of the effects, we gather all pre-treatment periods that are greater than or equal to six months prior to treatment into one dummy. Similarly, we assemble all post-treatment periods that are greater than or equal to six months following treatment into another dummy. Hence, the coefficient τ_{j} for $j = \geqslant$-6, -5, ... , -1 captures the pre-treatment trend of the impact of editorial reviews on user reviews, while the coefficient ω_l for $l = 0$, 1, 2, ... , 5, $\geqslant$6 captures the effect of editorial reviews in each post-treatment period. Consistent with prior work, we set a period prior to the time of treatment as the baseline by normalizing the coefficient of $Pre_{\mathrm{it}}(-1)$ to zero.

As for the second part of the analysis at the review level, the specification of our relative time model is:

$$Y_{\mathrm{i}kt} = \sum_{\mathrm{j}} \tau_{\mathrm{j}} Pre_{\mathrm{i}kt}(\mathrm{j}) + \sum_{l} \omega_l Post_{\mathrm{i}kt}(l) + \alpha_{\mathrm{i}} + \gamma_t + \zeta_{\mathrm{it}} + \epsilon_{\mathrm{i}kt}. \tag{2.4}$$

As before, $Y_{\mathrm{i}kt}$ represents one of the textual characteristics of the reviews (i.e., topics, sentiment, and readability) for restaurant i at the posted time of the k^{th} review in month t. The term $Pre_{\mathrm{i}kt}(\mathrm{j})$ is an indicator function that equals 1 if the posted time of the k^{th} review in month t is j month(s) prior to the treatment, and, similarly, we consider the posted time of each business editorial review as the treatment time. That is, if one user review was posted 0-30 day(s) earlier than the first editorial review, then the indicator $Pre_{\mathrm{i}kt}(1) = 1$. Consistently, the term $Post_{\mathrm{i}kt}(l)$ is an indicator function that equals 1 if the posted time of the k^{th} review in month t is l month(s) after the written of editorial review. That is, $Post_{\mathrm{i}kt}(0) = 1$ if the user review was posted 0-30 day(s) after the editorial review. α_{i}, γ_t, and ζ_{it} denote restaurant-fixed effect, time-fixed effect, and relative-time-fixed effect, respectively.

2.4.3 Mechanism Analysis

In this subsection, we present additional analyses to understand further the underlying mechanisms for the change of the reviewers' review-writing behaviors caused by editorial reviews. The purpose of our mechanism analysis is to supplement our main analyses, which focus on whether and to what degree editorial reviews causally influence subsequent user reviews with respect to quantity- and quality-related characteristics. In this regard, we seek to examine how and why the subsequent reviews are changed. Recall that one of the theoretical constructs that is usually used in the literature in the context of online reviews is herding and/or differentiation. In particular, prior works have consistently demonstrated that the review generation, rating, and writing behaviors of subsequent reviewers are all influenced by their earlier peers [5]–[8]. Because editorial reviews are presented in a similar manner to those generated by consumers in the context of this study, we expect that a similar influence may exist between editors and subsequent non-editor users. Specifically, on the one hand, the editorial reviews might trigger more users to write reviews (i.e., herding in regard to choosing which restaurant and whether to review) or to follow the editorial reviews in certain dimensions (e.g., topics, sentiment, length, readability) of composing reviews (i.e., herding in regard to how to write reviews). On the other hand, users may be discouraged to review for the restaurant reviewed by editorial reviews (i.e., differentiation regarding choosing which restaurant and whether to review) or review in a different way from the editorial reviews (i.e., differentiation regarding how to write reviews). At the review level, these mechanisms, if existing, should collectively result in differences in the similarity between editorial reviews and user reviews before and after editorial reviews.

Note that our main analyses already reveal the underlying mechanisms, partially through the observed changes in review volume. Based on the DID specification, an increase/decrease in review volume demonstrates that the same set of platform users who more/less review for the restaurants receive editorial reviews, which suggests the herding/differentiation of users toward editorial reviews in regard to choosing which restaurants to review and whether to review.[7] The mechanism that determines how users write reviews, however, remains unclear. To further uncover this perspective, we conduct additional empirical analysis at two levels.

1. First, we investigate the mechanism at the review level to attribute the effect of editorial reviews found in the main analysis to whether the subsequent reviews become more similar or dissimilar to editorial reviews. Despite not disentangling herding/differentiation in regard to how to write review from other mechanisms/factors, e.g., endogenous group forming [54], the answer is valuable from a normative perspective,

[7]↑Note that the analysis for review volume is at an aggregate (i.e., not individual) level and that the reviews for the treatment and control restaurants at any point in time come from the same set of platform users who have the same aggregate users' homophily (i.e., inherent taste). Hence, the effect of *EditorialReview* on volume is not specifically driven by homophily.

as it informs practitioners what the subsequent user reviews are like after posting an editorial review (e.g., more similar or dissimilar to the editorial review) and how to write editorial reviews for desired subsequent user reviews.

2. Second, we further consider the reviewer-level behavioral change triggered by the program and isolate herding/differentiation in regard to how to write a review from the other contextual factors and the endogenous group formation. For instance, this analysis examines whether a group of systematically different reviewers, who inherently write reviews that are similar or dissimilar to the editorial reviews, is more inclined to provide reviews for the restaurants with editorial reviews. If a user's reviews become more similar to editorial reviews after the user's exposure to the editorial review, we attribute that to herding behavior. In contrast, if a user's reviews become more polarized after exposure to editorial reviews, then such a change can be attributed to differentiation behavior. As compared to the review-level analysis, this analysis is more important from a theoretical perspective, as it attribute the impact of editorial reviews to individuals' herding/differentiation behaviors.

For both levels of analysis, we need to first measure the similarity between editorial reviews and user reviews. Specifically, we first quantify the characteristics of editorial reviews in a manner consistent with those of user reviews, using the methods noted in §2.3.2. Using the extracted the topic distributions, sentiment scores, and readability score for editorial reviews, we then construct (1) topic similarity, (2) sentiment/rating similarity, (3) length similarity, and (4) readability similarity to measure the content similarity in all review-level characteristics, as described below.

Topic Similarity Because each review is represented as a topic vector, we measure the topic similarity between editorial reviews and user reviews using cosine similarity. Cosine similarity assesses the cosine of the angle between two non-zero vectors. The cosine similarity between two topic vectors reflects how similar these reviews are in terms of covered topics. Specifically, if the cosine similarity equals 1, it means that these two reviews are identical in regard to the topic distribution. In contrast, if the cosine similarity equals 0, it means that these two reviews are completely different from each other with respect to topics. In our analysis, for each user review, we measure the topic similarity between that user review and the editorial review. For notation, we denote the topic similarity between the k^{th} user review for business i in month t and the editorial review for business i as $\Delta Topic_{\mathrm{i}kt}$. $T_{\mathrm{i}kt}^{UR}$ and T_{i}^{ER} is the topic vector of the user review and that of the editorial review, respectively. Thus, the cosine similarity between topic vectors of editorial reviews and user reviews is as follows:

$$\Delta Topic_{\mathrm{i}kt} = Cos(T_{\mathrm{i}kt}^{UR}, T_{\mathrm{i}}^{ER}), \tag{2.5}$$

Sentiment/Rating Similarity We evaluate the sentiment similarity between editorial reviews and user reviews in a manner that resembles how we measure topic similarity. In particular, we compare the user review with editorial reviews. Instead of using cosine similarity, we focus on the absolute difference of each sentiment score here. Hence, we specify the sentiment similarity as follows:

$$\Delta Sentiment_{\mathrm{i}kt} = |S_{\mathrm{i}kt}^{UR} - S_{\mathrm{i}}^{ER}|, \tag{2.6}$$

where $\Delta Sentiment_{\mathrm{i}kt}$ represents the difference of the sentiment scores between the k^{th} user review in month t for business i and the editorial review for business i. Note that we calculate positive score S^{pos}, neutral score S^{neu}, and negative score S^{neg} to represent the sentiment of each review body. $S_{\mathrm{i}kt}^{UR}$ is a vector of three values, i.e. $[S_{\mathrm{i}kt}^{UR,pos}$, $S_{\mathrm{i}kt}^{UR,neu}$, $S_{\mathrm{i}kt}^{UR,neg}]$, representing the sentiment scores of the k^{th} user review for business i at time t. Consistently, $\Delta Sentiment_{\mathrm{i}kt}$ and S_{i}^{ER} are vectors of three elements.

Note that the editorial reviews do not have an explicit rating. Research has established that rating reflects sentiment and the consistency between review (textual) sentiment and rating [55], and our analysis on the sentiment shows that the "treatment" user reviews tend to be more neutral and less positive due to herding to the editorial reviews, which aggregatively suggests lower sentiment and can explain the declines of the overall rating. We thus construct $\Delta Rating_{\mathrm{i}kt}$ as follows:

$$\Delta Rating_{\mathrm{i}kt} = |Rating_{\mathrm{i}kt}^{UR} - RatingEst_{\mathrm{i}}^{ER}|, \tag{2.7}$$

where $\Delta Rating_{\text{i}kt}$ is the similarity (measured as the absolute difference) between the rating of the k^{th} user review posted for business i in month t, $Rating^{UR}_{\text{i}kt}$ and an estimated rating of the editorial review for business i, $RatingEst^{ER}_{\text{i}}$, inferred from its sentiment in the review text. To construct the latter, we regress all of the instances of $Rating^{UR}_{\text{i}kt}$ on their sentiment variables $S^{UR,pos}_{\text{i}kt}$ and $S^{UR,neu}_{\text{i}kt}$ to extract the correlation between ratings and sentiments.[8] Then, we feed the trained model with the $S^{ER,pos}_{\text{i}}$ and $S^{ER,neu}_{\text{i}}$ to obtain $RatingEst^{ER}_{\text{i}}$. A smaller $\Delta Rating_{\text{i}kt}$, hence, means that the user reviews are more similar in rating and sentiment to the editorial review for the same restaurant.

Length Similarity Similar to the sentiment similarity, the length similarity between editorial reviews and user reviews is represented by:

$$\Delta Length_{\text{i}kt} = |ContentLength^{UR}_{\text{i}kt} - ContentLength^{ER}_{\text{i}}|, \tag{2.8}$$

where $\Delta Length_{\text{i}kt}$ is the content length similarity, measured as the absolute value of the difference between the content length of k^{th} user review posted for business i in month t, $ContentLength^{UR}_{\text{i}kt}$, and that of the editorial review, $ContentLength^{ER}_{\text{i}}$. A smaller $\Delta Length_{\text{i}kt}$ means that the user review is more similar to the editorial review regarding the content length.

Readability Similarity Similar to the specification above, the readability similarity between editorial reviews and user reviews can be represented by:

$$\Delta SMOG_{\text{i}kt} = |SMOG^{UR}_{\text{i}kt} - SMOG^{ER}_{\text{i}}|, \tag{2.9}$$

where $\Delta SMOG_{\text{i}kt}$ is the readability similarity between the k^{th} user review posted for business i in month t and the editorial review for business i; $SMOG^{UR}_{\text{i}kt}$ is the SMOG score of the k^{th} user review posted for business i in month t; and $SMOG^{ER}_{\text{i}}$ is the SMOG score of the editorial review for business i.

[8] ↑ $S^{UR,neg}_{\text{i}kt}$ must be dropped to avoid collinearity.

To reveal the review-level mechanism, we estimate the following equation:

$$Similarity_{ikt} = \beta EditorialReview_{ikt} + \theta_1 Prior\,Vol_{ikt} + \theta_2 Prior\,Rate_{ikt} + \alpha_i + \gamma_t + \zeta_{it} + \epsilon_{ikt}, \tag{2.10}$$

where subscript ikt represents the k^{th} user review for restaurant i in month t. $Similarity_{ikt}$ refers to (an element of) one of the following: $\Delta Topic$, $\Delta Sentiment$, $\Delta Rating$ $\Delta Length$ and $\Delta SMOG$. As before, $EditorialReview_{ikt}$ concerns whether an editorial review for restaurant i is provided as of the posted time of the k^{th} review in month t. $PriorVol_{ikt}$ and $PriorRate_{ikt}$ are the prior review volume and cumulative rating, respectively, for restaurant i by the posted time of the k^{th} review in month t, which we add to control for user review environment changes.[9] α_i, γ_t and ζ_{it} denote restaurant-fixed effect, time-fixed effect, and relative-time fixed effect, respectively. Using the similarity measure as the dependent variable, we show whether the content of user reviews that are posted after the editorial reviews is more similar to that of the editorial reviews than their before-editorial-review counterparts. Note that, for this part of the study, we utilize only a subset of data (41,491 observations), which includes only treatment restaurants because we are interested in the reviews pertaining to only such restaurants. The reviews of the non-treatment restaurants are not affected by any editorial review at any time period and, thus, do not have a paired editorial review to measure their degree of similarity. In addition, before feeding the data into the model, we follow [56] to use a bootstrapping method to achieve volume equalization. Because there are 28,497 after-treatment user reviews, we bootstrapped 28,497 before-treatment reviews. Such equalization rules out the alternative explanation of review volume difference for the changes in content similarity.[10]

The reviewer-level analysis follows the same process as the review-level analysis explained above. We narrow down the analysis scope to long-time users by keeping only those reviews by reviewers who post reviews both before and after the presence of editorial reviews. This

[9]↑Note that adding these two variables results in dropping each restaurant's first user review, which does not have $PriorVol_{ikt}$ and $PriorRate_{ikt}$. To show robustness, we also estimate with the full observation but without controlling for $PriorVol_{ikt}$ and $PriorRate_{ikt}$, as seen Appendix 2.8.4. The estimated coefficients for $EditorialReview$ remain similar.

[10]↑In Appendix 2.8.4, we show that an alternative specification without a bootstrap process, whereby an additional control variable, $ReviewVolume$, is added, leads to qualitatively similar results.

is due to the inherent difficulty in teasing out the effect of endogenous group forming (i.e., homophily) in cross-sectional data of individuals [54], [57]. The herding/differentiation behavior of the short-time reviewers is, thus, unidentifiable. After filtering out the recent users, we have repeated observations for each reviewer r. Using these data,[11] following the identification strategy of [58], we can tease out endogenous group forming through the regression as follows:

$$\begin{aligned} Similarity_{\mathrm{i}rkt} = \beta EditorialReview_{\mathrm{i}rkt} + \theta_1 Prior\,Vol_{\mathrm{i}kt} + \theta_2 Prior\,Rate_{\mathrm{i}kt} \\ + \theta_3 Experience_{rt} + \alpha_{\mathrm{i}} + \tau_r + \gamma_t + \zeta_{\mathrm{i}t} + \epsilon_{\mathrm{i}rkt}, \end{aligned} \tag{2.11}$$

where subscript $irkt$ represents the k^{th} user review for restaurant i in month t generated by reviewer r. Hence, $EditorialReview_{\mathrm{i}rkt}$ concerns whether an editorial review for restaurant i is provided as of the posted time of the k^{th} review in month t generated by reviewer r, and $\epsilon_{\mathrm{i}rkt}$ is the unexplained error term. τ_r is a reviewer-fixed effect that controls for users' inherent and persistent review-writing similarity to editorial reviews. To allow users' inherent review-writing similarity to evolve with their experience, we also control for the experience of reviewer r at time t, $Experience_{rt}$. All other specifications are the same as in Equation (2.10). Following this identification strategy, we use only a within-prior-user-review-environment (rating and volume), within-reviewer, within-experience, within-time, within-restaurant, and across-editorial-review variation in content similarity for the identification of reviewers' behavioral change.

It is worth noting that our mechanism analyses rely on an assumption that contextual effects and inherently changing taste/behaviors of individuals are reasonably controlled for. In other words, it is important to control for other exogenous factors and inherent changes in review-writing similarity to editorial reviews that are not due to the editorial review [5], [54], [57], [58]. Specifically, the assumptions are that $Experience_{rt}$ reasonably controls for such inherently changing taste/behaviors, and that the other controlling specification (i.e., $PriorVol_{\mathrm{i}kt}$, $PriorRate_{\mathrm{i}kt}$, α_{i}, γ_t, and $\zeta_{\mathrm{i}t}$) reasonably controls for contextual effects. As such,

[11]↑These data also are bootstrapped as we work on the review-level analysis.

the estimate of β is theoretically a close and upper bound for the behavioral change of review writing, influenced by editorial reviews [58].

2.5 Empirical Results

2.5.1 Relative Time Model

We first report the results of the relative time model to establish that the parallel trend assumption holds in our analysis so our Matching + DID specification is valid. As shown in Table 2.5, we find none of the coefficient of the pre-treatment dummies, Pre(j), to be statistically significant at p-value < 0.10. This result confirms that there is no detectable pre-treatment dissimilarity across reviews for restaurants that receive editorial reviews or for those that do not receive editorial reviews after the matching process; hence, the parallel trend assumption is satisfied.

Table 2.5. Results from the Relative Time Model

DV:	log(Review Volume)	Rating	log(Length)	Food Course	Dessert &Drink	Emoji	Service	Atmosph-ere	Positive	Neutral	Negative	SMOG
Pre (⩾-6)	-0.020	-0.007	0.036	0.002	-0.010	0.008	0.002	0.006	0.002	0.002	0.000	-0.072
	(0.013)	(0.037)	(0.055)	(0.007)	(0.006)	(0.006)	(0.008)	(0.007)	(0.004)	(0.004)	(0.002)	(0.101)
Pre(-5)	-0.021	-0.012	0.033	0.004	-0.004	0.005	0.005	0.008	0.002	0.002	0.001	-0.210
	(0.17)	(0.036)	(0.041)	(0.011)	(0.012)	(0.008)	(0.011)	(0.010)	(0.006)	(0.006)	(0.002)	(0.145)
Pre(-4)	0.019	-0.022	0.011	0.009	-0.010	0.003	0.003	0.001	0.000	0.005	-0.000	-0.213
	(0.015)	(0.033)	(0.038)	(0.011)	(0.012)	(0.008)	(0.011)	(0.010)	(0.006)	(0.006)	(0.002)	(0.144)
Pre(-3)	0.004	-0.023	0.023	0.016	-0.015	0.000	0.015	0.007	-0.004	0.006	0.001	-0.078
	(0.014)	(0.030)	(0.035)	(0.010)	(0.012)	(0.008)	(0.010)	(0.009)	(0.006)	(0.006)	(0.002)	(0.139)
Pre(-2)	0.010	-0.035	0.022	0.014	0.002	0.002	-0.005	0.007	-0.008	0.007	0.003	-0.071
	(0.013)	(0.027)	(0.032)	(0.009)	(0.011)	(0.007)	(0.010)	(0.009)	(0.006)	(0.006)	(0.002)	(0.129)
Pre(-1)						Baseline						
Post(0)	0.056***	-0.019	0.070**	0.005	0.005	0.002	-0.007	-0.001	-0.011**	0.011**	0.003	-0.042
	(0.011)	(0.023)	(0.028)	(0.009)	(0.010)	(0.007)	(0.009)	(0.008)	(0.005)	(0.005)	(0.002)	(0.118)
Post(1)	0.042***	-0.021	0.049*	0.009	0.010	0.000	-0.011*	-0.015*	-0.012**	0.012**	0.003	0.037
	(0.011)	(0.023)	(0.028)	(0.009)	(0.010)	(0.007)	(0.009)	(0.008)	(0.005)	(0.005)	(0.002)	(0.124)
Post(2)	0.032***	-0.023	0.049*	0.013	0.020**	-0.004	-0.016*	-0.020**	-0.013***	0.016***	0.002	0.090
	(0.012)	(0.024)	(0.029)	(0.009)	(0.010)	(0.007)	(0.009)	(0.008)	(0.005)	(0.005)	(0.002)	(0.126)
Post(3)	0.023*	-0.042*	0.048*	0.016*	0.021**	-0.009	-0.022**	-0.021**	-0.012**	0.011**	0.003	0.116
	(0.012)	(0.024)	(0.029)	(0.009)	(0.010)	(0.007)	(0.009)	(0.008)	(0.005)	(0.005)	(0.002)	(0.125)
Post(4)	0.022*	-0.056**	0.050*	0.018**	0.027**	-0.012*	-0.028***	-0.027***	-0.011**	0.009	0.002	0.232*
	(0.012)	(0.025)	(0.030)	(0.009)	(0.010)	(0.007)	(0.010)	(0.009)	(0.006)	(0.006)	(0.002)	(0.128)
Post(5)	0.021*	-0.044*	0.066**	0.020**	0.032***	-0.013*	-0.016*	-0.027***	-0.010	0.010	0.001	0.111
	(0.012)	(0.026)	(0.031)	(0.009)	(0.010)	(0.007)	(0.010)	(0.009)	(0.006)	(0.006)	(0.002)	(0.130)
Post(⩾6)	0.033***	-0.055**	0.031	0.025***	0.033***	-0.017***	-0.010**	-0.016**	-0.006	0.008*	0.000	0.089
	(0.09)	(0.027)	(0.022)	(0.007)	(0.008)	(0.005)	(0.007)	(0.006)	(0.004)	(0.004)	(0.002)	(0.094)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	48,027	48,027	48,027	86,185	86,185	86,185	86,185	86,185	86,185	86,185	86,185	86,185
Adj R-squared	0.293	0.137	0.157	0.436	0.436	0.290	0.166	0.200	0.096	0.091	0.059	0.081

2.5.2 Aggregated Review Characteristics

The estimation results for the impact of editorial reviews on aggregated characteristics of user reviews, which are performed at the restaurant-monthly level, are shown in Table 2.6. First, we find a strong and significant increase in the number of new user reviews (i.e., $ReviewVolume$) after the presence of editorial reviews. From an economic perspective, this translates to an increase of 3.1% user reviews per month for restaurants with one editorial review. The discussion of underlying mechanisms related to review volume (see §2.4.3) implies that platform users tend to write more review for restaurants that receive editorial reviews, suggesting the herding of users toward editors in regard to which restaurants to review. A possible next-layer mechanism, as shown in the literature, is that the posting of editorial reviews can increase consumers' awareness of the restaurant [59]. Hence, once receiving editorial reviews, the restaurants may attract more attention from consumers and receive more reviews. In addition, the content of editorial reviews may trigger users' willingness to express their opinions, especially when they are inspired by agreed-upon but not previously discussed topics or want to share their disagreement with the content of editorial reviews.

Table 2.6. Effect of Editorial Reviews on Aggregated Review Characteristics

DV:	log(Review Volume)	Rating	log(Length)
Editorial Review	0.031***	-0.022*	0.035**
	(0.006)	(0.013)	(0.014)
Restaurant FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes
Observations	48,027	48,027	48,027
Adj R-Squared	0.318	0.137	0.163

Note: robust standard errors in parentheses;* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Similarly, the existence of editorial reviews increases the *Length* of the subsequent user reviews for the same restaurants. All else equal, this finding suggests that user reviews written for restaurants with an editorial review are about 3.6% longer, on average, than are those written for restaurants without editorial reviews. A plausible interpretation is that editorial reviews serve as a starting point for discussions. The comprehensive content of editorial reviews may inspire reviewers to express their opinions about restaurants in a more thorough fashion. For instance, reviewers may provide only a simple evaluation of the food (e.g., delicious or not) before the existence of editorial reviews. After reading editorial reviews, however, reviewers may begin to assess the food from various perspectives, such as flavor, smell, appearance, and so forth. We will explore this possibility by examining the content of user reviews influenced by editorial reviews in §2.5.3 and §2.5.4.

Further, we find a negative effect of editorial reviews on the dependent variable Rating, although the statistical power of such an effect is weak. This indicates that the review valence decreases for restaurants that receive editorial reviews. At first glance, this result appears to be surprising, given that editorial reviews are often perceived to be similar to advertorial ones. Thus, one may expect that editorial reviews may positively influence the ratings of subsequent reviews. It turns out, however, that when the editorial reviews are neutral, this appears to drive subsequent reviewers to be more neutral, as evidenced by the decrease in average star ratings. We examine the change in reviewers' behavior following editorial reviews further in §2.5.3 and §2.5.4.

According to the results above, it appears that editorial reviews stimulate reviewers to not only contribute more reviews to the platform but also spend more effort in doing so, which results in greater length of content. From this perspective, editorial reviews can be considered a valuable addition that could stimulate greater user contribution to the overall review platform.

2.5.3 Characteristics of Textual Content

Next, we examine the changes in the textual content of user reviews that are influenced by the presence of editorial reviews. The analysis in this section helps us to understand more about the change in review-writing that follows editorial reviews. Table 2.7 presents the results from such an analysis, which is performed at the review level.

Recall that *Dessert & Drink* and *Food Course* occupy the two largest portions of the editorial review content (see Table 2.4) and that the proportion of these topics is larger in editorial reviews than in user reviews (i.e., editorial reviews more often discuss food and drink). The positive and statistically significant coefficients of $EditorialReview$ for *Dessert*

Table 2.7. Effect of Editorial Reviews on Textual Review Characteristics

DV:	Food Course	Dessert &Drink	Emoji	Service	Atmosphere	Positive	Neutral	Negative	SMOG
Editorial Review	0.009***	0.032***	-0.007***	-0.011***	-0.023***	-0.002**	0.002**	<0.001	0.028
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)	(<0.001)	(0.039)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	86,185	86,185	86,185	86,185	86,185	86,185	86,185	86,185	86,185
Adj R-Squared	0.436	0.368	0.290	0.166	0.200	0.059	0.054	0.021	0.090

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

& Drink and *Food Course* indicate that the presence of editorial reviews influences user reviews to more often discuss food and drink (i.e., *Dessert & Drink* and *Food Course*) as well. At the same time, the proportion of *Emoji*, *Service*, and *Atmosphere* shrinks. In short, the content structure of user reviews gradually resembles that of editorial reviews. It is worth noting, however, that a decrease in topic proportion does not mean a decrease in the content length for that topic because we normalize the content length to calculate the proportion. In §2.6.1, we consider length and topic proportion jointly to examine the net changes in topic content length and, thus, content variety.

In addition, the presence of editorial reviews greatly affects the sentiment of subsequent user reviews. Specifically, the treatment effect on neutral scores of user reviews is significantly positive at p-value < 0.05. This finding indicates that reviews become more neutral if they are contributed after the editorial reviews are posted for the focal restaurant. At the same time, the positive sentiment of user reviews decreases at p-value < 0.05. These results are consistent with the finding in the previous subsection, which suggests that user reviews become more neutral (i.e., the aggregated review valence decreases) and provide more detail about the change in the reviewers' behavior in regard to the review sentiment. Note that the neutral score of editorial reviews is 0.85, on average, and is higher than that of user reviews in our dataset. Following the neutral perspectives expressed in the editorial reviews, reviewers appear to be more objective when they evaluate restaurants, resulting in a slight decline of overall review valence. Finally, the readability of user reviews does not experience a significant change after editorial reviews are posted.

2.5.4 Underlying Mechanisms

In this section, we report results from our mechanism analysis to further our understanding regarding the underlying mechanisms of the changes in reviewer behavior observed in the previous subsections. The estimation results for the review-level analysis are provided in Table 2.8, which shows that the content of user reviews becomes more similar to that of editorial reviews in all aspects, including content topics, sentiment (rating), length, and readability. In particular, recall that the larger value of cosine similarity measure indicates a higher degree of similarity. The significantly positive coefficient for $\Delta Topic$ indicates that the topic distribution covered by reviews published after editorial reviews are becoming more similar to the editorial review content. Specifically, reviewers tend to focus more on topics such as desserts, drinks, and main courses, which are in line with the proportion of those topics in editorial reviews.

Table 2.8. Estimation Results of Review-level Mechanism Analysis

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.012***	-0.005***	-0.004***	0.001	-0.116***	-0.026***	-0.043***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.033)	(0.008)	(0.011)
Prior Vol	0.000***	-0.000***	-0.000	0.000	-0.000	0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Prior Rate	-0.002	0.001	0.002	-0.001	0.060	0.012	-0.145***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.043)	(0.011)	(0.015)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56,994	56,994	56,994	56,994	56,994	56,994	56,994
Adj R-Squared	0.205	0.165	0.165	0.071	0.189	0.618	0.143

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Further, recall that the similarity measures for sentiment, rating, length, and readability are constructed as an absolute value of difference. As such, a smaller value represents a larger degree of similarity. The negative and statistically significant coefficients of $\Delta Positive$, $\Delta Neutral$ and $\Delta Rating$ indicate that the sentiment (rating) of user reviews gravitates toward the sentiment of editorial reviews written for the same restaurant. Similarly, the coefficient for $\Delta Length$ is also negative and statistically significant, suggesting that the length of subsequent user reviews, as compared to those of prior user reviews, is closer to those of editorial reviews. Interestingly, the coefficient for $\Delta SMOG$ is also significantly negative, indicating that the readability of editorial and user reviews is also becoming more similar. Together with the insignificant impact found in our main results from the DID specification (§2.5.3 Table 2.7), this indicates a potential "spillover effect" of editorial reviews to the readability of reviews of the non-treatment restaurants. In other words, an editorial review might influence not only the readability of the subsequent user reviews of the same restaurant but also those of non-treatment restaurants. Such a spillover effect is also theoretically possible because readability is likely to be reviewer-dependent rather than restaurant-dependent. Therefore, a change in review-writing behavior in regard to review readability may not be exclusively applied to the reviews of the treatment restaurants alone.

In summary, we find strong evidence that restaurant reviews are starting to resemble editorial reviews after the editorial reviews' publication from the perspective of topics, sentiment, and readability. This suggests that review platforms can set "exemplar" editorial reviews to derive desired subsequent user reviews that resemble the example. Interestingly, in a supplemental analysis for a different purpose, we show insignificant resembling tendencies toward top-placed non-editor user reviews in all dimensions (see Appendix 2.8.4). This indicates that the effects of editorial reviews should be more than the salience effects of top-placed reviews, suggesting the existence of alternative mechanisms, e.g., herding, that are specific to editorial reviews.

The review-level analysis results provide support for herding behavior from consumers toward editors. We further validate these results by conducting additional analysis at the reviewer level, using long-term reviewers. The results are shown in Table 2.9. The $EditorialReview$ coefficients of all similarity measures are qualitatively consistent with those at the review-level analysis. Note that, in this analysis, the coefficients are identified from the within-reviewer variations. The coefficients thus are interpreted as that for a given user; his or her reviews after the editorial reviews for the same restaurants are more similar to the editorial reviews than were his or her reviews before, which aligns with the herding behavior of users in their review generation. This suggests that editorial reviews may exhibit gravitational powers that persuade reviewers to follow while contributing their own content. Therefore, by providing high-quality editorial reviews, the review platform can nudge reviewers to provide better quality feedback. Taken together with the finding of increased review volume triggered by editorial reviews, our analysis suggests the herding of users toward platform editors in terms of choosing which restaurants to review and how to write reviews.

Table 2.9. Estimation Results of Reviewer-level Mechanism Analysis

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.015***	-0.004**	-0.006***	-0.001	-0.198***	-0.036***	-0.033***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.045)	(0.014)	(0.013)
Experience	0.008***	-0.002*	-0.001*	0.000	-0.030**	-0.025***	-0.009**
	(0.001)	(0.001)	(0.001)	(0.000)	(0.013)	(0.004)	(0.004)
Prior Vol	-0.000***	-0.000***	-0.000***	-0.000	0.000	0.000	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	-0.012***	0.008***	0.006**	0.003***	-0.293***	0.164***	-0.200***
	(0.004)	(0.003)	(0.003)	(0.001)	(0.064)	(0.019)	(0.018)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51,372	51,372	51,372	51,372	51,372	51,372	51,372
Adj R-Squared	0.925	0.907	0.911	0.895	0.908	0.959	0.945

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

2.6 Robustness Checks

2.6.1 Content Length for Topics

In the main analysis, we consider the impact of editorial reviews on the proportion of each topic in subsequent user reviews. The results may, however, produce misleading interpretations about topic diversity. The decrease in the proportion of *Emoji*, *Service* and *Atmosphere* does not necessarily indicate decreased content length of such topics and, thus, decreased topic diversity after considering that the total content length increases, as shown in §2.5.2. To reveal the impact of editorial reviews on topic diversity, we investigate the impact of editorial reviews on the content length of each topic, i.e., *TopicLength*, of the subsequent user reviews. To calculate *TopicLength*, we multiply the review-level *ContentLength* with the relative topic proportion (i.e., *FoodCourse*, *Dessert & Drink*, *Emoji*, *Service*, and *Atmosphere*). This is in line with LDA topic modeling that extracts the relative proportion of each topic as the topic-length-to-content-length ratio. We then estimate the impact of editorial reviews on *TopicLength*, using the main DID specification in Equation 2.2, with dependent variable $log(TopicLength)$. As shown in Table 2.10, we find that the topic lengths for *Food Course* and *Dessert & Drink* significantly increase, whereas the topic lengths for the other topics do not change significantly. This suggests that, although the editorial reviews increase the content on *Food Course* and *Dessert & Drink*, they do not decrease the content on other topics. This indicates a net increase in content variety contributed by the increased length for *Food Course* and *Dessert & Drink*.

Table 2.10. Effect of Editorial Reviews on Topic Length

DV:	log(Food Course Length)	log(Dessert& Drink Length)	log(Emoji Length)	log(Service Length)	log(Atmosphere Length)
Editorial Review	0.066***	0.146**	-0.032	-0.034	-0.035
	(0.025)	(0.024)	(0.023)	(0.024)	(0.023)
Restaurant FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes
Observations	86,185	86,185	86,185	86,185	86,185
Adj R-Squared	0.323	0.255	0.109	0.204	0.135

Note: robust standard errors in parentheses;* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

2.6.2 Alternative Specifications for Ratings, Length, and Review Volume

For aggregate review characteristics (i.e., *Rating*, *Length*, and $ReviewVolume$), it is possible that editorial reviews for restaurants are provided in the middle of time T. Our current specification thus results in an underestimated treatment effect. Note that we also have review-level observations for *Rating* and *Length*, and the review-level analysis that uses Equation (2.2) does not suffer from this treatment effect. Hence, we conduct an additional review-level analysis for *Rating* and *Length*. We further alternatively specify the treatment in Equation (2.1) as $Lagged\,EditorialReview_{\mathrm{it}}$, in which, if restaurant i is treated in the middle of T, $Lagged\ Editorial\ Review_{\mathrm{i}T} = 0$ whereas $LaggedEditorialReview_{\mathrm{it}} = 1$, as t starts from $T + 1$. The results of both alternative specifications, shown in Table 2.11, are consistent with the main results in Table 2.6. In addition, note that both *Length* and $ReviewVolume$ are count variables, and we consider the alternative model framework of negative binomial regression. The results, also shown in Table 2.11, are consistent with the main results.

Table 2.11. Alternative Specifications for Ratings, Length, and Review Volume

	Restaurant-month Level			Review Level		Negative Binomial	
DV:	log(Review Volume)	Rating	log(Length)	Rating	log(Length)	Review Volume	Length
Editorial Review				-0.020*	0.027**	0.026**	0.023**
				(0.012)	(0.012)	(0.013)	(0.008)
Lagged Editorial Review	0.020***	-0.021*	0.030**				
	(0.006)	(0.012)	(0.014)				
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	48,027	48,027	48,027	86,185	86,185	48,027	48,027
Adj/Pseudo R-Squared	0.318	0.137	0.163	0.113	0.123	0.152	0.119

Note: robust standard errors in parentheses;* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

2.6.3 Controlling for Prior Review Environment in Regression

In our main analysis, we control for prior review environment, e.g., rating and review volume, in the matching process. An alternative approach to control for such factors is to add $PriorVol$ and $PriorRatings$ as control variables in the DID regressions in Equation (2.1) and Equation (2.2).[12] Note that we do not add prior editorial review characteristics because all of the data points can have, at most, one editorial review and no prior ones. The results, shown in Appendix Tables 2.14 and 2.15, are consistent with the main analysis results.

We note that, despite consistency, the above estimated effects are theoretically biased from the causal effects of editorial reviews. Controlling for the prior user reviews blocks any indirect causal effects of editorial reviews on focal user reviews through other user reviews in between. Consequently, the estimated treatment effects are closer to the direct causal effects. To correct for this bias, for all of the user reviews of the controlled restaurants and of the treatment restaurants before their treatment, we keep the same specification as above, whereas for the "treated" user reviews, we set $PriorVol$ and $PriorRatings$ to be the values

[12]↑As a consequence, each restaurant's first review, which has no prior reviews, must be dropped, and we see a decreased number of observations.

at the treatment time. This improved specification prevents blocking of the noted indirect causal effects. The estimation results, shown in Appendix Tables 2.16 and 2.17, are again consistent with our main results.[13]

2.6.4 Additional Analysis for Restaurants with Advertising Contracts

Our main analysis excludes the restaurants with advertising contracts to ensure that our results are not confounded by these contracts. To demonstrate the generalizability of our results, we conduct two additional analyses, using those excluded restaurants. The first analysis uses the restaurants with advertising contracts only, and the second pools such restaurants with the restaurants for the main analysis (i.e., restaurants never with advertising contracts). The general empirical framework is the same as our main analysis, except that we include the indicator of whether restaurant i at time t has an advertising contract Ad_{it} as an additional covariate in the second stage of the two-stage PSM and as a control variable in the mechanism analysis. As shown in Appendix Tables 2.18-2.25, the impact of editorial reviews remains similar to the main analysis and the mechanism analysis at all levels.

2.6.5 Including Heterogeneous Treatment Effects of Sentiment

In this section, we consider the heterogeneity effects of editorial reviews. Note that restaurants in our main analysis all have the same editorial review volume of one. We thus focus on the heterogeneity in the sentiment of these editorial reviews. Specifically, we add demeaned $NegativeScore$ and $NeutralScore$ of the treatment editorial reviews[14] and interact them with the treatment. The results are shown in Appendix Tables 2.26 and 2.27. For mechanism analysis, we also include these two variables to examine their varying impact. The results are shown in Appendix Tables 2.28 and 2.29. The effects of $EditorialReview$ (i.e., the average treatment effect in the main specification) are consistent with our main results. Interestingly, we find that most effects of the editorial review sentiments (i.e., heterogeneous treatment effects of sentiment in the main specification) are insignificant or marginally sig-

[13]↑We also apply this alternative specification to the mechanism analysis in Appendix 2.8.4. The estimation results, shown in Appendix Tables 2.40 and 2.41, remain qualitatively the same.
[14]↑$PositiveScore$ is dropped to avoid collinearity.

nificant except those of *Food Course*. These results suggest that either editorial reviews are highly homogeneous in sentiments or that the heterogeneity does not impose an additional impact on most dimensions of the subsequent user reviews. The former explanation is consistent with the summary statistics shown in Table 2.4: The variation of editorial reviews sentiments is roughly one-third of that of user reviews sentiments. We further leverage an alternative identification scheme that controls for heterogeneity through matching. Specifically, we construct a compound sentiment score, following [48], for editorial reviews and then keep only those restaurants with similar compound sentiment scores (i.e., in the range of average compound score $\pm$ 0.05) as an additional matching stage to erase the sentiment heterogeneity. We then use those matched restaurants for the main specification and mechanism analysis. As shown in Appendix Tables 2.30-2.33, the effects of $EditorialReview$ remain similar to the main results.

2.6.6 Alternative Matching Methods

In this subsection, we demonstrate that our results are not driven solely by the specific matching algorithm (i.e., the nearest neighbor matching) used in PSM. Here, instead, we adopt the entropy balancing approach, which is widely used in the literature [e.g., 60], [61], and update the estimation results using the main DID regression specification. The results, shown in Appendix Tables 2.34 and 2.35, are consistent with our main estimation results, indicating their robustness to matching algorithms. Also note that the results for mechanism analysis (shown in Appendix Tables 2.36 and 2.37) are not sensitive to the matching approach.

2.7 Conclusions

Third-party reviews, such as those written by expert reviewers, have a strong impact on a consumer's purchasing decisions and a company's sales. Such an economic implication also is observed for consumer reviews. Although prior studies have shown how consumers reviews affect subsequent reviews, there is virtually no previous work that examines the influence of third-party reviews on consumer reviews. Our study addresses this gap in the literature by collaborating with a large restaurant review platform in Asia to investigate the impact of editorial reviews, a form of third-party reviews, on subsequent consumer reviews in terms of content quantity and quality.

By leveraging a quasi-experiment setup, we find empirical evidence that editorial reviews significantly influence subsequent user reviews and users' review-writing behavior. Specifically, editorial reviews increase the quantity of subsequent user reviews and reshape their textual content. Restaurants that have editorial reviews receive more consumer reviews that not only are longer but also resemble the editorial review(s) in regard to the topic, sentiment/rating, length, and readability in the subsequent period. The increased length comes from augmented content on the topics that appear more in editorial reviews, while the content for the other topics does not decrease, suggesting a net increase in the content variety. The reviewer-level analysis for long-term reviewers further attributes the changes in their reviews to their behavioral changes under identifying assumptions. This finding, together with that of the increased review volume, suggests users' herding behavior toward platform editors not only in their choice of which restaurants to review but also in how to write reviews.

Our work makes a substantial contribution to academic literature and practice. From the perspective of academic research, we contribute to the extant literature on herding effects in the specific context of UGC by conceptualizing and identifying users' herding behaviors toward platform editors. Such a connection extends the scope of our understanding of peer influence on the UGC platform, as prior literature focuses only on the influence of peer users. It is particularly difficult, however, to extrapolate these prior findings and apply them to the context of editorial reviews because editorial reviews are significantly different from consumer reviews, and the platform control over these two types of feedback is also vastly different. In addition, we contribute to the literature on third-party reviews by revealing their additional economic value through examining their impact on WOM. This helps marketers to further comprehend the value of third-party reviews. Such an effect of third-party reviews is increasingly important due to the ubiquity of digitized UGC and eWOM in online platforms. In addition, our study contributes to the literature on the management of UGC by platforms, as our findings suggest the effectiveness of using an editorial review program to manage UGC in terms of content quantity and quality. We are among the first to demonstrate empirically the influence of editorial reviews on subsequent consumer reviews and reveal the underlying mechanisms of such an influence.

From a practical perspective, our research also provides insights that could help platform managers to enhance their reviewers' contributions and improve the quality of review content, both of which are crucial to the platform's competitive advantage and long-term sustainability. Herding by peer reviewers toward editors indicates that platforms could use editorial reviews to influence and manage the content generation of its users. Because the platform can more exogenously manage the content of editorial reviews and determine for which businesses to provide editorial reviews, as compared with consumer reviews, platform managers could leverage editorial reviews not only to encourage more users to write reviews for targeted businesses but also to influence the topic, sentiment/rating, length, and readability of their written expressions. Specifically, high-quality editorial reviews could lead platform users to provide more content, cover desired topics more, and use a more neutral tone and readable writing style when contributing reviews to the platform. Such high-quality content would further reduce information seekers' cost of searching and improve their overall experience on the platform. Therefore, the managerial implications of our research are especially valuable for relatively mature review platforms that have accumulated enough review volume and have shifted their focus to the quality of the content and the reading experience of users.

Being one of the first works to investigate the interplay between editorial reviews and consumer reviews, our work is subject to several limitations, which also represent avenues for future research. First, the editorial/consumer reviews in our dataset are exclusively for experience goods. It would be interesting to determine whether the effects we observe remain the same for reviews of search goods. Second, identifying herding effects is inherently challenging in observational data [57] because we can isolate only such effects among long-term users under identifying assumptions. Further works that utilize different research methodologies, such as randomized experiments, may challenge such assumptions and reexamine herding effects among all users. Alternatively, follow-up surveys or small-group interviews may further reveal the users' state of mind and uncover why such changes occur. Third, the review platform in our study is located in Asia, and more than 95% of the users are from Asia. A future study could investigate whether herding/differentiating behavior differs in another culture. Finally, future research that has access to panel data for traffic and sales may be able to further attribute the increased volume of user reviews, i.e., herding in which restaurants to review, to the awareness and conversion changes of users and to confirm whether the indirect impact of editorial reviews on consumer reviews generates cascading effects on product sales.

2.8 Appendix

2.8.1 Statistics of the Data

We supply relevant statistics to populate our model due to a non-disclosure agreement with the collaborating platform (including both summary statistics and distributional statistics). Specifically, Figure 2.4 to Figure 2.15 show probability density function (PDF) of all our dependent variables. Additionally, Table 2.12 and Table 2.13 show summary statistics for treated group and control group respectively.

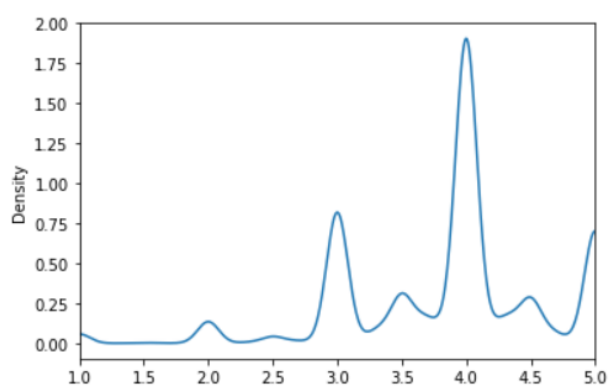


Figure 2.4. Probability Density Function of Aggregated Rating by Month

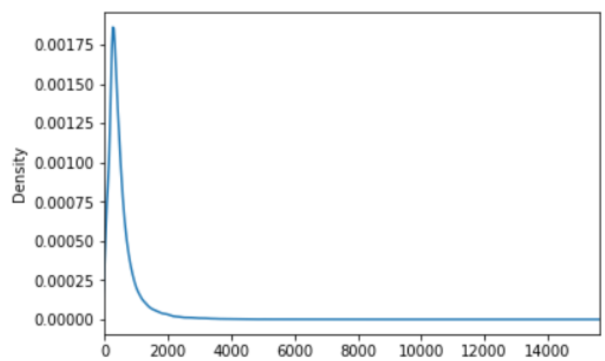


Figure 2.5. Probability Density Function of Aggregated Length by Month

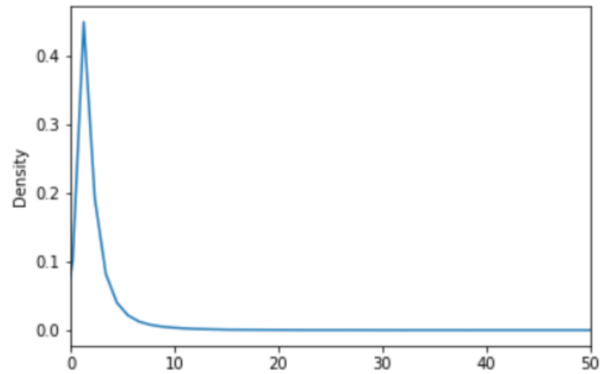


Figure 2.6. Probability Density Function of Volume (Monthly)

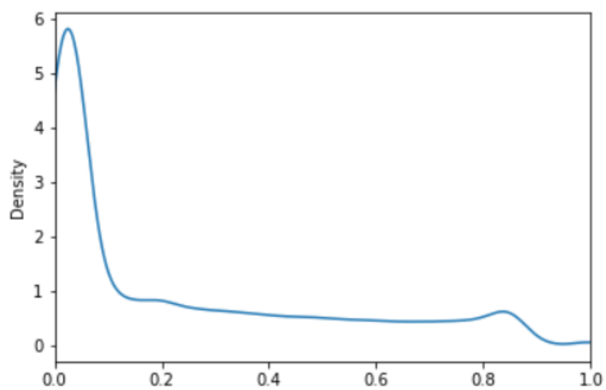


Figure 2.7. Probability Density Function of Topic1

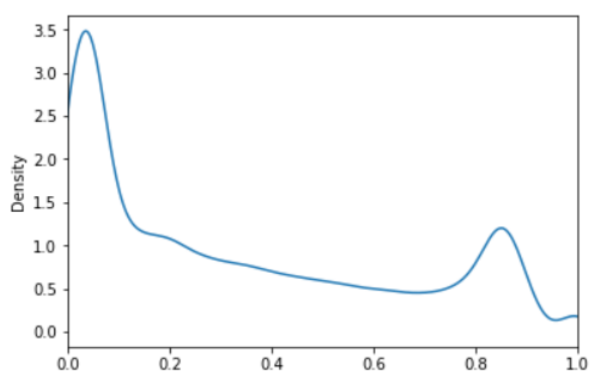


Figure 2.8. Probability Density Function of Topic2

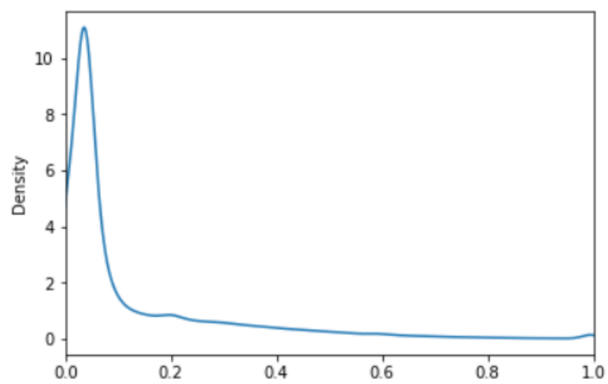


Figure 2.9. Probability Density Function of Topic3

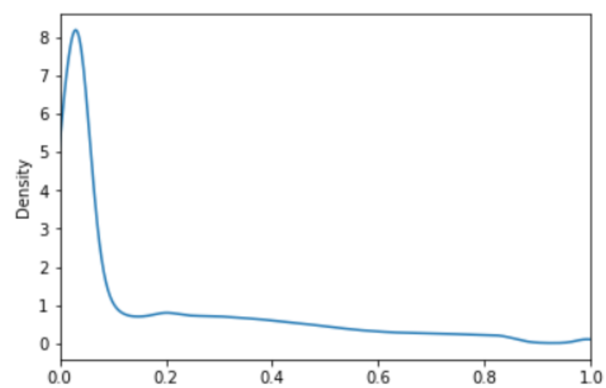


Figure 2.10. Probability Density Function of Topic4

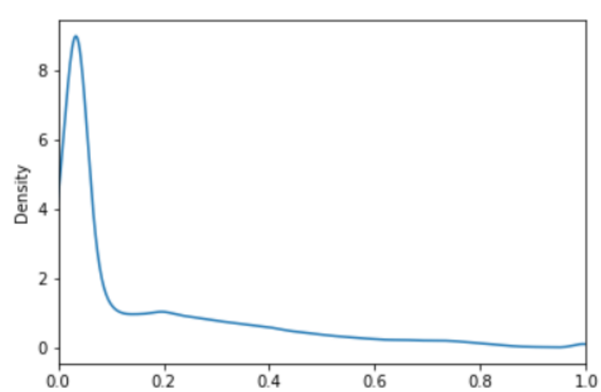


Figure 2.11. Probability Density Function of Topic5

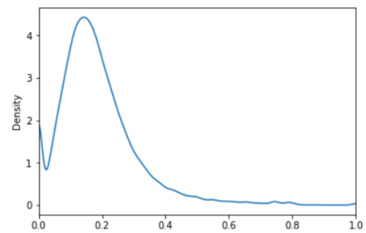


Figure 2.12. Probability Density Function of Positive Scores

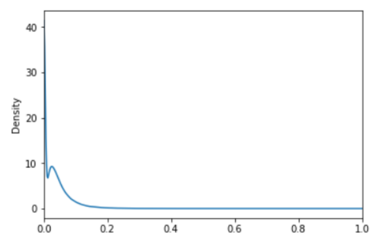


Figure 2.13. Probability Density Function of Negative Scores

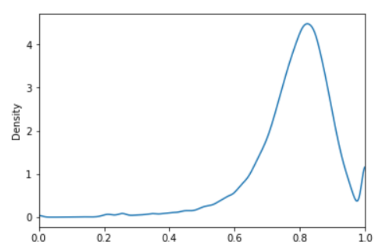


Figure 2.14. Probability Density Function of Neutral Scores

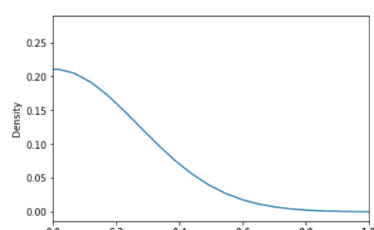


Figure 2.15. Probability Density Function of SMOG Scores

Table 2.12. Summary Statistics (Treated Group)

Variables	Mean	Std. Dev.	Min	Max
Before Treatment				
Monthly Review Volume†	2.131	3.784	1.000	214.000
Monthly Rating†	3.856	0.798	1.000	5.000
Rating	3.876	0.874	1.000	5.000
Monthly Length†	496.356	524.192	0.000	9,597.000
Length	518.229	643.062	0.000	14,103.000
Food Course	0.289	0.295	0.000	0.920
Dessert & Drink	0.381	0.320	0.019	0.939
Emoji	0.054	0.050	0.012	0.711
Service	0.156	0.220	0.013	0.892
Atmosphere	0.120	0.180	0.012	0.877
Positive Score	0.195	0.140	0.000	1.000
Negative Score	0.030	0.049	0.000	1.000
Neutral Score	0.775	0.137	0.000	1.000
SMOG Score	5.909	3.182	0.000	13.800
After Treatment				
Monthly Review Volume†	2.452	3.468	1.000	194.000
Monthly Rating†	3.763	0.751	1.000	5.000
Rating	3.851	0.877	1.000	5.000
Monthly Length†	563.351	595.456	0.000	13,707.000
Length	488.846	569.504	0.000	20,748.000
Food Course	0.298	0.292	0.013	0.915
Dessert & Drink	0.416	0.321	0.018	0.940
Emoji	0.050	0.046	0.011	0.717
Service	0.150	0.211	0.011	0.883
Atmosphere	0.106	0.160	0.010	0.892
Positive Score	0.183	0.119	0.000	1.000
Negative Score	0.030	0.044	0.000	1.000
Neutral Score	0.787	0.117	0.000	1.000
SMOG Score	6.476	2.757	0.000	17.500
At Treatment Time				
Prior Vol*	14.490	24.503	1.000	310.000
Prior Rate*	3.929	0.559	1.000	5.000

Note: †: a monthly-level variable; *: cumulative variable; otherwise, review level or non-cumulative.

Table 2.13. Summary Statistics (Control Group)

Variables	**Mean**	**Std. Dev.**	**Min**	**Max**
Before Treatment				
Monthly Review Volume†	1.104	3.987	1.000	532.000
Monthly Rating†	3.829	0.792	1.000	5.000
Rating	3.851	0.878	1.000	5.000
Monthly Length†	395.028	515.798	0.000	15,645.000
Length	488.738	569.374	0.000	42,748.000
Food Course	0.152	0.216	0.000	0.998
Dessert & Drink	0.207	0.306	0.000	0.997
Emoji	0.214	0.225	0.000	0.998
Service	0.198	0.235	0.000	0.998
Atmosphere	0.229	0.226	0.000	0.998
Positive Score	0.182	0.124	0.000	1.000
Negative Score	0.031	0.046	0.000	1.000
Neutral Score	0.789	0.123	0.000	1.000
SMOG Score	6.239	2.897	0.000	17.500
After Treatment				
Monthly Review Volume†	0.986	3.867	1.000	221.000
Monthly Rating†	3.826	0.781	1.000	5.000
Rating	3.774	0.913	1.000	5.000
Monthly Length†	389.531	520.134	0.000	41,600.000
Length	517.690	591.492	0.000	20,748.000
Food Course	0.151	0.214	0.000	0.998
Dessert & Drink	0.212	0.304	0.000	0.997
Emoji	0.216	0.228	0.000	0.998
Service	0.196	0.234	0.000	0.998
Atmosphere	0.227	0.225	0.000	0.998
Positive Score	0.179	0.134	0.000	1.000
Negative Score	0.033	0.041	0.000	1.000
Neutral Score	0.795	0.128	0.000	1.000
SMOG Score	6.722	3.117	0.000	17.500
At Treatment Time				
Prior Vol*	19.448	24.584	1.000	247.000
Prior Rate*	3.893	0.503	1.000	5.000

Note: †: a monthly-level variable; *: cumulative variable; otherwise, review level or non-cumulative.

2.8.2 Additional Model-Free Evidence Figures

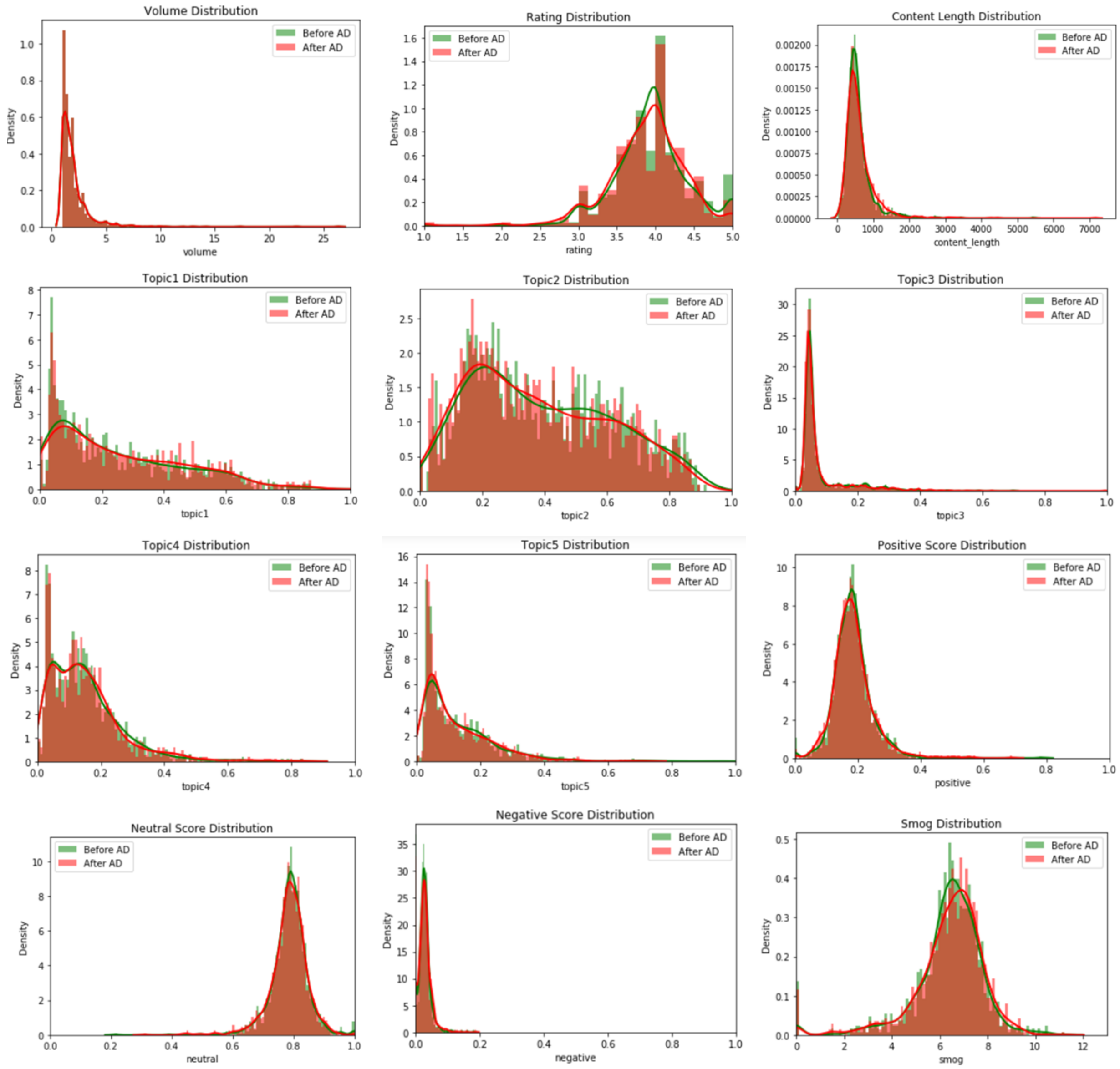


Figure 2.16. Review Characteristics Before and After Advertising Contracts

2.8.3 Additional Tables of Robustness Checks

Table 2.14. Controlling for Prior Reviews: Effect on Aggregated Review Characteristics

DV:	log(Review Volume)	Rating	log(Length)
Editorial Review	0.032***	-0.023*	0.034**
	(0.007)	(0.014)	(0.016)
Prior Vol	-0.000	0.000**	-0.001***
	(0.000)	(0.000)	(0.000)
Prior Rate	0.031***	-0.312***	0.016
	(0.007)	(0.015)	(0.016)
Restaurant FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes
Observations	44,699	44,699	44,699
Adj R-Squared	0.324	0.147	0.156

Note: robust standard errors in parentheses;* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.15. Controlling for Prior Reviews: Effect on Textual Review Characteristics

DV:	Food Course	Dessert &Drink	Emoji	Service	Atmosph-here	Positive	Neutral	Negative	SMOG
Editorial Review	0.011***	0.033***	-0.005**	-0.015***	-0.024***	-0.003*	0.004**	0.001	0.028
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.041)
Prior Vol	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Prior Rate	0.003	-0.005	0.006**	-0.002	-0.003	0.000	-0.001	0.000	-0.043
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.045)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	82,857	82,857	82,857	82,857	82,857	82,857	82,857	82,857	82,857
Adj R-Squared	0.437	0.371	0.293	0.166	0.201	0.058	0.054	0.021	0.084

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.16. Improved Controlling for Prior Review: Effect on Aggregated Review Characteristics

DV:	log(Review Volume)	Rating	log(Length)
Editorial Review	0.033***	-0.027*	0.036**
	(0.007)	(0.015)	(0.016)
Prior Vol	0.000	0.000	-0.001***
	(0.000)	(0.000)	(0.000)
Prior Rate	0.011	-0.352***	-0.003
	(0.008)	(0.017)	(0.019)
Restaurant FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes
Observations	44,699	44,699	44,699
Adj R-Squared	0.323	0.147	0.156

Note: robust standard errors in parentheses;* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.17. Improved Controlling for Prior Review: Effect on Textual Review Characteristics

DV:	Food Course	Dessert &Drink	Emoji	Service	Atmosphere	Positive	Neutral	Negative	SMOG
Editorial Review	0.009***	0.033***	-0.005**	-0.014***	-0.024***	-0.003*	0.004**	0.001	0.028
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.042)
Prior Vol	-0.000	-0.000**	0.000	0.000	0.000**	-0.000	0.000	0.000	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Prior Rate	0.005	-0.008*	0.008***	-0.002	-0.004	-0.002	0.001	0.001	-0.055
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)	(0.001)	(0.052)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	82,857	82,857	82,857	82,857	82,857	82,857	82,857	82,857	82,857
Adj R-Squared	0.436	0.370	0.289	0.167	0.197	0.058	0.054	0.021	0.085

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.18. Restaurants with Advertising Contracts: Effect on Aggregated Review Characteristics

DV:	log(Review Volume)	Rating	log(Length)
Editorial Review	0.039***	-0.016*	0.036*
	(0.011)	(0.008)	(0.022)
Restaurant FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes
Observations	19,428	19,428	19,428
Adj R-Squared	0.337	0.145	0.138

Note: robust standard errors in parentheses;* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.19. Restaurants with Advertising Contracts: Effect on Textual Review Characteristics

DV:	Food Course	Dessert &Drink	Emoji	Service	Atmosphere	Positive	Neutral	Negative	SMOG
Editorial Review	0.008**	0.007*	-0.006***	-0.010**	-0.009***	-0.004**	0.004**	<0.001	0.068
	(0.004)	(0.004)	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)	(0.050)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	44,990	44,990	44,990	44,990	44,990	44,990	44,990	44,990	44,990
Adj R-Squared	0.408	0.410	0.336	0.107	0.207	0.090	0.080	0.018	0.095

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.20. Restaurants with Advertising Contracts: Review-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.015***	-0.005***	-0.004**	-0.000	-0.155***	-0.022**	-0.045***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.044)	(0.010)	(0.015)
Prior Vol	-0.000***	0.000***	0.000***	-0.000***	0.003***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	-0.007*	0.001	0.001	0.000	0.117*	-0.003	-0.118***
	(0.004)	(0.002)	(0.002)	(0.001)	(0.060)	(0.013)	(0.020)
Ad	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	37,328	37,328	37,328	37,328	37,328	37,328	37,328
Adj R-Squared	0.200	0.119	0.120	0.035	0.143	0.681	0.150

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.21. Restaurants with Advertising Contracts: Reviewer-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.013***	-0.019***	-0.016***	-0.001	-0.181***	-0.108***	-0.045***
	(0.005)	(0.003)	(0.003)	(0.001)	(0.070)	(0.020)	(0.021)
Experience	0.007***	-0.002*	-0.001*	-0.002	-0.139**	-0.036***	-0.085***
	(0.003)	(0.001)	(0.001)	(0.002)	(0.080)	(0.015)	(0.026)
Prior Vol	0.000***	0.000	0.000	-0.000	0.001	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	-0.003	-0.003	-0.003	-0.003*	0.168**	-0.011	0.096***
	(0.005)	(0.003)	(0.003)	(0.001)	(0.080)	(0.022)	(0.024)
Ad	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,660	21,660	21,660	21,660	21,660	21,660	21,660
Adj R-Squared	0.888	0.854	0.863	0.823	0.856	0.957	0.923

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.22. Pooled Restaurants: Effect on Aggregated Review Characteristics

DV:	log(Review Volume)	Rating	log(Length)
Editorial Review	0.047***	-0.025**	0.021*
	(0.005)	(0.011)	(0.012)
Restaurant FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes
Observations	64,102	64,102	64,102
Adj R-Squared	0.334	0.141	0.160

Note: robust standard errors in parentheses;* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.23. Pooled Restaurants: Effect on Textual Review Characteristics

DV:	Food Course	Dessert &Drink	Emoji	Service	Atmosph-here	Positive	Neutral	Negative	SMOG
Editorial Review	0.008***	0.021***	-0.006***	-0.006***	-0.017***	-0.002**	0.002**	<0.001	0.023
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.031)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	124,684	124,684	124,684	124,684	124,684	124,684	124,684	124,684	124,684
Adj R-Squared	0.426	0.392	0.310	0.149	0.210	0.070	0.064	0.020	0.091

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.24. Pooled Restaurants: Review-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.004**	-0.003***	-0.003***	0.000	-0.079***	-0.016**	-0.047***
	(0.002)	(0.001)	(0.001)	(0.000)	(0.027)	(0.007)	(0.009)
Prior Vol	-0.000**	0.000*	0.000***	-0.000	0.001***	0.000***	0.001***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Prior Rate	-0.005**	0.001	0.002	-0.000	0.069*	0.003	-0.135***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.036)	(0.009)	(0.012)
Ad	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	89,306	89,306	89,306	89,306	89,306	89,306	89,306
Adj R-Squared	0.203	0.148	0.149	0.059	0.171	0.652	0.147

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.25. Pooled Restaurants: Reviewer-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.008***	-0.007***	-0.007***	-0.001	-0.201***	-0.070***	-0.051***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.046)	(0.013)	(0.014)
Experience	0.004***	-0.001**	-0.001*	-0.000***	-0.012*	-0.009***	-0.004**
	(0.000)	(0.000)	(0.001)	(0.000)	(0.007)	(0.002)	(0.002)
Prior Vol	-0.000	-0.000*	-0.000	-0.000**	0.000	0.001**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	-0.007*	0.001	-0.003	-0.001	-0.172***	0.084***	-0.017
	(0.004)	(0.002)	(0.002)	(0.001)	(0.058)	(0.017)	(0.017)
Ad	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	66,980	66,980	66,980	66,980	66,980	66,980	66,980
Adj R-Squared	0.920	0.894	0.898	0.881	0.901	0.961	0.940

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.26. Heterogeneous Effect of Editorial Reviews on Aggregated Review Characteristics

DV:	log(Review Volume)	Rating	log(Length)
Editorial Review	0.031***	-0.023*	0.121***
	(0.006)	(0.013)	(0.015)
Negative Score	-1.569	-9.058	-8.965
	(6.461)	(14.571)	(16.048)
Neutral Score	0.014	0.048	-0.029
	(0.158)	(0.356)	(0.392)
Restaurant FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes
Observations	48,027	48,027	48,027
Adj R-Squared	0.318	0.137	0.163

Note: robust standard errors in parentheses;* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.27. Heterogeneous Effect of Editorial Reviews on Textual Review Characteristics

DV:	Food Course	Dessert &Drinks	Emoji	Service	Atmosph-here	Positive	Neutral	Negative	SMOG
Editorial Review	0.014**	0.031***	-0.009*	-0.012*	-0.020***	-0.007*	0.005**	0.001	-0.005
	(0.007)	(0.008)	(0.005)	(0.007)	(0.007)	(0.004)	(0.002)	(0.002)	(0.004)
Negative Score	-0.442***	0.024	0.017	0.268	0.133	0.075	-0.145	0.071*	-0.145
	(0.160)	(0.183)	(0.123)	(0.165)	(0.148)	(0.095)	(0.094)	(0.037)	(0.094)
Negative Score	0.019**	0.001	0.000	-0.013	-0.008	-0.008	0.010*	-0.002	0.010*
	(0.010)	(0.011)	(0.007)	(0.010	(0.009)	(0.006)	(0.006)	(0.002)	(0.006)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	86,185	86,185	86,185	86,185	86,185	86,185	86,185	86,185	86,185
Adj R-Squared	0.436	0.368	0.290	0.166	0.200	0.059	0.054	0.021	0.054

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.28. Heterogeneous Effect of Editorial Reviews: Review-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.013***	-0.005***	-0.004**	-0.000	-0.149***	-0.104***	-0.087***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.076)	(0.019)	(0.026)
Prior Vol	0.000***	-0.000***	-0.000	0.000	-0.000	0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Prior Rate	-0.002	0.001	0.002	-0.001	0.057	0.014	-0.147***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.011)	(0.013)	(0.015)
Negative Score	0.219**	-0.214***	-0.070	-0.049*	-4.574***	-1.419***	-1.142**
	(0.097)	(0.064)	(0.064)	(0.027)	(1.516)	(0.378)	(0.508)
Neutral Score	-0.021***	0.003	-0.003	0.002	0.170	-0.124***	0.089***
	(0.007)	(0.004)	(0.004)	(0.002)	(0.103)	(0.026)	(0.035)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56,994	56,994	56,994	56,994	56,994	56,994	56,994
Adj R-Squared	0.206	0.165	0.165	0.071	0.189	0.618	0.144

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.29. Heterogeneous Effect of Editorial Reviews: Reviewer-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.060***	-0.011***	-0.012***	-0.003	-0.386***	-0.149***	-0.075**
	(0.007)	(0.004)	(0.003)	(0.002)	(0.145)	(0.032)	(0.030)
Experience	0.008***	-0.002*	-0.001*	0.000	-0.035***	-0.024***	-0.008**
	(0.001)	(0.001)	(0.001)	(0.000)	(0.013)	(0.004)	(0.004)
Prior Vol	-0.000***	-0.000***	-0.000***	-0.000	0.001	0.000	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	-0.011***	0.008***	0.006**	0.003***	-0.298***	0.064***	-0.202***
	(0.004)	(0.003)	(0.003)	(0.001)	(0.064)	(0.019)	(0.018)
Negative Score	-0.816***	-0.416***	-0.197**	-0.248***	10.403***	-1.091*	-0.061
	(0.131)	(0.082)	(0.082)	(0.035)	(2.022)	(0.613)	(0.585)
Neutral Score	-0.063***	0.016***	0.013**	0.005**	0.621***	0.036	0.238***
	(0.009)	(0.006)	(0.006)	(0.002)	(0.143)	(0.043)	(0.041)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51,372	51,372	51,372	51,372	51,372	51,372	51,372
Adj R-Squared	0.925	0.907	0.911	0.895	0.908	0.959	0.945

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.30. Restaurants with Similar ER: Effect on Aggregated Review Characteristics

DV:	log(Review Volume)	Rating	log(Length)
Editorial Review	0.024***	-0.008**	0.031**
	(0.006)	(0.004)	(0.015)
Restaurant FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes
Observations	45,047	45,047	45,047
Adj R-Squared	0.315	0.137	0.165

Note: robust standard errors in parentheses;
* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.31. Restaurants with Similar ER: Effect on Textual Review Characteristics

DV:	Food Course	Dessert &Drinks	Emoji	Service	Atmosph-here	Positive	Neutral	Negative	SMOG
Editorial Review	0.008***	0.033***	-0.006***	-0.011***	-0.024***	-0.003*	0.004**	0.001	-0.023
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.042)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	81,362	81,362	81,362	81,362	81,362	81,362	81,362	81,362	81,362
Adj R-Squared	0.423	0.362	0.280	0.172	0.193	0.059	0.054	0.020	0.094

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.32. Restaurants with Similar ER: Review-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.013***	-0.005***	-0.004**	-0.001	-0.077**	-0.114***	-0.049***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.035)	(0.008)	(0.012)
Prior Vol	0.000***	-0.000***	-0.000**	0.000	-0.000	0.000**	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Prior Rate	-0.003	-0.001	0.001	-0.000	0.024	0.000	-0.156***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.045)	(0.011)	(0.015)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	52,406	52,406	52,406	52,406	52,406	52,406	52,406
Adj R-Squared	0.217	0.179	0.175	0.078	0.198	0.622	0.152

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.33. Restaurants with Similar ER: Reviewer-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.006**	-0.006***	-0.007***	-0.002	-0.107**	-0.029**	-0.068**
	(0.003)	(0.002)	(0.002)	(0.002)	(0.047)	(0.014)	(0.013)
Experience	0.008***	-0.002*	-0.001*	0.000	-0.045***	-0.025***	-0.011***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.014)	(0.004)	(0.004)
Prior Vol	-0.000	-0.000***	-0.000***	0.000	0.002	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	-0.010**	0.007***	0.005*	0.003***	-0.384***	0.185***	-0.191***
	(0.004)	(0.003)	(0.003)	(0.001)	(0.064)	(0.018)	(0.018)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47,356	47,356	47,356	47,356	47,356	47,356	47,356
Adj R-Squared	0.931	0.918	0.920	0.903	0.913	0.962	0.951

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.34. Entropy Balancing: Effect on Aggregated Review Characteristics

DV:	log(Review Volume)	Rating	log(Content Length)
Editorial Review	0.032***	-0.026*	0.033**
	(0.005)	(0.014)	(0.013)
Restaurant FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes
Observations	573,473	573,473	573,473
Adj R-Squared	0.192	0.097	0.104

Note: robust standard errors in parentheses;* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.35. Entropy Balancing: Effect on Textual Review Characteristics

DV:	Food Course	Dessert &Drinks	Emoji	Service	Atmosph-here	Positive	Neutral	Negative	SMOG
Editorial Review	0.008***	0.029***	-0.007***	-0.009***	-0.021***	-0.002**	0.002**	<0.001	0.035
	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.001)	(0.001)	(<0.001)	(0.040)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	817,349	817,349	817,349	817,349	817,349	817,349	817,349	817,349	817,349
Adj R-Squared	0.325	0.297	0.261	0.122	0.178	0.047	0.051	0.020	0.075

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.36. Entropy Balancing: Review-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.014***	-0.004***	-0.003**	-0.000	-0.078**	-0.116***	-0.036***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.029)	(0.007)	(0.010)
Prior Vol	0.000***	-0.000***	-0.000**	0.000**	-0.001*	0.000**	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Prior Rate	-0.003	0.000	0.002	0.001	0.104***	0.021**	-0.135***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.038)	(0.009)	(0.013)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	76,258	76,258	76,258	76,258	76,258	76,258	76,258
Adj R-Squared	0.220	0.178	0.180	0.093	0.204	0.630	0.159

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.37. Entropy Balancing: Reviewer-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.011***	-0.005***	-0.003***	-0.001	-0.148***	-0.030***	-0.056**
	(0.002)	(0.001)	(0.001)	(0.001)	(0.030)	(0.009)	(0.009)
Experience	0.010***	-0.001*	-0.002***	-0.000	-0.034***	-0.021***	-0.013***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.009)	(0.003)	(0.003)
Prior Vol	-0.000***	-0.000***	-0.000***	-0.000	0.001**	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	0.004	0.016***	0.008***	0.002**	-0.146***	0.158***	-0.219***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.042)	(0.013)	(0.012)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	68,916	68,916	68,916	68,916	68,916	68,916	68,916
Adj R-Squared	0.960	0.951	0.952	0.945	0.952	0.978	0.971

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

2.8.4 Additional Mechanism Analysis Results

No/Alternative Control for Prior User Review Environment

We first show that the results are not sensitive to adding $PriorVol$ and $PriorRatings$ as control variables by repeat our mechanism analysis after dropping these two variables. The estimation results, shown in Table 2.38 and Table 2.39, are consistent with our main mechanism analysis results. Second, we change to the second approach of controlling for prior reviews. For all the user reviews of the control restaurants and of the treated restaurants before their treatments, we keep the same specification of $PriorVol$ and $PriorRatings$ as the main mechanism analysis, whereas for the "treated" user reviews, we set $PriorVol$ and $PriorRatings$ to be the values at the treatment time. The results remain the same, as shown in Table 2.40 and Table 2.41.

Table 2.38. No Control for Prior Reviews: Review-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial review	0.011***	-0.003***	-0.004***	-0.000	-0.159***	-0.023***	-0.034***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.031)	(0.008)	(0.011)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	62,368	62,368	62,368	62,368	62,368	62,368	62,368
Adj R-Squared	0.211	0.160	0.160	0.057	0.187	0.607	0.142

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.39. No Control for Prior Reviews: Reviewer-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.011***	-0.006***	-0.006***	-0.001	-0.163***	-0.036***	-0.041***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.045)	(0.013)	(0.013)
Experience	0.008***	-0.001***	-0.002***	-0.001***	-0.076***	-0.020***	-0.009**
	(0.001)	(0.001)	(0.001)	(0.000)	(0.014)	(0.004)	(0.004)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51,946	51,946	51,946	51,946	51,946	51,946	51,946
Adj R-Squared	0.923	0.909	0.914	0.906	0.907	0.957	0.944

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.40. Alternative Control for Prior Review: Review-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.005**	-0.008***	-0.002***	0.001	-0.131***	-0.025***	-0.036***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.033)	(0.008)	(0.011)
Prior Vol	0.000***	-0.000***	-0.000***	0.000**	-0.007***	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	0.000	-0.006**	-0.004*	0.001	-0.058	-0.002	-0.403***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.053)	(0.013)	(0.018)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56,994	56,994	56,994	56,994	56,994	56,994	56,994
Adj R-Squared	0.206	0.167	0.167	0.071	0.190	0.618	0.149

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.41. Alternative Control for Prior Review: Reviewer-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.016***	-0.004**	-0.005***	-0.001	-0.224***	-0.036***	-0.033***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.045)	(0.013)	(0.010)
Experience	0.008***	-0.002***	-0.002***	-0.000	-0.058***	-0.022***	-0.004**
	(0.001)	(0.001)	(0.001)	(0.000)	(0.013)	(0.004)	(0.003)
Prior Vol	-0.000***	-0.001***	-0.001***	-0.000	-0.001	-0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	-0.012***	0.012***	-0.001	-0.005***	-0.595***	0.234***	-0.119***
	(0.004)	(0.004)	(0.004)	(0.002)	(0.095)	(0.029)	(0.022)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51,372	51,372	51,372	51,372	51,372	51,372	51,372
Adj R-Squared	0.925	0.912	0.916	0.906	0.907	0.957	0.925

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

No/Alternative Control for Review Volume

We first show that the results are not sensitive to the bootstrap approach by repeating our mechanism analysis without the bootstrap. The estimation results, shown in Table 2.42 and Table 2.43, are consistent with our main mechanism analysis results.

Alternative to the use of bootstrap, we include $ReviewVolume$ as an additional control variable in Equation (2.10) to rule out the alternative explanation of increased volume of user

Table 2.42. No Control for Review Volume: Review-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.007***	-0.004**	-0.003***	0.000	-0.158***	-0.023**	-0.024*
	(0.003)	(0.002)	(0.001)	(0.001)	(0.041)	(0.011)	(0.014)
Prior Vol	0.000***	-0.000	-0.000	0.000**	-0.000	-0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Prior Rate	-0.002	0.002	0.003	-0.001	0.033	0.005	-0.181***
	(0.004)	(0.002)	(0.002)	(0.001)	(0.058)	(0.015)	(0.020)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	37,764	37,764	37,764	37,764	37,764	37,764	37,764
Adj R-Squared	0.192	0.144	0.136	0.029	0.147	0.610	0.110

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.43. No Control for Review Volume: Reviewer-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.015***	-0.004**	-0.006**	-0.001	-0.087*	-0.028**	-0.036***
	(0.005)	(0.002)	(0.003)	(0.003)	(0.051)	(0.014)	(0.018)
Experience	0.005*	-0.001*	-0.001*	-0.000	-0.038*	-0.004*	-0.008*
	(0.001)	(0.000)	(0.001)	(0.001)	(0.022)	(0.002)	(0.004)
Prior Vol	-0.000	-0.000	-0.000***	-0.000	0.001	-0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.001)	(0.001)
Prior Rate	-0.017	0.003	0.006**	0.006	-0.219	-0.008	-0.162**
	(0.015)	(0.009)	(0.003)	(0.004)	(0.231)	(0.066)	(0.067)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,375	31,375	31,375	31,375	31,375	31,375	31,375
Adj R-Squared	0.466	0.392	0.402	0.260	0.908	0.748	0.601

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

reviews for the estimation results. As shown in Table 2.44 and Table 2.45, the estimation results are, once again, consistent with the main mechanism analysis results.

Salience Effect

We test whether subsequent user reviews become more similar to any top-placed reviews, i.e., the salience effect. As the effects of editorial reviews have been shown, we focus on the effects of top-placed non-editorial reviews. Specifically, we select restaurants with top-placed user reviews without editorial reviews because user reviews for restaurants with editorial

Table 2.44. Alternative Control for Review Volume: Review-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial review	0.007***	-0.003*	-0.003**	0.000	-0.082**	-0.027***	-0.026*
	(0.003)	(0.002)	(0.002)	(0.001)	(0.041)	(0.011)	(0.014)
Prior Vol	0.000**	0.000	0.000	0.000*	0.000	0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Prior Rate	-0.002	0.001	0.002	0.001	0.011	0.003	-0.191***
	(0.004)	(0.002)	(0.002)	(0.001)	(0.058)	(0.015)	(0.021)
Review Volume	-0.000**	0.001***	0.001***	-0.000	0.016***	0.001*	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.001)	(0.001)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	37,764	37,764	37,764	37,764	37,764	37,764	37,764
Adj R-Squared	0.192	0.146	0.138	0.029	0.148	0.610	0.112

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

Table 2.45. Alternative Control for Review Volume: Reviewer-level Mechanism Analysis Results

DV:	Δ Topic	Δ Positive	Δ Neutral	Δ Negative	Δ SMOG	Δ Length	Δ Rating
Editorial Review	0.015***	-0.004**	-0.005***	0.000	-0.152*	-0.035***	-0.043*
	(0.006)	(0.002)	(0.001)	(0.003)	(0.086)	(0.013)	(0.023)
Experience	0.005*	-0.001**	-0.002***	-0.000	-0.039*	-0.008*	-0.014*
	(0.003)	(0.001)	(0.001)	(0.001)	(0.023)	(0.004)	(0.008)
Prior Vol	-0.000	-0.000	-0.000	-0.000	0.002	-0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.001)	(0.001)
Prior Rate	-0.017	0.004	0.006	0.005	-0.210	-0.006	-0.163**
	(0.015)	(0.009)	(0.009)	(0.004)	(0.231)	(0.066)	(0.067)
Review Volume	-0.000	0.000	0.000	0.000	0.010	0.002	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)	(0.003)	(0.003)
Restaurant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31,375	31,375	31,375	31,375	31,375	31,375	31,375
Adj R-Squared	0.466	0.392	0.402	0.260	0.334	0.748	0.601

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

reviews cannot be top-placed. Because user reviews are chronically sorted, at the time of writing a new user-review, the review listed next to it is the top-placed review. Therefore, for each review of the selected restaurant, we calculate the absolute value of difference in each user review characteristic between each pair of neighboring user reviews to capture the salience effect. Note that reviews may be inherently similar even without the salience effect. For a clear comparison, we calculate the same absolute value of difference between each user review and the 10^{th} previous user review and set these measures as a no-salience-effect benchmark (i.e., inherent similarity) since the first page only contains 6 reviews, and a review on the second page should no longer affect a new review through the salience effect. We then for each review characteristic contrast the measures containing the salient effect with the no-salience-effect benchmarks using a hedonic regression (equivalent to the student T-test) to statistically test for the salience effects. The results are shown in Table 2.46. We find that the salience effects of top-placed user reviews are not significant regarding any characteristics. Thus, the effects of editorial reviews is more than the salience effects of top-placed reviews.

Table 2.46. Salience Effect of Top-placed User Reviews

DV:	ΔRating	ΔLength	ΔPositi-ve Score	ΔNeutr-al Score	ΔNegati-ve Score	ΔSMOG	ΔFood Course	ΔDesse-rt&Drink	ΔEmoji	ΔServi-ce	ΔAtm-osphere
Salience Effect	0.010	-9.445	0.003	-0.003	-0.001	-0.195	-0.002	0.005	-0.003	0.003	-0.003
	(0.016)	(11.072)	(0.002)	(0.002)	(0.001)	(0.153)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	19,272	19,272	19,272	19,272	19,272	19,272	19,272	19,272	19,272	19,272	19,272

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

3. LET ARTIFICIAL INTELLIGENCE BE YOUR SHELF WATCHDOG: THE IMPACT OF AI-POWERED SHELF MONITORING ON PRODUCT SALES

3.1 Introduction

Artificial intelligence (AI) and machine learning (ML) are increasingly being used to develop foundations for the digital operating models within firms [62]. Firms such as Ant Financial, Amazon, and Netflix have demonstrated the value of such digital operating models. Even the literature has established the effectiveness of AI-driven models relative to human-driven ones [e.g., 63]–[68]. In particular, the benefits to implementing AI have been recognized as larger when AI is used in conjunction with humans [e.g., 69], [70]. While the impact of AI has been generally well studied, the mechanism that explains how AI offers these benefits has been less studied.

In terms of AI-human interactions, there are two main ways that AI impacts businesses. For one, AI replaces human tasks, as demonstrated by robotic process automation, which has been a primary focus of the literature that explores AI's impact. The other is that AI can also augment human tasks by expanding the scale and therefore, the scope of humans' decision-making. A rudimentary example of such a system is that of spreadsheets allowing accountants to shift their focus from the arithmetic of cash flow calculations to providing business insights. Even though this aspect of AI has been noted, except for some recent papers [e.g., 69], the literature has not extensively investigated issues related to the scale/scope augmentation of AI. Our paper focuses on this dimension by providing insights in the context of shelf management.

Shelf management is critical for fast-moving consumer goods (FMCGs) manufacturer. The ways in which products are displayed on shelves directly affect customer choice and serve as a key determinant of product sales [71], as an inappropriate shelf layout has the potential to kill a product/manufacturer [e.g., 72], [73]. As a result, retailers contractually agree with manufacturers regarding how shelf space will be used to showcase product(s). A typical contract between physical retailers and manufacturers specifies both shelf location and space, as well as the number of products to be displayed. Even restrictions such as displaying products at eye-level or making product labels visible are imposed by these contracts. To verify contract terms, manufacturers send their delegates to monitor shelf display in each store (i.e., retailers' compliance). In the past, delegates who represent manufacturers have successfully executed such contracts [e.g., 74], [75]; however, that dynamic is changing, as manufacturers attempt to reach a more global audience. In emerging retail markets, manufacturers are increasingly selling their products through a large number of independent retailers that have a higher degree of heterogeneity. Enforcing non-unified shelf display standards across independent retailers requires high scalability. Thus, more conventional shelf management practices that rely solely on human judgment lose their effectiveness. A natural solution to this problem would be to consider AI/ML to overcome these limitations. With respect to this solution, three key questions arise: a) Does the use of AI improve shelf management and, thus, product sales? b) If yes, what is the mechanism that explains how AI helps manufacturers generate improved outcomes? and c) What is the long-term success from implementing AI/ML? To answer these questions, we collaborated with a world-leading FMCGs manufacturer, Danone S.A. (hereafter referred to as "the manufacturer"), and conducted systematic, analytical experiments.

Our collaborative research includes three stages. In Stage 1, we address the overall impact and economic value for launching AI-powered shelf monitoring of product sales by using a quasi-experiment scheme and analyzing observational data. In Stage 2, we conduct a randomized field experiment to more precisely quantify the causal effect of launching AI-powered shelf monitoring; to do so, we include an additional treatment group that first is subject to AI-powered shelf monitoring and thereafter is terminated. This design is consistent with the literature on monitoring [e.g., 76] and allows us to examine the persistence of AI-powered monitoring, so we may understand the long-term adoption effect of this program. In Stage 3, we attempt to disentangle the underlying mechanism, and this process consists of four parts. First, we assume that the incremental value derived from AI-powered monitoring comes from the higher scalability of AI in dealing with heterogeneous objects as compared to solely using humans, and this higher scalability allows delegates to expand their monitoring scope to more heterogeneous retailers. Therefore, we expect a more salient impact when using AI-powered monitoring, as the correspondingly monitored shelves will be more heterogeneous. We test this assumption by looking at a heterogeneous treatment effect of AI-powered monitoring considering differentiated store types by their relative degree of contract heterogeneity. Second, we assume that the observed increase in product sales results from retailers' better compliance with shelf display requirements, and we therefore verify this expectation by using shelf photos to analyze the change in retailers' compliance. Third, we tease out the effect of only taking shelf photos. Finally, we qualitatively examine our quantitative findings and our underlying mechanism by interviewing the manufacturer's delegates.

Our analysis reveals several notable findings. First, all of our analyses consistently show that retail stores that implement AI-powered shelf monitoring programs boost their product sales. However, the gains in incremental sales drop after the termination of the program. We also show a diminished worth: treatment stores that launched AI-powered shelf monitoring programs sustained higher sales even when these monitoring programs were terminated, as compared to control stores that continued to use human-based shelf monitoring. These findings collectively suggest that the launch of an AI-powered shelf monitoring program greatly affects retailers' compliance behaviors, even after the monitoring is terminated.

More interestingly, we find that improved sales derived from AI-powered shelf monitoring are largely attributed to independent retail stores, which have more heterogeneous contracts and are more dominant in emerging markets. In contrast, these improved sales are insignificant for chain stores, which typically have more homogeneous contracts and are more common in mature markets. This finding supports our expectation as well as supports our decision to encourage Danone to experiment with this program in its emerging market. This finding further supports the scalability of AI and shows how this scalability can be expanded to repetitive tasks with heterogeneous objects, which are typically more challenging for humans to complete.

We further explores the causal chain between AI use and improved sales, so we may examine the change in retailers' compliance after the launch of AI-powered shelf monitoring. Specifically, we hypothesize that the use of AI could lead to more effective monitoring by delegates, and that this improved monitoring could result in retailers' better compliance with shelf display requirements, which would yield improved sales. Also, we find that the benefit of launching this AI-powered shelf monitoring system comes from AI rather than from taking photos. In other words, a placebo monitoring system that only takes shelf photos only will not boost sales.

Our study offers several contributions to the literature and to practitioners. With respect to the literature, we first extend the scope of the emerging literature stream on the business value/impact of AI by examining the effectiveness of AI-powered monitoring in shelf management contexts, and we do so by finding evidence that AI can be designed to complement human tasks. Second, we investigate behavioral persistence with respect to AI-powered shelf monitoring. Third, we explore the business circumstance/scope under which AI is more applicable due to its scalability. Finally, our results add to the literature stream on shelf management by focusing on the perspective of manufacturers (i.e., how manufacturers could effectively manage their products on retailers' shelves).

Our study further generates several managerial insights for practitioners, especially FMCG manufacturers, regarding their use of AI in their monitoring practices. The incremental value of AI-powered shelf monitoring indicates that practitioners could leverage AI to manage retailers' shelf displays, especially in contexts in which retailers are exhibiting a high degree of heterogeneity in terms of shelf conditions and contracts, which not only accounts for the majority in emerging markets but also accounts for those that exist in developed scenarios. The insights we derive with respect to AI-powered shelf monitoring programs' persistent influence and also with respect to our cost-benefit analysis should help practitioners understand how to apply AI and enact policies to ensure long-term success. More broadly, the rise of the sharing economy renders handling heterogeneous objects and tasks an increasingly salient challenge for managers seeking to address scalability. For example, Airbnb and Uber must manage millions of individual hosts and vehicles, respectively, that can be quite different from each other. Our findings suggest that such managers should use AI when their business scope involves managing heterogeneous objects, so they may benefit from the incremental value that AI scalability affords.

The rest of the paper is organized as follows. In Section 2, we survey the literature related to our study. In Section 3, we briefly describe Danone and our collaboration with this manufacturer. In Sections 4, 5, and 6, we elaborate upon the three steps of our investigation (i.e. quasi-experiment, field experiment, and mechanism analysis, respectively). In Section 7, we show a robustness check, and we use Section 8 to provide a cost-benefit analysis and conclude our paper.

3.2 Literature Review

Our research is closely to three literature streams. First, we extend the literature stream of retailer management and the business value of AI. Second, we further extend the body of literature on monitoring, particularly the emerging stream about using AI as a monitoring tool. Lastly, we contribute to the literature on shelf management. We review the mentioned literature streams as follows.

3.2.1 Retailer management and AI

Collaboration between manufacturers and retailers (i.e., retailer management) has been well recognized [e.g., 77]. A McKinsey report [78] categorizes collaboration efforts into four types: 1) supply-chain flows and processes, 2) demand planning and fulfillment, 3) promotion strategy, and 4) in-store layout/visual merchandising (including shelf management). The literature has demonstrated the value of IT in aiding the first three of the four types of collaborations. For instance, [79] shows that IT has significantly improved logistics strategy and practice. Similarly, [80] demonstrate an improved efficiency in the supply chain of an automatic replenishment program enabled by IT. [81] reveal that IT enables strategic pricing decisions for both manufacturers and retailers. However, to the best of our knowledge, the literature on improving visual merchandising is scarce, perhaps because visual merchandising/shelf management has been largely executed by people (e.g. delegates), thereby limiting traditional IT's role in this context.

The introduction of AI and its enhanced scalability [62] might change this domain of visual merchandising. The benefits of AI have been acknowledged in previous and recent studies in a variety of domains. For example, in finance, [66] show that AI applications could improve fraud detection and asset management. In healthcare, studies have explored how AI can help doctors diagnose cancers [e.g., 65], [82] and help hospitals improve operation efficiency [63]. In marketing, AI can be used to understand consumers' emotional status [67] and provide better customer service [68]. In addition, AI is thought to be most suited for tasks that need a lot of text, audio, image, and video processing [64]. Therefore, leading FMCGs manufacturers are seeking effective AI-based applications to manage their visual merchandising. Following [69], we help our collaborator, Danone S.A., develop and experiment with an AI-human monitoring assemblage (i.e., an image-recognition based monitoring system to assist delegates) to verify the compliance of its retailers with respect to shelf displays.

Thus, our investigation on the effectiveness of AI-based applications for managing retailers' shelf displays responds to a gap in our understanding of how IT helps retailer management, especially visual merchandising. In addition, we extend the scope of the emerging literature stream on business value/impact of AI to shelf management and draw generalizable insights into the business scope to indicate how AI-based monitoring might generate incremental values.

3.2.2 Monitoring

Monitoring, defined as the implementation of observation systems that gather information about individuals' activity or tasks [83], has been empirically studied in several contexts, including healthcare and production lines, among others [e.g., 76], [83], [84]. The effect of monitoring on process compliance captured in previous studies is mostly positive. For example, [76] find that individual electronic monitoring can dramatically improve caregivers' dispenser usage. [84] provides evidence that monitoring in onboard computers could improve truckers' driving performance. Similarly, [85] argue that monitoring can decrease the rate at which employees shirk because the perceived cost of opportunistic behavior is increased. [86] finds that monitoring can reduce teacher absenteeism. Further, [87] show that monitoring can not only improve the quality of customer service but also reduce employee theft.

In the labor-intensive retail industry, shelf management requires retailers' compliance with contracts, for which monitoring is needed. Such monitoring used to be executed by humans and can be biased [88]; however, with the development of technologies, the process of monitoring can be automated. For example, a radio frequency identification (RFID)-based system can be deployed to automatically monitor caregivers' hand hygiene compliance [76]. Similar RFID systems have been adopted to monitor inventory record inaccuracy [89]. AI may also be utilized to facilitate the automation of monitoring.

However, the role of AI in automated monitoring is not clear. We extend this literature by providing real-world empirical evidence that captures the positive impact of AI-powered monitoring in retailing, thereby showing that the incremental value of AI-powered monitoring comes from its improved scalability when objects being monitored are more heterogeneous.

Additionally, in behavior analysis, the persistence of behavioral interventions is worth understanding. Some interventions, in instance, result in long-term behavioral changes that can last even after the interventions are stopped, whereas others only produce short-term behavioral changes that fade once the interventions are stopped [90]. Behavioral persistence has been studied in several contexts. For example, [91] find that financial incentives for exercise can positively reinforce this behavior and that this behavioral change will last for a long time even after the incentives are removed (i.e., a long-term impact). Likewise, [92] show that social comparison-based home energy reports can reduce residents' energy usage. Even if reports are discontinued, the effect of reports is relatively persistent, decaying from 10 - 20 percent per year. Conversely, [93] argue that the level of labor effort increases due to gift exchange, but that the impact also wanes quickly. Finally, [66] provide empirical evidence that the impact of sending technical advice letters to household concerning water usage and consumption will stop after the treatment is ended.

However, the persistence of monitoring remains understudied. To the extent of our knowledge, the most relevant paper that shows the sustainability of the monitoring of employees is [76]. Specifically, in the context of healthcare, they study caregivers' hand hygiene behavior in hospitals by using observational data from Proventix, a company that offers a monitoring service. They find that healthcare workers' dispenser usage dropped severely following the removal of electronic monitoring. Further, the rate of dispenser usage was even lower than before monitoring was introduced.

Through a large-scale field experiment, we contribute to this stream of literature by inspecting the persistence of AI-powered monitoring on retailers' compliance behaviors. Since the additional monitoring effort typically incurs costs, the effect's persistence could determine how manufacturers should apply AI-powered monitoring (e.g., whether there is any need to continuously enforce such monitoring).

3.2.3 Shelf Management

Shelf management is the practice of managing existing shelf space, including deciding the amount of space to allocate to an item and determining the location of the item on a shelf. The existing literature in operations research focuses on the optimal product selection and shelf-space allocation from retailers' side [e.g., 94]–[96]. Specifically, numerous optimization models have been developed for retailers subject to constraints like store capacity [e.g., 96], product availability [e.g., 97], [98], and the strategic interactions between retailers and manufacturers [e.g., 99], [100]. Furthermore, the literature has proposed solutions for retailers to deal with shelves that are out-of-stock (OOS) and inventory record inaccuracy (IRI), such as using third-party audits [101], utilizing RFID [89], and implementing information-sensitive replenishment systems [102].

In fact, shelf management (i.e., managing products on retailers' shelves) is also crucial for manufacturers since a shelf display could greatly affect product sales [e.g., 103], [104] and, thus, manufacturers' revenue. For example, [71] show that for manufacturers that produce frequently purchased and branded grocery products, the number of product facings on retailers' shelves could significantly impact sales. Likewise, [73] argue that shelf layout (i.e., location and number of facings) have both a direct and indirect impact on sales. Further, [103] finds that additional shelf space allocated to a manufacturer could lead to higher product sales. However, this rule is not applied to manufacturers that produce staple products and impulse brands with low consumer acceptance.

Prior studies have advanced our understanding of shelf management from retailers' perspective and also illustrated the importance of shelf management for manufacturers. However, much remains to be understood regarding how manufacturers could effectively manage their products on retailers' shelves. Our research adds additional insights to shelf management from the manufacturers' point of view by suggesting AI-powered shelf monitoring systems should be used to increase retailers' compliance rates in shelf display.

3.3 Research Context

3.3.1 Background

We collaborate with Danone S.A, a world-leading multinational food product manufacturer. Its product line includes specialized nutrition, waters, and beverages, as well as dairy and plant-based products, which are FMCGs. The manufacturer follows the common practices of retailing in the FMCG industry. In particular, because FMCGs are typically pushed spatially close enough to potential consumers to achieve market penetration, manufacturers build a large-scale retail network consisting of an enormous set of brick-and-mortar retail stores, including delis, convenience stores, grocery stores, supermarkets, and so forth, all of which are evenly and densely distributed in every market.

To sell products in retail stores, our manufacturer negotiates with each retailer (e.g., chain-level merchandise managers or store owners) and enacts shelf space rental contracts with them. The contract ensures the display of products on the shelf and clearly specifies the quantity of the products to be displayed, along with which part of the shelf is rented to display those particular products. It is worth noting that the shelf space rental contracts are usually more than one year in duration. In our context, such contracts for each retail store are time-invariant during our study period.

To ensure that the shelf space rental contract is executed properly, a large number of delegates are utilized to check the shelf display. The delegates are full-time employees of our focal manufacturer who are only responsible for our collaborative manufacturer and the products it manufactures. Also, the delegates serve as important connections between the manufacturer and retailers. For example, delegates need to manually check whether the shelf space rental contract is enforced properly. If any non-compliance is found, then delegates can provide retailers with relevant feedback and impose penalties. In addition to shelf management, their responsibilities also include visiting stores, delivering ordered products to stores, taking orders, and collecting feedback from retailers on (at least) a weekly basis.

One delegate is responsible for around 150 retail stores (e.g., independent stores, chained grocery stores, chain supermarkets) in a specific coverage geographic area.

Establishing shelf space rental contracts that are solely monitored by delegates and that are designed to boost sales has evolved over the past century as a standard practice that has proven effective in mature markets in Europe and North America. However, due to differences in market structure, manufacturers now find it challenging to replicate this conventional practice in emerging markets, such as in Asia, Africa, and Latin America. As compared to those in mature markets, retail structures in emerging markets are more fragmented and decentralized. Chain stores, which are the major players in mature markets, only account for a small share of emerging markets; instead, manufacturers largely depend on independent retail stores owned by different owners to distribute products. Unlike chain stores, which can enact a unified standard and, thus, sign homogeneous contracts for product displays with a manufacturer, these independent retail stores have a higher degree of heterogeneity in shelf conditions (e.g. shelf layout, shelf availability). Thus, a contract is uniquely specified for each store.

The heterogeneity of the contracts makes the traditional practice of shelf monitoring potentially challenging. When contracts are more homogeneous, a delegate's examination might receive some scalability because the next display might be similar to the current one observed. The more shelves a delegate has checked, the faster the next monitoring will be, indicating a marginally diminished time cost. Indeed, an experienced delegate could coarsely check a shelf in minimal time; however, when contracts exhibit heterogeneity across retailers, delegates must treat each store independently, indicating a linearly increasing time cost with the growing number of stores under a delegate's responsibility. As a result, shelf monitoring in emerging markets that is conducted solely by delegates might have less scalability. Because the number of stores for delegates to monitor follows the similar practice to that in mature markets, delegates in emerging markets, facing the same time constraint, might be hard-pressed to fulfill the same monitoring task as their counterparts do in established markets.

Because of insufficient monitoring, retailers who behave like rational cheaters [85] might not comply with standards specified in contracts for the following reasons. First, it takes effort for retailers to comply with shelf space rental contracts. For instance, the owners of small independent stores might sometimes fail to place products correctly, and they might even prefer sacrificing profit to complying with their contracts. Second, retailers have different utility functions from that of our manufacturer. Our manufacturer only cares about its sales of focal products. However, retailers want to maximize sales for all products in the same category by optimizing product placement. Third, when shelf monitoring is not sufficiently effective, non-compliance behaviors can be easily ignored, which will not result in any penalties to retailers. As a result of retailers' failure to execute product display plans as specified in contracts, the sales of products by the manufacturer might be negatively affected.

To facilitate shelf monitoring in emerging markets, the manufacturer decided to explore the opportunities afforded by an AI-powered shelf monitoring program that is based on a mobile app[1]. This app is installed on delegates' company/working smartphones, and no CCTV infrastructure is required in any store by this system. Only delegates (rather than store owners) can use this mobile app (i.e., AI-powered shelf monitoring system) to take photos of shelves while visiting the (treatment) stores for which they are responsible. In addition to taking real-time photos of shelves, this app can also be used to upload photos to cloud storage such that manufacturer's products can be automatically detected. The detection, fulfilled by an image recognition model using AI-based computer vision technology, also identifies the number and location of each product. With this AI-powered shelf monitoring system, AI assists delegates to supervise a retailer's shelf display, which provides the manufacturer with a direct channel to monitor its rented shelves.

The introduction of AI-powered shelf monitoring dramatically changes the process of shelf management. In particular, delegates must no longer manually check to see whether a shelf display complies with a contract. Instead, the delegate can now use the app to take shelf photos and determine retailers' compliance based on AI-generated reports. At the same time, the store owners/employees can realize changes in the monitoring process by observing delegates' visits and practices in their stores or by inferring from receiving changed (e.g., more accurate) feedback from delegates about their compliance. The manufacturer expects that, with the assistance of AI, delegates can increase their capacity for shelf monitoring

[1]↑The cost of developing and maintaining this AI-powered shelf monitoring system is around 155,000 USD per year.

under time constraints, thereby making retailers more likely to display products as specified in their contracts with the manufacturer, all of which implies better sales.

3.3.2 Research Stages

Since AI-powered monitoring is relatively new, there is no existing anecdotal or academic evidence about its effectiveness. The manufacturer, thus, seeks the help of researchers to understand its incremental value. The investigation of AI-powered shelf monitoring programs' effectiveness consists of three stages. We briefly explain the context of those stages as follows.

Stage 1. We note that, prior to our participation in a field experiment, the manufacturer has applied AI-powered shelf monitoring in a group of selected retail stores. Therefore, we first collect the corresponding data from this period from Danone, so we may empirically investigate the effect of AI-powered shelf monitoring as a quasi-experiment. In the pilot study, the manufacturer selected around 2,500 retail stores in six cities in China, one of its emerging markets, to implement this AI-powered shelf monitoring program. The selection of treated stores is based on certain criteria (e.g., store location, gross floor area), and the majority of treatment stores are independent retail stores (e.g., delis, grocery stores), rather than chain stores (e.g., supermarkets). From the manufacturer's perspective, the monitoring of independent retail stores is more worthwhile because the disorganized shelf display that is often characteristic of independent retail stores makes it more difficult to monitor without the help of AI. In April 2019, the delegates who were responsible for the selected treatment stores were required to use the app, which was built by the manufacturer, to take time-stamped shelf photos once a month. Given that one delegate usually oversees around 150 stores, she/he could be in charge of both treatment and control stores simultaneously. With respect to the shelf monitoring of control stores, delegates would continue their conventional practice (i.e., human-based shelf monitoring).

Stage 2. Though quasi-experiments are powerful inferential tools, it is subject to several limitations. For example, it does not have an actual randomization of assigning treatment and control groups, which can thus undermine the validity of causal inference in observational settings [105]. To precisely gauge the causal impact of AI-powered shelf monitoring, we conduct a field experiment with the manufacturer. The field experiment further enables us to investigate the impact of AI-powered shelf monitoring at a deeper level. In particular, so we may better understand the persistence of the shelf monitoring impact, we create an experimental group with stores that were once monitored through AI-powered shelf monitoring, but that have no longer been monitored by AI after some time. We note that control stores and stores that stopped AI monitoring still received the conventional human-based monitoring. The field experiment includes around 3,800 retailers randomly selected from Danone's retailing network in six major cities in China and lasts from May 2019 to August 2019. The details of the experiment setup, randomization, and data collection are explained in Section 5.

Stage 3. We conduct our mechanism analysis in four phases. First, we extend our models in prior stages by considering the heterogeneity of store types, so we may test our proposed underlying mechanism that drives the realized change in product sales. Second, we study the change of retailers' compliance to complete our casual chain between the use of AI and improved sales. Third, we tease out the effect of only taking photos. Finally, following [106], we conduct a qualitative study by interviewing delegates, so we may further examine the underlying mechanism behind the main effect of launching AI-powered shelf monitoring on product sales that is captured in previous quantitative studies. In August 2019, our partner manufacturer assigned one of its employees as a research assistant to help us conduct the interviews [107].

3.3.3 Role of AI

Since the amount of allocated space and the position of products on shelves are well established as important determinants of sales [e.g., 71], [73], [103], [104] and are the two main aspects of shelf space rental contracts, our manufacturer uses the AI-powered shelf monitoring system and applies the following image recognition techniques to extract these two qualitative metrics from shelf photos.[2]

[2]↑The accuracy is greater than 98%.

Number of Facings. The concept of "facing" in retail refers to products that are placed at the front edge of a shelf with the brand label turned forward. For each photo, the number of facings is identified through the following steps. First, the app gathers substantial 360-degree product photos of our focal brand and others. Second, all product photos are transformed into constructs that can be logically analyzed by the computer using a feature detection algorithm. Third, the app trains a neural network with thousands of relevant photos and then obtains a classification model. Finally, delegates' shelf photos are fed into this classification model to recognize our focal products and then calculate the number of facings. A larger number of facings indicates that more space is allocated to the brand.

Position on the Shelf. Given that products on lower shelves usually receive less consumer attention than those placed at eye level, our manufacturer wants to examine the position of products on shelves in each shelf photo. The position of a product is measured by the vertical distance between the product and the ground floor. In general, if the product is placed at the height of 120 - 180 cm (i.e., 48 - 70 inches), it is considered to be in compliance with the manufacturer's shelf display standards. We collect information about the product's position on the shelf by using the following steps. First, we train a neural network to recognize tiers of shelves. Second, after identifying our focal products, the app can then calculate the shelf number upon which the product is placed. Third, the location of the product can be inferred, based on both the shelf height and product height.

3.4 Stage 1: Quasi-experiment

We leverage the manufacturer-conducted quasi-experiment to estimate the impact of AI powered shelf monitoring on product sales.

3.4.1 Data

Danone provided the store-level data from January 2017 to August 2019, which covers the period of its quasi-experiment. The data set contains information that pertains to each retail store including the monthly sales of each specific product in the product line of the manufacturer, store location (i.e., latitude and longitude), gross floor area, the number of cashiers, whether the store employs a point-of-sales (POS) system, and whether the store provides shopping baskets or carts. Further, for the stores that have been monitored by the AI-powered tool, we have access to the shelf photo-level data, which contains a photo's time stamp, the number of facings, and the shelf location of the focal product.

In total, the dataset includes records from 42,097 retail stores in the observation window. The quasi-experiment is conducted throughout April 2019 and 2,217 stores are treated with the AI-powered shelf monitoring program; the remaining stores continue to have the conventional human-based shelf monitoring as before. The dataset is constructed into a panel structure at the store-level such that each observation is a retail store and each time period is a month. Our main dependent variable $In(Sales)_{\mathrm{it}}$ denotes the log-transformed focal product sales in Chinese Yuan (¥) for store i in month t. Meanwhile, our independent variable of interest is the treatment indicator, $Monitoring_{\mathrm{it}}$, which indicates the status of the AI monitoring program for store i in month t. Later in this paper, we will apply a propensity score matching (PSM) scheme to refine our data. We will then provide descriptive statistics of our data after we complete the PSM process.

3.4.2 Identification Strategy

We note that the retail stores that the manufacturers selected as treatment stores may be significantly different from other control stores. Therefore, we use the propensity score matching (PSM) method to create a group of treatment stores that started AI-powered shelf monitoring in April 2019 ("treatment group") and a group of control stores that are similar to the ones in the treatment group but that did not adopt AI monitoring ("control group"). By doing so, we structure our data to mimic a controlled experiment, which allows us to control on multiple confounders and to tease out the causal effects of AI-powered shelf monitoring more precisely. Following that, we use the difference-in-difference (DID) technique as the regression specification for our quasi-experimental setting. This method is commonly used to estimate the effect of an exogenous treatment by comparing changes in outcomes across treatment and control groups over time [e.g., 108]. DID is suitable in our research context because we fundamentally observe a quasi-experimental design in which the treatment (i.e., AI-powered shelf monitoring) is applied to a subset of retail stores, whereas many other control stores do not receive this treatment. For treatment stores, product sales are expected to be affected by AI after the launch of this program, whereas, there is no such effect for control stores. In addition, the sales data also cover a long period before the treatment for all stores in both the treatment and control groups.

Propensity Score Matching

As previously mentioned, the selection of treatment stores that adopt AI monitoring is not purely exogenous. Since the manufacturer selected treatment stores based on several store characteristics, treatment stores may be different from their control counterparts. Therefore, we use PSM to eliminate the systematic differences between the treatment group and the control group [109]. Essentially, each store that launches the AI-powered shelf monitoring program is paired with a control store that is similar to the treated one in terms of its probability of being treated, which allows for a fair comparison between these two groups of retail stores. Because store characteristics rarely change over time, we utilize static matching [51] in our study. We first approximate store i's overall propensity score as a logistic function of a vector of store characteristics. The covariates in the vector consist of static characteristics including store location (i.e., latitude and longitude), gross floor area, the number of cashiers, whether the store employs a point-of-sales (POS) system, and whether the store provides shopping baskets or carts. These variables cover most of the observable and recorded characteristics of a store to the manufacturer. Our baseline matching utilizes one-to-one matching with replacement to derive the closest matched control store. Finally, we obtain 2,217 pairs of comparable stores with 2,217 treatment stores and 584 unique control stores in total. Summary statistics are provided in Table 3.1. In total, the dataset comprises 78,428 observations for 2,217 treatment stores and 584 control stores from January 2017 to April 2019. To evaluate the success of PSM, we show summary statistics for the treatment group and control group respectively in Table 3.2.

Table 3.1. Summary Statistics

Variables	Mean	Std. Dev.	Min	Max
Sales	936.102	622.676	0.000	12,419.800
Monitoring	0.028	0.166	0.000	1.000

Table 3.2. Summary Statistics

Variables	Mean	Std. Dev.	Min	Max
treatment group				
Sales	940.982	597.268	0.000	12,419.800
Control Group				
Sales	917.579	640.626	0.000	11,057.000

Difference-in-Difference

Next, we utilize the difference-in-difference approach to uncover the causal effects of AI-powered shelf monitoring on store sales. In particular, our model specification for the quasi-experiment setting is:

$$In(Sales)_{\mathrm{it}} = \alpha_{\mathrm{i}} + \gamma_t + \beta Monitoring_{\mathrm{it}} + \epsilon_{\mathrm{it}}, \tag{3.1}$$

in which $In(Sales)_{\mathrm{it}}$ refers to the log-transformed focal product sales of store i in month t. The independent variable of interest is the treatment indicator, $Monitoring_{\mathrm{it}}$, which is one if store i is in the treatment group and month t is after the launch of AI-powered shelf monitoring (i.e. April 2019). To account for the time-invariant, unobserved factors (e.g., store-level competition) that are specific to the retail store and that may affect product sales[3], we incorporate a store-fixed effects term α_{i}. To account for the time-variant unobservable factors (e.g., seasonality) that affect all retail stores, we include a time-fixed effects term γ_t. ϵ_{it} represents the error term. This specification identifies the impact of AI-powered shelf monitoring on focal product sales by the key parameter β. To reduce heteroscedasticity concerns, we leverage robust standard errors clustered at the store level [52].

[3]↑The competition level is essentially determined by the shelf displays of competing products, which are specified in the shelf space rental contracts between the competing manufacturers and individual stores. Since most competing FMCG manufacturers follow the practice of signing year-long contracts, we assume the persistent (store-specific) competition level within our observation window, especially after erasing seasonality through time-FE.

Relative Time Model

The validity of our previous identification strategy that leverages the matching and DID model relies critically on the pre-treatment parallel trend assumption (i.e., that there is no significant difference between treatment and control stores before the treatment). To test this assumption, we utilize the relative time model with the leads and lags periods [53]. With this model, we add a series of time dummies that indicate the relative chronological distance between observation time and treatment time. The specification of our relative time model is as follows:

$$In(Sales)_{\mathrm{it}} = \alpha_{\mathrm{i}} + \gamma_t + \sum_{\mathrm{j}} \tau_{\mathrm{j}} Pre_{\mathrm{it}}(\mathrm{j}) + \sum_{l} \omega_l Post_{\mathrm{it}}(l) + \epsilon_{\mathrm{it}}, \tag{3.2}$$

in which $In(Sales)_{\mathrm{it}}$ represents our focal product sales for store i in month t. α_{i} and γ_t denote store-fixed effects and time-fixed effects, respectively. The newly added term $Pre_{\mathrm{it}}(\mathrm{j})$ is an indicator function that equals one if month t is j month(s) prior to the treatment. Similarly, the term $Post_{\mathrm{it}}(l)$ is an indicator function that equals one if month t is l month(s) after the launch of the AI-powered shelf monitoring program. To estimate all the effects, we gather all pre-treatment periods that are greater than or equal to six months prior to treatment into one dummy. Hence, the coefficient τ_{j} for $j = \geqslant$-6, -5, ... , -1 captures the pre-treatment trend of our focal product sales while the coefficient ω_l for $l = 0$ captures the effect of shelf monitoring in a post-treatment period. Consistent with prior work, we set a period prior to the time of treatment as the baseline by normalizing the coefficient of $Pre_{\mathrm{it}}(-1)$ to zero.

3.4.3 Results

Relative Time Model

In Figure 3.1, we first plot the average product sales for matched stores (vertical) over the periods (horizontal). The plot shows that the sales of matched treatment and control stores follows the same trend until the intervention of AI-powered shelf monitoring, which validates the parallel pre-treatment trends assumption that is needed for our standard Matching + DID specification.

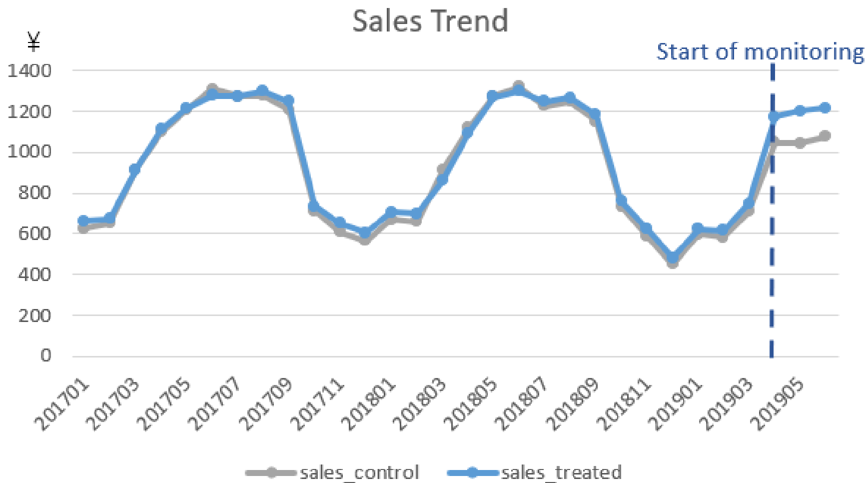


Figure 3.1. Parallel Trends in Matched Sample

We further report the results of the relative time model to establish that the parallel trend assumption holds in our analysis. As shown in Table 3.3, we find that none of the coefficients of the pre-treatment dummies (i.e., Pre(j)) are statistically significant at p-value < 0.10. This result confirms that there is no detectable pre-treatment dissimilarity across sales for stores that launch AI-powered shelf monitoring or for those that do not launch AI monitoring after the matching process; hence, the parallel trend assumption is satisfied.

The Effect of Launching Shelf Monitoring

The estimation results for the impact of AI-powered shelf monitoring on product sales, which are performed at the store-level, are shown in Table 3.4. We find a strong and significant increase in the focal product sales (i.e., $In(Sales)_{\text{it}}$) after the presence of AI-powered shelf monitoring. From an economic perspective, this translates to about a 15% increase in product sales per month for stores with AI monitoring, comparing to those that solely used human-based monitoring. This finding suggests the effectiveness of using AI-powered shelf monitoring to boost the sales of the manufacturer's products, which implies a better monitored shelf. One possible explanation for the change in $In(Sales)$ is that AI enhances the possibility for delegates to effectively manage shelf displays, thereby prompting retailers to comply with standards outlined in contracts. Higher retailers' compliance rates could

Table 3.3. Relative Time Model

DV:	lnSales(Yuan)
Pre (⩾-6)	-0.026
	(0.019)
Pre(-5)	-0.019
	(0.026)
Pre(-4)	-0.008
	(0.026)
Pre(-3)	-0.016
	(0.026)
Pre(-2)	0.023
	(0.026)
Pre(-1)	Baseline
Post(0)	0.118***
	(0.026)
Business FE	Yes
Time FE	Yes
Observations	78,428
Adj R-Squared	0.627

Note: robust standard errors in parentheses;
* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

yield higher product sales. The detailed mechanism analysis is elaborated upon in Section 6.

Table 3.4. Effect of Shelf Monitoring on Product Sales

DV:	lnSales(Yuan)
Monitoring	0.141***
	(0.019)
Store FE	Yes
Time FE	Yes
Observations	78,428
Adj R-Squared	0.627

Note: robust standard errors in parentheses;
* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

3.5 Stage 2: Field Experiment

Though the prior quasi-experiment provides an estimate of the causal effect of AI-powered shelf monitoring on sales, the robustness of the finding is subject to the concern of confounding bias (e.g., unobserved time-varying, store-specific factors) because we do not randomly assign retail stores to either the treatment or control groups. The validity of the causal inference is, therefore, under question [105]. To fundamentally address such concerns, we execute a randomized field experiment with the manufacturer to estimate precisely the causal effect of AI-powered shelf monitoring on sales.

In addition to checking the robustness of our previous finding, we further leverage the field experiment opportunity to investigate the effect's persistence and, thus, the long-term sustainability of the program. The persistence of the monitoring effect is an important dimension in the literature stream on behavior analysis [e.g., 66], [76], [91], as it helps researchers and policy makers to understand a program's long-term impact. We note that the additional monitoring effort typically incurs an additional cost; a monitoring program that has persistence, for example, as derived from the habit forming of the monitored object [e.g., 91], [92] indicates that a policy-maker could decrease the enforcement of or even terminate the monitoring policy over time; a policymaker might do so to minimize costs and make operations more sustainable, especially when the additional costs related to program enforcement is high. On the other hand, a monitoring program that has little or no persistence effect or that has an adverse effect once the program is terminated [e.g., 66], [76], [93] implies the need to enforce the monitoring program persistently, perhaps even with increasing intensity.

To investigate whether AI-powered monitoring requires persistent implementation and, thereby, requires long-term costs, we further investigate the persistence effect over the long run. In particular, we examine the product sales of retailers who are only treated by the program for a given period and who no longer make use of it thereafter. On the one hand, if retailers' habits are formed throughout the treatment period [91], the withdrawal of AI monitoring should have little or no immediate influence on product sales. If the launch of the AI-powered shelf monitoring program changes retailers' beliefs about the importance of shelf display, then a high level of compliance and, thus, a high level of product sales should persist after terminating shelf monitoring. On the other hand, if retailers are "rational cheaters" [85], one might expect that the removal of shelf monitoring would reset sales to where they were before the treatment.

We, therefore, conduct a field experiment with the cooperation of the manufacturer. In particular, using the large network of retailers in six cities in China where the delegates are equipped with the AI-powered shelf monitoring program, we first randomly assign 3,808 eligible subjects (i.e. retail stores) into one of three experimental groups, as presented in Table 3.5. The two treatment groups are differentiated by the launch and termination of AI-powered shelf monitoring. Throughout the experiment, 1,230 stores are assigned to a control group that uses human-based monitoring and never receives any AI-powered shelf monitoring. The other 2,578 stores are assigned to the two treatment groups. In May 2019, we informed the delegates of the 2,578 treatment stores to begin applying the program. In June 2019, we informed the delegates for 60 randomly selected treatment stores to terminate the AI monitoring program and switch back to human-based monitoring, while AI monitoring continued undisturbed in the remaining 2,518 treatment stores. The sample size of the group that terminated the AI monitoring (i.e. Group 2) is smaller than the group that continued the AI monitoring (i.e. Group 1) because of the manufacturer's concern that the frequent requests for changed practices might spark resistance among delegates. Except for the change of shelf monitoring, everything else remained the same.

Table 3.5. Experimental Groups

Experimental Group	Treatment	No. of Stores
Group 0	Hold-out group. No AI-powered shelf monitoring.	1,230
Group 1	Have AI-powered shelf monitoring from May 2019 to August 2019.	2,518
Group 2	Start AI-powered shelf monitoring in May 2019 and terminate it in June 2019.	60

3.5.1 Data

We tracked monthly sales for all 3,808 retail stores from May 2019 to August 2019. We use these data to examine how sales changed after the launch and termination of AI-powered shelf monitoring, which indicates a change in retailers' compliance over the long term. Here, we set one of the independent indicator variables $RemoveMonitoring_{\mathrm{it}}$ equal to one if the treatment store i terminated the AI monitoring by month t; otherwise, it equals zero. For the remaining treatment stores that continued AI monitoring before their final observations in our data and control stores, $RemoveMonitoring$ always equals zero. We provide summary statistics in Table 3.6. In total, the data set consists of 15,232 observations for 3,808 retail stores (1,230 stores in Group 0; 2,518 stores in Group 1; 60 stores in Group 2) from May 2019 to August 2019.

Table 3.6. Summary Statistics

Variables	**Mean**	**Std. Dev.**	**Min**	**Max**
Sales	1,379.732	893.741	37.170	16,611.440
Monitoring	0.981	0.135	0.000	1.000
RemoveMonitoring	0.019	0.135	0.000	1.000

3.5.2 Identification Strategy

The announcement of either terminating or continuing AI-powered shelf monitoring was made in a single day at the same time. This four-month field experiment allows us to create an exogenous shock to retailers' compliance behavior and thus product sales within a very short period. [4] The identification of the causal effect is, therefore, straightforward.

[4]↑We note that the partner manufacturer was not conducting any other parallel experiments; thus, our randomized experiment was the only one taking place from May 2019 to August 2019.

Terminate vs. Continue AI-powered Shelf Monitoring

To estimate the average treatment effect of AI monitoring termination on product sales, we apply the following regression model on the observations in Group 1 and Group 2:

$$In(Sales)_{\mathrm{it}} = \alpha_{\mathrm{i}} + \gamma_t + \beta RemoveMonitoring_{\mathrm{it}} + \epsilon_{\mathrm{it}}, \tag{3.3}$$

in which the dependent variable is the log of the volume of product sales for store i in month t. Our key exogenous variable is the termination of the AI-powered shelf monitoring program, which is represented by the treatment indicator $RemoveMonitoring$. If $RemoveMonitoring_{\mathrm{it}}$ equals one, it means that the store i has terminated AI monitoring by month t; if $RemoveMonitoring_{\mathrm{it}}$ equals zero, it means that the store i continued AI monitoring in month t. Its coefficient β can be interpreted as the change in product sales that is caused by terminating AI monitoring, compared with the product sales of the stores that continued AI monitoring. We also include a store fixed effect α_{i} and a month fixed effect γ_t as controls for unobserved factors.

Our experimental AI-powered shelf monitoring termination guarantees that the treatment indicator $RemoveMonitoring_{\mathrm{it}}$ is uncorrelated with the residual ϵ_{it} and that the ordinary least squares (OLS) estimator will satisfy the assumption of conditional mean independence, which is needed for convergence to the true parameter. Additionally, regressions could produce estimates with higher efficiency than those obtained from pairwise t-tests [110].

Terminate vs. Never Launch AI-powered Shelf Monitoring

To better capture the impact of the termination of AI-powered shelf monitoring, we further compare product sales in treatment stores that ceased AI monitoring activities with those in the control group (i.e., those that kept using human-based monitoring and never launched AI monitoring) by including all the observations of Group 0, 1, and 2. Building on Eq.(1), we add an additional variable, $RemoveMonitoring$, to examine how product sales changed in response to the termination of AI monitoring and then compare it with control group stores:

$$In(Sales)_{\mathrm{it}} = \alpha_{\mathrm{i}} + \gamma_t + \beta_1 Monitoring_{\mathrm{it}} + \beta_2 RemoveMonitoring_{\mathrm{it}} + \epsilon_{\mathrm{it}}. \tag{3.4}$$

The specification is identical to that for Eq.(1) except that we include two indicator variables, $Monitoring$ and $RemoveMonitoring$, to identify the effects of launching and terminating AI-powered shelf monitoring separately. $Monitoring$ equals one when AI monitoring is active and zero when it is not. The indicator variable $RemoveMonitoring$ equals one if the store has been treated and is now removed from the AI monitoring list; otherwise, $RemoveMonitoring$ equals zero. Thus, $RemoveMonitoring$ captures the OLS-estimates for the differences in product sales between control group stores and the Group 2 stores when their AI-powered shelf monitoring is discontinued. The key parameters of interest in this equation are β_1 and β_2.

3.5.3 Results

We assess the validity of our randomization by conducting t-tests for mean comparisons across observable store-level covariates. As shown in Table 3.7, we do not find any statistically significant differences between any pair of experimental groups, indicating that the randomization is at work.

Table 3.7. Randomization Check

Test Group	Sample Size	Store Characteristics			
		Floor area	Number of cashiers	POS system	Shopping baskets or carts
0	1,230	0.000	0.000	0.000	0.000
1	2,518	34.900	3.520	0.002	-0.001
2	60	20.100	-1.510	0.001	-0.003
p value for joint test Group 0 = 1 = 2		0.187	0.346	0.838	0.719

To respect NDA, the table provides demeaned values obtained by subtracting the mean value of the treatment group from that of the control group. Demeaning preserves the difference in mean value between test groups as well as the t-test.

We report the visualized evidence and the mean comparisons by experimental groups for our outcome of interest. Figure 3.2 presents the bar graphs of average product sales across experimental groups. The error bars in the graphs reflect 95% confidence intervals. We also report the results of t-tests with group mean comparisons in Table 3.8. We find that product sales are significantly larger in Group 1 (launched and continued AI-powered shelf monitoring), as compared to both Group 0 (not launched) and to Group 2 (launched and terminated AI monitoring). Additionally, we find that terminating AI monitoring significantly decreases product sales, relative to continuing AI monitoring. Finally, we also observe significantly higher product sales in Group 2, relative to Group 0. These findings suggest that the impact of the program might persist, although its strength might subside.

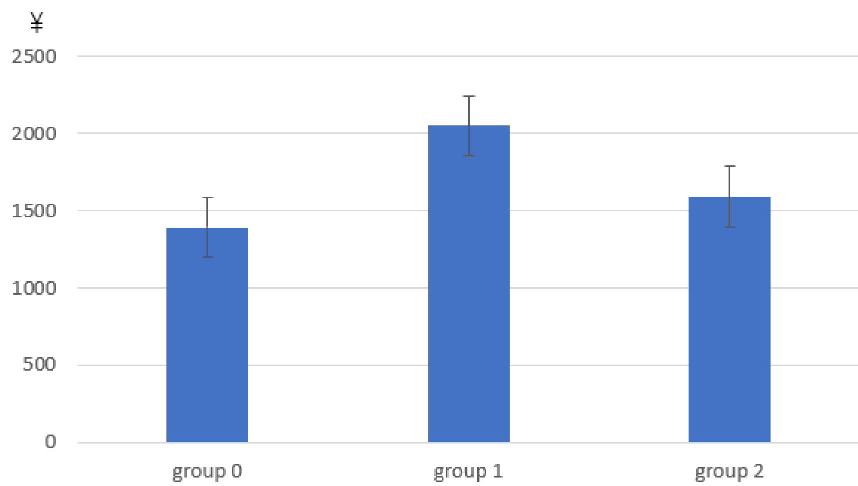


Figure 3.2. Mean Product Sales by Group

Table 3.8. Mean Comparison in Product Sales

Two-sample t-Tests	Mean Contrast	Diff.	t	p-value
Group 0 vs. Group 1	1394.850 vs. 2051.226	-656.376	-10.262	0.000
Group 0 vs. Group 2	1394.850 vs. 1592.589	-197.739	-5.194	0.000
Group 1 vs. Group 2	2051.226 vs. 1592.589	458.637	6.252	0.000

The treatment effects of termination identified from the difference between Group 1 and Group 2, as shown in Table 3.9, concurs with our previous finding. The termination of AI-powered shelf monitoring results in decreased $In(Sales)$. Quantitatively, product sales for stores that terminated AI-powered shelf monitoring are about 20% less than those that continued the program. This finding suggests that retailers operate as rational cheaters that change their compliance behaviors once the AI-powered monitoring is removed. Retailers might become lax with respect to shelf arrangement, which mfurther leads to disorganized product display and, thus, decreased sales.

Table 3.9. Average Treatment Effect of Terminating Shelf Monitoring

DV:	InSales(Yuan)
RemoveMonitoring	-0.229***
	(0.030)
Store FE	Yes
Time FE	Yes
Observations	10,312
Adj R-Squared	0.746

Note: robust standard errors in parentheses; * $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

As obtained in Eq.(4), product sales after AI monitoring termination, as compared to the baseline sales of the control group stores that never launched AI monitoring, are reported in Table 3.10. Consistent with previous results, a positive and significant coefficient on $Monitoring$ indicates that sales of the focal product increase following the launch of AI-powered shelf monitoring. More importantly, the positive and significant coefficient on $RemoveMonitoring$ suggests that Group 2 stores' sales after the termination are still higher, as compared to those of control group stores that never launch AI-powered shelf monitoring. Comparing the coefficient of $RemoveMonitoring$ with that of $Monitoring$, we could conclude that the termination, on average, decreases sales; however, increased sales obtained from launching the AI-powered shelf monitoring are still partially maintained. This finding, once again, indicates a partially persistent program effect. One possible explanation is that the mere existence of AI-powered shelf monitoring makes retailers realize that delegates are capable of effective shelf monitoring, and they therefore try to maintain shelf displays in compliance with contracts, at least for a while. However, such a program effect will not last over the long run as retailers are rational cheaters. Due to the diminishing nature of the program effect legacy, our findings also suggest that the manufacturer might better continuously execute the AI-powered shelf monitoring practice for the purpose of fostering more sales only if the marginal cost of applying such a program over the longer term is tolerable.

Later, we conduct a cost-benefit analysis to generate further practical insights regarding the incremental value of AI-powered shelf monitoring.

Table 3.10. Effect of Terminating Shelf Monitoring Relative to Baseline

DV:	lnSales(Yuan)
Monitoring	0.166***
	(0.008)
RemoveMonitoring	0.084**
	(0.033)
Store FE	Yes
Time FE	Yes
Observations	15,232
Adj R-Squared	0.730

Note: robust standard errors in parentheses;
* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

3.6 Stage 3: Mechanism Analysis

Thus far, the analysis does not provide insights on why product sales are boosted with the introduction of AI-powered shelf monitoring. In this section, we attempt to explain the underlying mechanism of the observed effect using four phases. First, we verify a proposed hypothetical explanation by extending our empirical analysis to find the differentiated treatment effects considering retail stores with different degrees of heterogeneity. Second, we utilize shelf photos to further explain the close relationship between AI and our observed increase in retail stores' sales. After that, we tease out the effect of only taking photos. Finally, we adopt a qualitative research approach and interview delegates of the manufacturer, so we may further examine our proposed explanation.

3.6.1 Effect by Types of Retail Stores

Our analyses have consistently shown a positive average effect of the launch of AI-powered shelf monitoring on product sales. We propose that this positive effect stems from delegates' enhanced ability to monitor more heterogeneous shelf conditions enabled by the scalability of AI.

Comparing to that of human beings, AI has exceptional working memory and tremendous calculation speed. Therefore, AI is capable of surpassing humans' ability to process routine tasks. For example, [68] find that AI chatbots are four times more effective than inexperienced workers at generating customer purchases. In addition, AI is shown to be efficient in processing heterogeneous objects [111]. In short, AI should competently manage repetitive but heterogeneous tasks.

Shelf monitoring can be considered routine work, since the monitoring task requires simply comparing the shelf display with the contract terms. In addition, as we describe in our research context, shelf monitoring in emerging markets must respond to more heterogeneous contracts among many independent retail stores. We hypothesize that AI could exceed human performance in coping with more heterogeneous shelf conditions.

Specifically, we expect that, given the time constraint and a fixed number of shelf monitoring tasks, human-based monitoring might be as effective as AI for more homogeneous contracts, whereas AI-powered monitoring might process heterogeneous contracts more efficiently than humans. This expectation might be explained by the scalability nature of AI: since the AI-based model is pre-trained, the marginal cost of handling a different contract might be minimal; however, a human delegate might need more time to learn and process the information when faced with a new contract that differs from earlier ones. The marginal cost of monitoring a new contract is, thus, linear. As a result, when facing time constraints and heterogeneous contracts, human-based monitoring might be coarse and inadequate, which gives retailers the opportunity not to follow contracts. This further leads to unmet sales targets, as stipulated in contracts. In contrast, AI empowers delegates to fulfill shelf monitoring for heterogeneous contracts in a limited time, making retailers more likely to enforce contracts, which results in higher product sales.

According to this proposed underlying mechanism, we expect that the positive effect of launching AI-powered shelf monitoring is mostly attributed to retail stores with more heterogeneous shelf conditions (e.g., independent stores). Because chain stores are associated with a homogeneity of contracts, human monitoring might perform as well in such stores because of the lower marginal time cost for an additional (albeit familiar and unified) contract.

Therefore, we might expect either an insignificant or significant effect but less magnitude for AI-powered shelf monitoring on the sales of stores with more homogeneous shelf conditions.

To test the proposed mechanism explanation, we use data from stores in our randomized field experiment and estimate the differential impact of AI-powered shelf monitoring on product sales in stores with different types of shelf conditions. In particular, we follow the standards determined by our manufacturer and divide stores into three general groups, based on the degree of contract heterogeneity (i.e., low, middle, and high). The classification standard is related to some business characteristics, such as store type (i.e., chain or independent) and the scale of stores. For example, large-scale chain stores (e.g., Walmart) or nationwide convenience stores (e.g., 7-Eleven) are considered to have low degrees of heterogeneity in shelf space rental contracts. Typically, medium-size local stores or regional chain stores with fewer than 5 branches are believed to have a middle level of contract heterogeneity. Finally, small-scale independent stores (e.g., street vendors) have the most heterogeneous contracts.

According to this classification, we add two interaction terms (i.e. $Monitoring \times Middle$ and $Monitoring \times High$) and replicate the DID model to capture the heterogeneous effects regarding the store types. Specifically, our model specification is:

$$In(Sales)_{\mathrm{it}} = \alpha_{\mathrm{i}}+\gamma_t+\beta_1 Monitoring_{\mathrm{it}}+\beta_2 Monitoring_{\mathrm{it}}\times Middle_{\mathrm{i}}+\beta_3 Monitoring_{\mathrm{it}}\times High_{\mathrm{i}}+\epsilon_{\mathrm{it}}, \tag{3.5}$$

in which all the specifications follow those in Eq.(1), except that we include two indicator variables $Middle_{\mathrm{i}}$ and $High_{\mathrm{i}}$, which equal 1 when store i belongs to a specific store type (i.e., stores with middle-/high-level contract heterogeneity).

For this analysis, we only focus on stores in Group 0 (i.e., stores that never launch the AI-powered shelf monitoring system) and Group 1 (i.e., stores that have continued using AI since May 2019). Our study period is from Jan 2017 to Aug 2019. Therefore, we have 119,936 observations from 3,748 stores in total. We report our results in Table 3.11. As seen, the introduction of AI has no significant impact on stores with lower levels of contract heterogeneity. More importantly, the impact of AI-powered shelf monitoring becomes larger as the degree of contract heterogeneity increases. This result further bolsters our claim that AI can better facilitate the shelf management of more heterogeneous contracts than can humans. The main reason is that stores with more homogeneous contracts (e.g., chain stores) are more likely to comply with shelf space rental contracts before the launch of AI. This explanation is theoretically possible. To be specific, we believe that human-based monitoring can be as effective as AI-powered monitoring for more homogeneous contracts. That is to say, human-based monitoring is relatively sufficient for stores like chain stores. Therefore, the introduction of AI cannot greatly affect chain stores' compliance rates. However, human-based shelf monitoring is shown to be insufficient for stores like independent stores. Since AI-powered shelf monitoring could process heterogeneous contracts more efficiently than humans, independent stores' compliance rates should improve after the launch of AI. Since the majority of retail stores in emerging markets are stores with high-level contract heterogeneity where no unified contracts can be signed, AI-powered shelf monitoring empowers delegates in terms of scalability and expands the monitoring scope of delegates to heterogeneous retail stores, which ultimately drives an incremental increase in sales.

Table 3.11. Effect of Launching AI Monitoring by Degree of Contract Heterogeneity

DV:	lnSales(Yuan)
Monitoring	0.103
	(0.093)
Monitoring × Middle	0.163*
	(0.094)
Monitoring × High	0.302***
	(0.093)
Store FE	Yes
Time FE	Yes
Observations	119,936
Adj R-Squared	0.734

Note: robust standard errors in parentheses;
* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

3.6.2 Change in Retailers' Compliance

We believe that the observed increase in sales is mainly due to the use of AI because AI could lead to more effective monitoring by delegates; in turn, better monitoring could result in retailers' better compliance in shelf display, thus improving sales. To better uncover this causal chain from the use of AI to improved sales, we leverage shelf photo-level data for 2,518 treatment stores in Group 1 of the field experiment and consider the impact of AI on retailers' compliance behavior in terms of shelf display. To measure retailers' compliance, we analyze the output photos and reports of the AI system. Notably, even though we have neither the photo nor data of compliance before the first photo was taken, we can nevertheless use the compliance shown in the first shelf photo of each retail store taken by the AI system to indicate the treatment of the AI system (i.e., a baseline), since by the time the first photo was taken, the shelf display and condition still reflected its original conditions without any AI intervention. In other words, the retailers at that time had not realized that delegates were using the AI system and also had not had the required time to take action in response to the policy change. We note that the AI system use was never directly revealed or notified to any retailer in both our quasi- and field experiments.

Given this setup, we then define and calculate retailers' compliance rates by comparing the shelf display shown in shelf photos and the required shelf display determined in shelf space rental contracts. In particular, we consider facing compliance rates that reflect the ratio of displayed number of products and that number required in the contract, as well as the position compliance rate that indicates whether the position of the displayed products is as required. We then apply linear and generalized linear regressions on these two variables, respectively, to test the difference of compliance before and after the use of AI system following:

$$Compliance_{\mathrm{i}t} = \alpha_{\mathrm{i}} + \gamma_t + \beta After\,AI_{\mathrm{i}t} + \epsilon_{\mathrm{i}t}, \tag{3.6}$$

in which $AfterAI_{\mathrm{i}t}$ equals 0 if the picture t of store i is its first ever picture, and equals 1 otherwise. This is the main independent variable of interest. The dependent variables include *Facing Compliance Rate* and *Position Compliance Rate*. Store fixed effect and time fixed effect are added using α_{i} and γ_t respectively. The results, shown in Table 3.12, suggest that both compliance rates significantly increase after the launch of AI-powered shelf monitoring. This finding provides direct evidence that shelf display actually follows contract requirements more closely (i.e., compliance rates increase) after the AI monitoring. This finding aligns with increased sales being caused by AI-powered systems, as we observed with our main analysis result.

Table 3.12. Compliance Rates Analysis

DV:	Facing Compliance Rate	Position Compliance Rate	Facing Compliance Rate	Position Compliance Rate
After AI	0.015*	0.025***	0.135**	0.314***
	(0.009)	(0.010)	(0.049)	(0.100)
Store FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	10,072	10,072	10,072	10,072
Adj R-Squared	0.257	0.255	0.135	0.333

Note: robust standard errors in parentheses;
* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

We shall, however, acknowledge the data limitation in this extension. Because there is no electronic record of retailers' compliance behaviors (e.g., photos) for the time periods before the launch of AI-powered shelf monitoring and for all retailers in the control groups, we could not replicate a treatment-control to estimate the causal effects of AI-based monitoring. Therefore, we summarize our interpretation of this results as simply "compliance increases after the launching of the AI-powered shelf monitoring."

3.6.3 Effect of Taking Photos

We further conduct a mechanism analysis to show that without AI, namely taking shelf photos only, does not improve sales. To do so, we aim to identify the effect of a treatment wherein the shelf photos were taken just as in standard AI-assisted shelf monitoring contexts, but for which the photos were neither processed by AI, nor would the delegates receive the AI-generated report about the compliance to use in their monitoring practices. Therefore, delegates were forced to monitor without AI assistance, just as their previous practice (i.e., human-based monitoring); however, the only difference now included that of taking pictures. The difference between this treatment's effect and that of the control group can be attributed to the effect of taking photos on shelves without either the AI analysis or the enabled monitoring. We name this effect as "the effect of taking photos only", for brevity's sake.

To investigate this effect, we leverage the interrupted use of the system at the first month of launching the AI-powered shelf monitoring system in our field experiment. Specifically, although delegates were required to start taking photos using the app for the treatment retailers in the experiment, the AI model hadn't been fully prepared to use. Delegates in the first month would not immediately receive the report generated by the back-end AI, which would reveal whether the shelf display in the photo complied with the contract. Instead, the AI-generated reports were delayed for two to four weeks. This delay, however, provides us with additional variation in the treatment. Treatment retailers may only be affected by delegates' taking photos, but by no other steps in the AI-based shelf monitoring (e.g., delegates' using the AI-generated compliance report to monitor and impose penalties), for various times in the month. Such variation allows us to separate the effect of taking photos only from that of the rest in the AI-based shelf monitoring on sales.

Notably, we do not know the exact delays for each individual retailer. Specifically, we do not know when the delegates use the reports once available to contact retailers and inform them about compliance check results, despite our knowing when pictures were taken and when reports were available. Failing to account for the delays may lead to an underestimate of the treatment effect, since untreated units at a specific time could be mischaracterized as treated, and vice versa. Therefore, we identify the upper and lower bounds of this special treatment with two extreme cases. The first case assumes all the AI-generated reports during the upgrade are delayed by two weeks. This assumption essentially results in an overestimated effect of the special treatment using the DID specification as our main analysis and, thus, provides us with our upper bounds. The second case assumes all the AI-generated reports during the upgrade are delayed by four weeks, which results in an underestimated effect and, hence, a lower bound of the interested effect.

More specifically, the empirical models are close to our main DID regression specification in Eq.(1), shown as:

$$In(Sales)_{\mathrm{it}} = \alpha_{\mathrm{i}} + \gamma_t + \beta_1 Photo_{\mathrm{it}} + \beta_2 AI_{\mathrm{it}} + \epsilon_{\mathrm{it}}, \tag{3.7}$$

in which $In(Sales)_{it}$, again, refers to the log-transformed focal product sales of store i in month t. We denote the proportion of month t of retailer i that was treated by "taking photo only" as $Photo_{\mathrm{it}}$, and the proportion of month t of retailer i that was treated by the rest of the AI-based shelf monitoring as AI_{it}. Both variables are bounded by 0 to 1, with 0 indicating not any treatment, which applies to stores in the control group as well as treatment stores before the launch of the AI-powered shelf monitoring system for both variables, and 1 indicating that the full month is treated (i.e., for any time period in the month, there must be a completed photo-taking or the rest of the AI-based shelf monitoring before that). Accordingly, the value of 0.5 for AI_{it}, for example, indicates that only the second half of the month t are after the delegates' conducting the rest of the steps of AI-based shelf monitoring. Notably, without the system interruption, $Photo_{\mathrm{it}}$ and AI_{it} are exactly identical (and the combination of them is also identical to the treatment dummy $Monitoring_{\mathrm{it}}$ in Eq.(1) and (4) of the main models) and, thus, not separately identifiable. In other words, the interruption that happened at the beginning provides us with the identification power to separate the effect of $Photo_{\mathrm{it}}$ from that of AI_{it}. We also include store fixed effect α_{i} and month fixed effect γ_t as controls for unobserved heterogeneity. β_1 identifies the impact of taking photos only on focal product sales, and β_2 captures the impact of the rest in AI-based shelf monitoring. To reduce heteroscedasticity concerns, we leverage robust standard errors clustered at the store level.

For this analysis, we include retailers in Group 0 and Group 1 of the field experiment, which never launches the AI-powered shelf monitoring system and keeps using AI since May 2019, respectively. Notably, the treatment-control assignments are randomized. The sales observation window is from Jan 2017 to Aug 2019. As a result, we have 119,936 observations for 3,748 stores in total. The estimation results are shown in Table 3.13.

We note that the action of taking photos has no significant impact on product sales with respect to both the upper bound value and the lower bound value, suggesting that the true effect of taking photo only should be insignificant. The positive effect of the AI-based monitoring, however, has all been absorbed by the effects of AI_{it}. Collectively, these results suggest that the observed increase in sales after the launch of the AI-powered shelf monitoring system should be attributed to AI rather than to the practice of only taking photos. That is, a placebo app with the feature of taking shelf photos only will not improve product sales.

Table 3.13. Average Effect of Taking Photos and Applying AI

	Upper Bound	Lower Bound
Photo	-0.052	-0.022
	(0.033)	(0.024)
AI	0.265***	0.175***
	(0.034)	(0.027)
Store FE	Yes	Yes
Time FE	Yes	Yes
Observations	119,936	119,936
Adj R-Squared	0.734	0.734

Note: robust standard errors in parentheses;
* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

3.6.4 Interview

Using a qualitative study approach, we offer supplementary evidence to help us uncover the underlying mechanism. Specifically, we conduct interviews with randomly sampled delegates who implement AI-powered shelf monitoring to draw additional insights with respect to this mechanism.

Procedure

More than 30 delegates who participated in the AI-powered shelf monitoring program were randomly selected and interviewed over a five-day period. Before interviews began, we carefully designed seven questions based on our discussion with the manufacturer, and we trained our research assistant (i.e., an employee from Danone) to conduct phone interviews. The details of the interview procedure follow. First, our research assistant briefly outlined the expectations of this particular interview to each participant (i.e. selected delegates). After this initial introduction, participants were asked to answer all interview questions and were also encouraged to share their thoughts by offering explicit explanations. The questions used in the interviews are summarized in Table 3.14. Such questions can lead participants to describe their opinions about the use of AI and AI-powered shelf monitoring at work. The interviews were conducted individually, and each interview was recorded using a digital recorder and then transcribed thereafter. Finally, the research assistant sent us all complted records.

Table 3.14. Interview Questions

Number	Question
1	Do you think the use of AI makes your job of shelf monitoring easier or harder?
2	How do you think this AI-powered shelf monitoring changes your cooperation with Danone?
3	How do you think this AI-powered shelf monitoring changes your cooperation with retail stores?
4	How do you think this AI-powered shelf monitoring changes product sales in retail stores?
5	How do you think this AI-powered shelf monitoring changes your monitoring process?
6	Do you feel any difference while using this AI-powered shelf monitoring system in different types of stores (i.e., stores with different degrees of contract heterogeneity)?
7	Do you have any comments or suggestions regarding this AI-powered shelf monitoring program?

Results

We obtained responses from 28 delegates who answered either the first six or all questions in the interview.

Our first question is related to the impact of the new technology, AI, on shelf monitoring. Few delegates were unsatisfied with the current AI technology because the manufacturer had developed two separate mobile apps (one for taking orders and one for shelf monitoring) for delegates to use, which complicates their typical practice. However, and more importantly, most delegates considered that AI/image recognition technique facilitates their capacity to check products placed on shelves. Specifically, AI makes shelf monitoring more effective by replacing manual checks. In addition, delegates reported that keeping reports generated from shelf photos allowed them to not provide more accurate and valuable feedback to retailers, but also urge retailers to comply with shelf display contracts from the outset.

Our second question is whether the AI-powered shelf monitoring produced any variation in the relationship between delegates and the manufacturer. Our results show that around 80% of delegates (i.e. 22 out of 28) believed that the AI-powered shelf monitoring benefited their relationship with the manufacturer. Specifically, the shelf monitoring program enabled by AI increased delegates' ability to better fulfill their duties, which were required by the manufacturer (i.e. monitor shelf display in a large number of independent retail stores).

Our third question concerned whether the collaborative relationship between delegates and retail stores changed due to the AI-powered shelf monitoring. Ninety percent of delegates (i.e., 25 out of 28) thought that this program helped them manage retail stores, especially in terms of shelf display. In particular, the AI-powered shelf monitoring pushed more retailers to follow the shelf space rental contract consciously (i.e., better compliance).

Our fourth question is about the direct connection between AI-powered shelf monitoring and product sales. Again, the majority of delegates (i.e. 25 out of 28) noticed major growth in product sales after the launch of AI-powered shelf monitoring. Various explanations for this phenomenon were proposed. For example, some delegates mentioned that the presence of AI-powered shelf monitoring enlarged their capacity to fulfill monitoring, thus making retailers more willing to comply with shelf display contracts. This may be because noncompliant behaviors of retailers were not only easily identified, but also well documented. Additionally, the manufacturer's efforts to manage shelves encouraged retailers to pay more attention to shelf displays.

Our fifth and sixth question are related to the mechanism of how the AI-powered shelf monitoring works in our context. In their responses, 90% of delegates mentioned that AI-generated reports improved their capacity to find retailers' non-compliant behaviors, which indicates that AI does improve retailers' compliance. Further, around 80% of delegates thought that the use of AI was more helpful when monitoring stores with more heterogeneous shelf conditions (i.e., high degree of contract heterogeneity), such as independent stores. This result confirms our claim that AI could generate more incremental value when addressing more heterogeneous objects because of scalability.

Discussion

The interviews complement our quantitative studies (i.e. quasi-experiment and field experiment) by providing evidence regarding a plausible mechanism underlying the change in product sales caused by AI-powered shelf monitoring. In short, the introduction of AI-powered shelf monitoring boosted product sales in emerging markets because of the following causal chain. AI (especially the AI-generated report) made it possible for delegates to fulfill shelf monitoring in emerging markets with higher degrees of heterogeneity in retail stores, which is difficult to achieve for human delegates who must contend with time constraints. The improved monitoring leads to retailers' higher compliance with contracts, and better presented products lead to improved sales.

3.7 Robustness Check

In this section, we demonstrate that our results for our quasi-experiment are not driven solely by the specific matching algorithm (i.e., the nearest neighbor matching) used in PSM. Here, instead, we adopt the coarsened exact matching approach, which is widely used in the literature [e.g., 112], [113], and we update our estimation results using the main DID regression specification. Our results, shown in Table 3.15, are consistent with our main estimation results, indicating their robustness to matching algorithms.

Table 3.15. CEM: Effect of Shelf Monitoring on Product Sales

DV:	lnSales(Yuan)
Monitoring	0.412***
	(0.020)
Store FE	Yes
Time FE	Yes
Observations	81,396
Adj R-Squared	0.669

Note: robust standard errors in parentheses;
* $p < 0.1$; ** $p < 0.05$; and *** $p < 0.01$.

3.8 Implications and Conclusion

3.8.1 Cost-Benefit Analysis

Our findings provide important policy implications regarding shelf management in emerging markets. Our focal manufacturer with markets in a number of developing countries recently proposed and implemented the AI-powered shelf monitoring system to improve shelf management. A key question is whether AI-powered shelf monitoring is a more strategic monitoring strategy to implement in terms of profitability. To address this question, we conduct a cost-benefit analysis to present the economic efficiency of this system and provide the estimates of net changes in profit.

The additional cost of developing and maintaining the AI-powered shelf monitoring system is not small, as it amounts to roughly 155,000 USD per year for the use of all delegates in our focal market[5]. According to our estimation result from our field experiment in Table 3.10, the launch of AI-powered shelf monitoring could increase product sales by 18%. This estimation is based on the DID design in a 4-month period in China. As a result, unless we make additional assumptions, our estimates should be understood as the Chinese market's short-term average treatment effect. To construct a cost-benefit analysis for longer terms and for other emerging markets, we must assume that the coefficients can be extrapolated to out-of-sample time windows and markets. We wish to underline that the benefit we present here should be interpreted with this assumption in mind because it is an untestable assumption in our study. To learn how much benefit could be brought by this AI monitoring system, we must know the yearly sales of our focal product in the Chinese market. For this analysis, we use the total sales in year 2020 (i.e., around 160 million USD) to obtain the total benefit for this policy, which is 28.8 million USD. Finally, we calculate the benefit-cost ratio using this number as the policy's benefit (≈ 185.81). We note that our benefit estimate is likely to be a lower bound estimate because our focal product in year 2020 experienced a slowdown in consumption amid the pandemic. As a result, the benefit-cost ratio is almost certainly a lower bound estimate. Our result suggests that the AI-powered shelf monitoring system is a profitable policy, even with our lower bound estimate of the policy's benefit.

3.8.2 Conclusion

Collaborating with Danone, we use three investigative steps (i.e., quasi-experiment, field experiment, and mechanism analysis) to explore the business value of using AI-based shelf monitoring. We show that the introduction of AI-powered shelf monitoring leads to a sig-

[5]↑The cost in monitoring by delegates is essentially part of the labor compensation (i.e., salary) for the delegates, which does no change throughout our study period.

nificant performance upgrade for delegates who manage shelves, which results in retailers' improved compliance with respect to shelf display. The improvement of retailers' compliance gives rise to a 15% - 18% increase in product sales, which is shown to be attributed by stores with high contract heterogeneity (e.g., independent stores) rather than stores with low contract heterogeneity (e.g., chain stores). This observation indicates that the scalability of AI can expand our manufacturer's monitoring scope to stores with more heterogeneous contracts. In addition, we observe that, after the AI-monitoring was terminated, product sales dropped; even then, however, sales remained higher than original levels when humans monitored shelves. These results indicate that the program effect is only partially persistent. We provide two plausible explanations for such positive change on product sales. Though the AI-powered shelf monitoring was terminated, the implementation of AI made it possible for retailers to realize the manufacturer's ability to validate the shelf management; in turn, retailers may have better perceived the importance of shelf display affecting sales and continued complying.

This research makes a substantial contribution to both the literature and practice. With respect to the literature, we contribute to the extant literature on the business value of AI by extending the scope to shelf management and providing evidence that AI can be designed to complement human processes, rather than replacing them. We also study the mechanism by which AI offers benefits by exploring the business circumstances under which AI-powered shelf monitoring is more appropriate than human-based monitoring, due to its greater scalability. More broadly, we explore the business scope under which AI generates incremental economic value/impact. In addition, we contribute to the literature on monitoring by extending the context to retailing, and we contribute to the literature on behavioral persistence by considering the change of product sales not only after the launch of AI-powered shelf monitoring, but also after the termination of such monitoring. By doing so, we help manufacturers understand how to best implement AI (e.g. shelf monitoring in our context) in a longer time horizon. Lastly, our study contributes to the literature regarding shelf management by focusing on the manufacturer's side.

Also, our research provides several valuable managerial insights for practitioners, especially FMCG manufacturers. First, the incremental value of AI-powered shelf monitoring indicates that practitioners could leverage AI to manage retailers' shelf displays, especially when retailers face a high degree of contract heterogeneity. Such heterogeneous retail stores not only account for the majority in emerging markets, but also exist in some developed markets as well. Second, our insights about the persistence of AI-powered shelf monitoring as well as our cost-benefit analysis could help practitioners understand how to apply AI correctly for long-term success. Finally, we extrapolate and discuss the broader implications for leveraging AI to cope with business challenges. Given the rise of the sharing economy, doing repetitive tasks with heterogeneous objects becomes an increasingly salient challenge in terms of scalability. For example, Airbnb must manage numerous house listings while Uber must manage various vehicles. Our findings suggest that AI use can expand the business scope when it involves managing heterogeneous objects.

4. WHEN DONATION MEETS REWARD: AN EMPIRICAL EXAMINATION OF CONTRIBUTION DYNAMICS IN CROWDFUNDING

4.1 Introduction

Crowdfunding is a viable method of raising capital through the collective effort of individual backers. It does not only greatly eliminate the hassle of searching for backers, but also allows fundraisers to get direct access to the market even before the products are well-developed. According to Statista, crowdfunding platforms helped global ventures raise nearly $30 billion in 2019. Therefore, it is worth studying the crowdfunding process.

In the crowdfunding process, individual backers play multiple roles simultaneously. To be specific, they may act as donors who invest their money to encourage fundraisers. At the same time, they may also act as rational reward buyers who maximize their utility by selecting final rewards. Additionally, all backers are socially influenced by each other. Such multifaceted roles of backers result in complex contribution dynamics.

Recent studies have investigated peer influence in donation-based and reward-based crowdfunding separately where backers only play two roles mentioned above [e.g., 114]–[116]. However, the literature is silent on the cross-channel peer influence, that is, the potential interplays among various crowdfunding behaviors (i.e., donating and reward pledging) when they coexist. Under that circumstance, backers can play all roles mentioned above.

Reward pledging is found to be driven by high project quality and high prospects of success [e.g., 14], [117]. Besides, [116] show that the prosocial motive to help fundraisers is another important driving factor. Since peers' donation might affect the prosocial motive of potential backers, we expect it will exhibit a different peer influence on backers' reward-pledging behavior other than the herding effect found in the two-role case. Similarly, peers' reward pledging could pass signals about project quality to subsequent donors and then influence their donating decisions. Therefore, we expect that the coexistence of donating and reward pledging might reshape the fundraising process significantly by the cross-channel peer influence. Understanding the potential interaction of donation and reward pledge is critical because it greatly affects what strategies the fundraisers and crowdfunding platforms should undertake to harness the power of peer influence.

In the crowdfunding process, fundraisers can apply various strategies to make a promising project. Such strategic behaviors of fundraisers could interact with backers' crowdfunding behaviors (i.e., donating and reward pledging) during both the project setup phase and afterward. For example, fundraisers could strategically set up and present the project information (e.g., detailed description and progress). Since, as shown by [118], the narratives of projects are used by investors to supplement additional objective sources of information when making financial decisions, such behavior could influence backers later on and affect the project's final success. In addition, during the fundraising phase, fundraisers might strategically contribute to their projects publicly. This contribution might convey signals that can be perceived by and affect other backers' contribution dynamics. The interaction between fundraisers' strategic behaviors and backers' crowdfunding behaviors will further complicate contribution dynamics.

The potential impact of fundraisers' behaviors on backers' contribution behaviors is not that straightforward. For project narratives, subjective content may attract backers by striking a chord. At the same time, subjective project narratives may repel backers since they cannot convey much objective project information to signal its quality. In addition, fundraisers' contributions might also be perceived differently by different potential backers. On the one hand, it may be considered as a signal showing that this project does not need much help and thus crowd the backers out. On the other hand, the contributions from fundraisers may show fundraisers' great effort and commitment to the project and thus encourage more backers to support it.

To our best knowledge, prior literature has yet considered the interplay between strategic behaviors from the fundraiser side and crowdfunding behaviors from the backer side. We, thus, take a further and giant leap to investigate how the fundraiser's project narrative and contribution behavior could affect the progress of the fundraising by backers. Note that, since fundraisers could directly manage their project narrative and fundraising behavior, on the practical side, our investigations could generate direct managerial implications for fundraisers to lead their project to success.

In short, both academics and crowdfunding practitioners must understand the dynamic contribution patterns caused by backers' multiple roles and fundraisers' strategic behaviors, which could provide substantive implications for individual backers, fundraisers, and crowdfunding platforms.

Therefore, in this work, we try to answer the following research questions to enhance our understanding of the contribution dynamics in crowdfunding. First, what effects do peers' prior contribution behaviors have on later backers' contribution decisions, especially the interplay of donating and reward pledging? Second, during the project setup phase, how could fundraisers set up their project' narratives (i.e., project descriptions) to attract more donations and reward pledge? Third, after the setup, what is the impact of contribution made by fundraisers themselves on the behaviors of potential backers?

To address those questions, we employ a proprietary dataset of detailed contribution records from a leading reward-based crowdfunding platform in China. Similar to all other reward-based platforms, backers could pledge money by selecting reward items and paying pre-determined reward prices. One unique design of this platform is that backers could also donate a customized amount to support crowdfunding projects without redeeming rewards. Since all backers' identities are revealed, any users could observe whether the fundraiser has donated/pledged his/her own project. The dataset consists of complete information on contributions conducted from November 2014 to May 2018 for 4,091 crowdfunding projects on our focal platform.

In the initial step of our analysis, we establish the specific project attributes that backers value most and so impact the aggregate demand for the projects. Beyond the directly observable quantitative characteristics provided on the project page (e.g., remaining budget, number of total backers, etc.), potential backers also tend to value project quality characteristics embedded in the project description, such as the subjectivity of project title and project introduction. We incorporate natural language processing (NLP) techniques to infer these textual features. In the second step, following [119] we use demand estimation techniques (i.e., the random coefficient logit model) to quantify the economic influence and relative importance of project characteristics. Furthermore, we quantify the impact of peer's crowdfunding behaviors and fundraisers' strategic behaviors on contribution dynamics.

Our analysis reveals several notable and interesting findings. First, we find evidence in support of the herding effect. As individuals observe others pledging/donating more money, the amount they are inclined to pledge/donate increases. Surprisingly, empirical evidence of negative cross-channel peer influence between donating and reward pledging is found. That is, more prior donating behaviors could lead to fewer reward pledging behaviors and vice versa. Second, projects with more subjective titles and introductions will repel potential donors. One possible explanation is that donors are more skeptical of subjective than objective narratives while evaluating projects [120]. Potential donors may perceive projects with the subjective description as less trustworthy because those projects do not convey much factual information that is useful. This finding yields a significant managerial insight: fundraisers could provide project narratives with more objective information to attract more donors and get more donations. Third, we find that if a fundraiser contributes to his/her own project, it will increase backers' intention to donate. One possible reason is that fundraisers' contributions can arouse donors' sympathy. At the same time, since reward buyers pay more attention to the creditworthiness of fundraisers and the quality of future products/services [14], fundraisers' contribution has no significant impact on the reward pledging of subsequent backers.

There are two key literature contributions of this paper. First, our study attempts to contribute to the stream of literature that explores the role of peer decisions in online crowdfunding. Though separately, peer influence in reward pledging and charitable giving have been widely studied, the potential interplay between these two types of crowdfunding behaviors (i.e., reward pledging and donating) has yet been investigated. We examine the dynamics of donating, reward pledging, and their interplays on a crowdfunding platform that allows both donation and reward pledge. Second, we add to the signaling literature. In the context of crowdfunding, the information between fundraisers and backers is asymmetric. Specifically, fundraisers are aware of the real quality of their projects while potential backers are lack of such information [121]. Therefore, it is necessary for fundraisers to pass enough signals to potential backers if they want to reach their funding goals. On our focal platform, fundraisers have three ways to convey project information to backers. First, a project page created by the fundraiser is available to the public. One typical project page includes detailed information of the fundraiser, a comprehensive description of the project, and real-time project status updates, etc. Second, fundraisers could actively reply to reviews posted by backers. Third, fundraisers are allowed to donate money or buy rewards from their own projects, which can be observed by potential backers in the contribution records on the project page. We investigate whether project narratives lead to different dynamics of backers' contributions. Besides, we further uncover the relationship between fundraisers' strategic contribution and subsequent backers' behaviors (i.e., both donating and reward pledging). This study provides significant managerial implications for practitioners on crowdfunding platforms. First, our findings suggest that donors are greatly affected by the project narratives offered by fundraisers. Fundraisers may attract more donors by conveying more factual project information and describe their projects in a more objective way. Second, we suggest that fundraisers who are in need of donations could fund their own projects as such behavior will increase other backers' motivation to donate. Third, our research points out a possible way for fundraisers who are interested in managing the herding effect. If fundraisers want to increase their chances of fundraising success, they should try to maximize the contribution they get in the early stage. For example, they could seek friends or family members for help at first. Last but not least, for the donation-based crowdfunding platform managers, we suggest incorporation of a feature that can reveal fundraisers' great effort to the project as it exhibits a positive impact on donating. For the reward-based crowdfunding platform managers, we suggest deleting the donation feature as it has a negative cross-channel influence on the reward pledging.

In the remainder of the article, we review relevant research about peer influence and signaling theory and summarize our contributions to the literature in Section 4.2. Next, we present details of our research context in Section 4.3 and empirical methods in Section 4.4. We discuss our results in Section 4.5. We finally conclude this paper in Section 4.6.

4.2 Literature Review

4.2.1 Peer Influence in Crowdfunding and Charitable Giving

There is an emerging stream of literature that has examined the concept of crowdfunding. Crowdfunding projects can be categorized into four types: equity-based, lending-based, donation-based, and reward-based [122]. Studies that have explored the role of peer decisions in online crowdfunding are most closely related to our research. For example, in the context of a lending-based market, [14] find the existence of herding behavior. In particular, lenders infer whether borrowers are trustworthy or not by observing peer lending decisions. On the contrary, small initial contributions are found to significantly decrease the number of backers and thus decrease the chances of success for a project [123].

The existence of peer influence, known as the direct influence on people by peers, has been widely identified in the context of charitable giving (i.e., donating). Shown as a strong motivator of donating decisions, peers' donating behaviors could have either a positive or negative impact on subsequent potential donors. On the one hand, both [124] and [125] find that individual donation increases with previous peer donation. On the other hand, [114] have documented negative peer influence. That is, higher donating frequencies could result in lower subsequent donating amounts. A possible underlying mechanism is that donors experience a decrease in the utility of donating when they perceive themselves as less important to the fundraiser.

Though separately, peer influence in reward pledging and charitable giving have been widely studied, the potential interplay between these two types of crowdfunding behaviors (i.e., reward pledging and donating) has yet been investigated. Our study attempts to contribute to this stream of literature by studying the dynamics of donating, reward pledging, and their interplays on a crowdfunding platform that allows both donation and reward pledge.

4.2.2 Information Asymmetry and Signaling in Crowdfunding

The signaling theory is proposed by [126] as an answer to one of the main causes of market failure – information asymmetry. In the context of crowdfunding, the information between fundraisers and backers is asymmetric. Specifically, fundraisers are aware of the real quality of their projects while potential backers are lack of such information [121]. Therefore, it is necessary for fundraisers to pass enough signals to potential backers if they want to reach their funding goals. On our focal platform, fundraisers have three ways to convey project information to backers. First, a project page created by the fundraiser is available to the public. One typical project page includes detailed information of the fundraiser, comprehensive description of the project, and real-time project status updates, etc. Second, fundraisers could actively reply to reviews posted by backers. Third, fundraisers are allowed to donate money or buy rewards from their own projects. Those fundraisers' contribution behaviors could also be observed by potential backers in the contribution records on the project page.

We add to the signaling literature in crowdfunding by investigating whether project narratives lead to different dynamics of backers' contribution using NLP to look into the information embedded in the text of project descriptions. Besides, we further show the impact of fundraisers' contribution on subsequent backers' behaviors (i.e., both donating and reward pledging).

4.3 Research Context and Data

4.3.1 Research Context

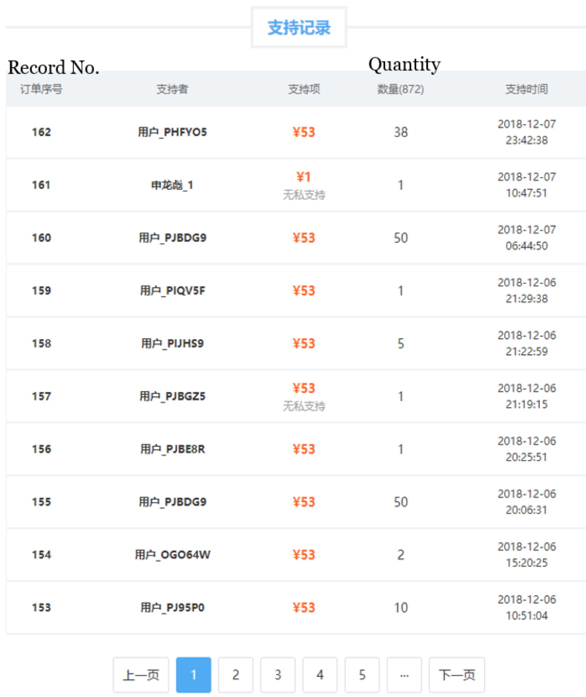
Our study is based on one of the largest Chinese platforms for reward-based crowdfunding. This platform, established in February 2013, enables anyone, without any limitation of location, to raise money for a venture or idea. When fundraisers submit their projects to the platform for crowdfunding, they must decide several project characteristics. Specifically, they have to decide project title, project introduction, the target amount, and the planned funding duration, etc. In addition, they need to determine a reward scheme (a menu of reward offering including reward items and prices) because fundraisers have to provide reward buyers with different rewards in return. In addition, one interesting feature of the platform is that fundraisers on this platform are allowed to donate money or buy rewards from their own projects.

When visiting a project page, a potential backer can easily get access to project details, cumulative prior contributions, and all contribution records. Once a backer has decided to contribute to a particular project, he/she can choose to either pledge money by selecting perks (i.e., reward items) or simply make donations without expecting to redeem rewards. After that, the information of this contribution will be automatically shown on the project page. Figure 1 presents a screenshot of contribution records for one project. As can be seen, the platform offers information on prior contribution behavior such as backer id, how much a backer contributes to this project, when the contribution happens, and whether it is donating. For the sake of simplicity, we multiply the price by the quantity of rewards to calculate the total amount of reward pledge for each reward pledging record. Since the default quantity of each donating record is one, we directly get the amount of donation by looking at the price. If the funding goal is met within a predetermined time window, reward buyers will get rewards delivered by fundraisers. If a project fails to meet its funding goal, all backers (i.e., reward buyers and donors) will get their money refunded.

4.3.2 Data

Data Description and Variable Definition

We obtain a proprietary dataset of the focal platform. The dataset consists of detailed information about 4,091 projects and 133,368 observations from November 2014 to May 2018. For each project, the data set contains project title, introduction, type, funding goal, fundraiser id, and all contribution records. For each contribution record, the data includes backer id, the amount of contribution, contribution type (i.e., reward pledge or donation) and the timing. We aggregate contribution data on a daily basis. Our key dependent variables are $Donation_{\mathrm{it}}$ and $RewardPledge_{\mathrm{it}}$. $Donation_{\mathrm{it}}$ represents the percentage of donation (i.e., donation divided by the funding goal) for the project i on a given day t. $RewardPledge_{\mathrm{it}}$



支持记录

订单序号 (Record No.)	支持者	支持项	数量(872) (Quantity)	支持时间
162	用户_PHFYO5	¥53	38	2018-12-07 23:42:38
161	申龙彪_1	¥1 无私支持	1	2018-12-07 10:47:51
160	用户_PJBDG9	¥53	50	2018-12-07 06:44:50
159	用户_PIQV5F	¥53	1	2018-12-06 21:29:38
158	用户_PIJHS9	¥53	5	2018-12-06 21:22:59
157	用户_PJBGZ5	¥53 无私支持	1	2018-12-06 21:19:15
156	用户_PJBE8R	¥53	1	2018-12-06 20:25:51
155	用户_PJBDG9	¥53	50	2018-12-06 20:06:31
154	用户_OGO64W	¥53	2	2018-12-06 15:20:25
153	用户_PJ95P0	¥53	10	2018-12-06 10:51:04

上一页 1 2 3 4 5 ... 下一页

Figure 4.1. Project Contribution Records

denotes the percentage of money (i.e., reward pledge divided by the funding goal) pledged to the project i on day t.

To understand how fundraisers should set up their projects' narratives, we extract a qualitative metric, subjectivity (in the range of 0-1), from project title and introduction by using NLP techniques. $Title\,Subjectivity$ and $Intro\,Subjectivity$ represent the subjectivity score of the project title and introduction respectively. We provide more details on the construction of Title Subjectivity and Intro Subjectivity later.

In addition, we study the impact of fundraisers' contribution behavior by using $Fundraiser\,Amount_{\text{it}}$. $Fundraiser\,Amount_{\text{it}}$ refers to the percentage of money (i.e., fundraisers' contribution divided by the funding goal) that the fundraiser supports his/her own project i on day t. We construct this variable through the following steps. First, we collect fundraiser id for each project. Second, by comparing fundraiser id and backer id, we could identify those contributions made by the fundraiser. Third, we calculate Fundraiser Amount for each project each day. Finally, we include several project characteristics associated with contribution behavior such as $Remaining\,Budget$, $Total\,Backer$ and $Type\,Philanthropy$.

Remaining Budget refers to the percentage of budget outstanding (i.e., remaining budget divided by the funding goal) as of a given day. *Total Backer* means the cumulative number of backers, both donors and reward buyers, by the given day. *Type Philanthropy* is denoted with 1 if project i belongs to philanthropy category. The type of each project is determined by fundraisers based on project details.

Natural Language Processing (NLP) for Project Narratives

Because the narratives are considered as an important determinant of backer' decisions [118], we apply sentiment analysis to extract textual characteristics of the project description (i.e., project title and introduction). Sentiment analysis is an NLP technique that could extract narratives' sentiment in a quantitative manner and thus offer us additional insights into the content of both project title and introduction. The primary goal of the sentiment analysis in our study is to determine whether sentences in project narratives express subjective opinions or not. In particular, we conduct the sentiment analysis to classify the subjectivity of each project description. The subjective project description is generally associated with fundraisers' opinions, emotion or judgment whereas objective project description consists of factual project information. In terms of sentiment analysis algorithm, we performed subjectivity classification using a lexicon built in the TextBlob package. This analysis enables us to label each project description with one subjectivity score in the range of 0 to 1. When the subjectivity score equals to one, it refers that this narrative is purely subjective. For instance, if the subjectivity score is 0.75, it means that mostly it is a subjective opinion and not factual information.

4.4 Empirical Analysis

Following [119] we use demand estimation techniques (i.e., the random coefficient logit model) to quantify the economic influence and relative importance of project characteristics. Furthermore, we quantify the impact of peer's crowdfunding behaviors and fundraisers' strategic behaviors on contribution dynamics.

4.5 Empirical Results

We first find that the subjectivity score of project titles and introductions greatly affects crowdfunding behaviors of donors. In particular, we find that projects with more subjective narratives will repel potential backers for donation. One underlying reason is that, backers are more skeptical of subjective than objective narratives while evaluating projects [120]. Potential donors may perceive projects with subjective titles as less trustworthy because subjective narratives cannot convey much factual information about projects that is useful to potential backers. This result yields a significant managerial implication: fundraisers should be careful about their project narratives in the project setup phase such as providing more objective content (e.g., project title and introduction) to attract more donors and donation.

Interestingly, we find that fundraisers' contributions will increase backers' willingness to donate. One possible explanation is that fundraisers' behavior is a signal to potential backers. Specifically, if fundraisers are observed to make contributions to their own projects, potential donors may believe that those fundraisers are making great efforts to the project. At the same time, fundraisers' contribution has no significant impact on the reward pledging of subsequent backers. This may be because reward buyers pay more attention to the creditworthiness of fundraisers and the quality of future products/service, compared to donors [14]. The signal conveyed by fundraisers' contribution is relatively not that important to reward buyers.

We also find that lagged amount of donation/reward pledge has a significant, positive impact on the amount of donation/reward pledge in the current period, which is in support of the herding effect. As individuals observe others donating/pledging more money, the amount they are inclined to donate/pledge increases. Above that, empirical evidence of negative interaction between donating and reward pledging is found. That is, more prior donating behaviors could lead to less reward pledging behaviors and vice versa.

4.6 Conclusions

Crowdfunding is considered as one of the primary ways of raising capital for ventures. However, the potential interplay of donating, reward pledging and fundraisers' behavior in crowdfunding have not received sufficient investigation. By leveraging data from a leading Chinese reward-based crowdfunding platform, we first find that projects with more subjective titles and introductions will repel potential backers for donation. We also find that fundraisers' contribution will increase backers' intention to donate, but it has no significant impact on backers' intention to reward pledge. These findings add to the signaling literature by offering new empirical evidence on the importance of signaling in markets where agents face information asymmetry. We are among the first to empirically uncover the relationship between fundraisers' strategic behaviors and backers' crowdfunding behaviors.

Further, consistent with prior literature, we find the herding effect in backers' crowdfunding behaviors and we further complement the prior studies about the dynamics of backers' behaviors on solely reward-based or donation-based crowdfunding platforms by investigating the interplay between these two types of crowdfunding behaviors (i.e., donating and reward pledging). We contribute to the literature on peer influence in crowdfunding by empirically showing negative interaction between donating and reward pledging on a crowdfunding platform that allows both behaviors. Specifically, more prior donating behaviors could lead to fewer reward pledging behaviors and vice versa.

This study provides significant managerial implications for practitioners on crowdfunding platforms. First, our findings suggest that backers are greatly affected by the project narratives offered by fundraisers. Fundraisers may attract more backers by conveying more project information and describe their projects in a more objective way. Second, we suggest that fundraisers should devote to their own project explicitly as it might increase the other backers' motivation to donate. Third, our research points out a possible way for fundraisers who are interested in managing the herd. If fundraisers want to increase their chances of funding success, they should try to maximize the contribution they get in the early stage. For example, they could seek friends for help at first. Last but not least, for the reward-based crowdfunding platform managers, we suggest non-incorporation of donation feature as it exhibits a negative peer influence on the pledging itself. They might also consider taking backers' dynamic behaviors into account. For instance, platforms should allow backers to sort projects based on the total number of backers.

5. CONCLUSIONS

In this dissertation, I focus on analyzing the design elements in the machine-platform-crowd transformation. In terms of platform, I look at how UGC platform could leverage editor-generated content to manage user-generated content. Online review is a decisive factor in consumers' decision-making process. Therefore, review platforms have used multiple strategies to manage user reviews. For example, platforms like BestBuy utilize financial incentives to attract new reviewers. Websites like Amazon place the most helpful reviews on the top of the page to enhance users' review reading experience. However, there exists few efforts from platforms to manage users' review content writing behaviors. We are wondering whether there is a way to manage the content of user reviews. So, we collaborate with a leading review platform in Asia. This platform is similar to Yelp, except that it provides editorial reviews written by the platform. With the access to this unique and rich dataset, we examine whether and how editorial reviews can affect the following user reviews. We analyze this from two dimensions. First, we look at the impact of editorial reviews on some numerical features such as the total number of reviews. Second, we extract textual features from large-scale unstructured reviews by utilizing Natural Language Processing (NLP) techniques (e.g., Topic modeling and Sentiment analysis). We find that editorial reviews have an overall positive impact on user reviews. To be specific, after seeing editorial reviews, users tend to post more and longer reviews. Also, they will resemble editorial reviews regarding of the topics, sentiment, and readability. While it is difficult for review platforms to harness the natural social influence among users, I propose that platforms can utilize editorial reviews to manage the quantity and quality of following user reviews.

Speaking of machine, I look at how machine can be used to help human in the context of retailing. Traditionally, when it comes to shelf management, manufacturers sign contracts with the retailers and then send delegates to ensure the execution of the contracts. However, this conventional practice loses its effectiveness in the globalized market. The main reason is that there are more small and independent retailers in emerging markets like China and those retailers have very heterogeneous shelf conditions. The question is whether we can use any advanced technology such as AI to help shelf monitoring in such situations. By collaborating with Danone, we conducted a large-scale field experiment as well as a qualitative study (i.e., interviews) in China. We find that AI-powered shelf monitoring significantly improves product sales in general. We further reveal that the positive effect shall be attributed to independent retailers rather than chained retailers. On a broader spectrum, this finding suggests the better scalability of AI in coping with more heterogeneous objects. Based on these findings, we can provide valuable insights on whether and how to leverage AI to cope with possible business challenges. Also, we confirm the importance of shelf display with solid empirical evidence by analyzing shelf photos using deep learning (e.g., image recognition).

On the part of crowd, I look at how crowdfunding platform could utilize the interplay of backers' behaviors. Backers' multiple roles and fundraisers' strategic behaviors could lead to complex dynamics of contribution behaviors on crowdfunding platforms. In order to provide substantive implications for crowdfunding practitioners, we need to understand the dynamic contribution patterns. Therefore, we work with one crowdfunding platform in China. Different from other reward-based crowdfunding platforms, this one also allows backers to donate. Using the random-coefficients logit model (i.e., BLP), we study the dynamics of donating, reward pledging and their interplays. We find a positive cross-channel influence between donation and reward pledge. Therefore, for the reward-based crowdfunding platform managers, we suggest incorporating the donation feature. Also, fundraisers' behaviors can affect the contribution dynamics. In order to give fundraisers some guides on setting a successful project, I utilize NLP techniques to analyze detailed project description. We show that projects described by more subjective content (i.e., title and introduction) significantly repel potential donors. We further show that fundraisers' contribution to their own projects might increase donor' intention to donate. So, for the donation-based crowdfunding platform managers, we suggest incorporation of a feature that can reveal fundraisers' great effort to the project as it exhibits a positive impact on donating.

REFERENCES

[1] C. Dellarocas, "The digitization of word of mouth: Promise and challenges of online feedback mechanisms," *Management science*, vol. 49, no. 10, pp. 1407–1424, 2003.

[2] W. Khern-am-nuai, K. Kannan, and H. Ghasemkhani, "Extrinsic versus intrinsic rewards for contributing reviews in an online platform," *Information Systems Research*, vol. 29, no. 4, pp. 871–892, 2018.

[3] G. Burtch, Y. Hong, R. Bapna, and V. Griskevicius, "Stimulating online reviews by combining financial incentives and social norms," *Management Science*, vol. 64, no. 5, pp. 2065–2082, 2017.

[4] P. B. Goes, M. Lin, and C. M. Au Yeung, "Popularity effect in user-generated content: Evidence from online product reviews," *Information Systems Research*, vol. 25, no. 2, pp. 222–238, 2014.

[5] Y. J. Lee, K. Hosanagar, and Y. Tan, "Do i follow my friends or the crowd? information cascades in online movie ratings," *Management Science*, vol. 61, no. 9, pp. 2241–2258, 2015.

[6] W. W. Moe and D. A. Schweidel, "Online product opinions: Incidence, evaluation, and evolution," *Marketing Science*, vol. 31, no. 3, pp. 372–386, 2012.

[7] W. W. Moe and M. Trusov, "The value of social dynamics in online product ratings forums," *Journal of Marketing Research*, vol. 48, no. 3, pp. 444–456, 2011.

[8] A. E. Schlosser, "Posting versus lurking: Communicating in a multiple audience context," *Journal of Consumer Research*, vol. 32, no. 2, pp. 260–265, 2005.

[9] K. Y. Goh, C. S. Heng, and Z. Lin, "Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content," *Information Systems Research*, vol. 24, no. 1, pp. 88–107, 2013.

[10] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of marketing research*, vol. 43, no. 3, pp. 345–354, 2006.

[11] X. M. Zhang and F. Zhu, "Group size and incentives to contribute: A natural experiment at chinese wikipedia," *American Economic Review*, vol. 101, no. 4, pp. 1601–15, 2011.

[12] S. Ransbotham, N. H. Lurie, and H. Liu, "Creation and consumption of mobile word of mouth: How are mobile reviews different?" *Marketing Science*, vol. 38, no. 5, pp. 773–792, 2019.

[13] G. Burtch, Q. He, Y. Hong, and D. D. Lee, "Peer awards increase user content generation but reduce content novelty," Working Paper. Available at SSRN 3465879., 2019.

[14] J. Zhang and P. Liu, "Rational herding in microloan markets," *Management science*, vol. 58, no. 5, pp. 892–912, 2012.

[15] H. Cai, Y. Chen, and H. Fang, "Observational learning: Evidence from a randomized natural field experiment," *American Economic Review*, vol. 99, no. 3, pp. 864–82, 2009.

[16] R. Cui, D. J. Zhang, and A. Bassamboo, "Learning from inventory availability information: Evidence from field experiments on amazon," *Management Science*, vol. 65, no. 3, pp. 1216–1235, 2019.

[17] C. Marsh, "Back on the bandwagon: The effect of opinion polls on public opinion," *British Journal of Political Science*, vol. 15, no. 1, pp. 51–74, 1985.

[18] K. Ozpolat, G. Gao, W. Jank, and S. Viswanathan, "Research note—the value of third-party assurance seals in online retailing: An empirical investigation," *Information Systems Research*, vol. 24, no. 4, pp. 1100–1111, 2013.

[19] G. Z. Jin and P. Leslie, "The effect of information on product quality: Evidence from restaurant hygiene grade cards," *The Quarterly Journal of Economics*, vol. 118, no. 2, pp. 409–451, 2003.

[20] J. Tirole, *The theory of industrial organization*. MIT press, 1988.

[21] Y. Chen, Y. Liu, and J. Zhang, "When do third-party product reviews affect firm value and what can firms do? the case of media critics and professional movie reviews," *Journal of Marketing*, vol. 76, no. 2, pp. 116–134, 2012.

[22] J. Eliashberg and S. M. Shugan, "Film critics: Influencers or predictors?" *The Journal of Marketing*, vol. 61, no. 2, pp. 68–78, 1997.

[23] S. Basuroy, S. Chatterjee, and S. A. Ravid, "How critical are critical reviews? the box office effects of film critics, star power, and budgets," *Journal of Marketing*, vol. 67, no. 4, pp. 103–117, 2003.

[24] S. K. Reddy, V. Swaminathan, and C. M. Motley, "Exploring the determinants of broadway show success," *Journal of Marketing Research*, vol. 35, no. 3, pp. 370–383, 1998.

[25] Y. J. Lee and Y. Tan, "Effects of different types of free trials and ratings in sampling of consumer software: An empirical study," *Journal of Management Information Systems*, vol. 30, no. 3, pp. 213–246, 2013.

[26] G. Shaffer and F. Zettelmeyer, "When good news about your rival is good for you: The effect of third-party information on the division of channel profits," *Marketing Science*, vol. 21, no. 3, pp. 273–293, 2002.

[27] Y. Chen and J. Xie, "Third-party product review and firm marketing strategy," *Marketing Science*, vol. 24, no. 2, pp. 218–240, 2005.

[28] D. Yin, S. Mitra, and H. Zhang, "Research note—when do consumers value positive vs. negative reviews? an empirical investigation of confirmation bias in online word of mouth," *Information Systems Research*, vol. 27, no. 1, pp. 131–144, 2016.

[29] E. Yazdani, S. Gopinath, and S. Carson, "Preaching to the choir: The chasm between top-ranked reviewers, mainstream customers, and product sales," *Marketing Science*, vol. 37, no. 5, pp. 838–851, 2018.

[30] D. Mayzlin, D. Yaniv, and C. Judith, "Promotional reviews: An empirical investigation of online review manipulation," *American Economic Review*, vol. 104, no. 8, pp. 2421–55, 2014.

[31] T. Lappas, G. Sabnis, and G. Valkanas, "The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry," *Information Systems Research*, vol. 27, no. 4, pp. 940–961, 2016.

[32] F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics," *Journal of marketing*, vol. 74, no. 2, pp. 133–148, 2010.

[33] M. Ma and R. Agarwal, "Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities," *Information systems research*, vol. 18, no. 1, pp. 42–67, 2007.

[34] P. J. Bateman, P. H. Gray, and B. S. Butler, "Research note—the impact of community commitment on participation in online communities," *Information Systems Research*, vol. 22, no. 4, pp. 841–854, 2011.

[35] L. Cabral and L. Li, "A dollar for your thoughts: Feedback-conditional rebates on ebay," *Management Science*, vol. 61, no. 9, pp. 2052–2063, 2015.

[36] Y. Chen, F. M. Harper, J. Konstan, and S. X. Li, "Social comparisons and contributions to online communities: A field experiment on movielens," *American Economic Review*, vol. 100, no. 4, pp. 1358–98, 2010.

[37] N. Huang, G. Burtch, B. Gu, Y. Hong, C. Liang, K. Wang, and B. Yang, "Motivating user-generated content with performance feedback: Evidence from randomized field experiments," *Management Science*, vol. 65, no. 1, pp. 327–345, 2019.

[38] X. Zeng and L. Wei, "Social ties and user content generation: Evidence from flickr," *Information Systems Research*, vol. 24, no. 1, pp. 71–87, 2013.

[39] Y. Yu, W. Khern-am-nuai, and A. Pinsonneault, "When paying for reviews pays off: The case of performance-contingent monetary rewards," *MIS Quarterly* (Forthcoming), 2020.

[40] N. Hu, P. A. Pavlou, and J. Zhang, "Why do product reviews have a j-shaped distribution?" *Communications of the ACM*, vol. 52, no. 10, pp. 144–147, 2007.

[41] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE Trans. Knowledge Data Engineer.*, vol. 23, no. 10, pp. 1498–1512, 2011.

[42] N. Archak, A. Ghose, and P. G. Ipeirotis, "Deriving the pricing power of product features by mining consumer reviews," *Management science*, vol. 57, no. 8, pp. 1485–1509, 2011.

[43] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *J. Machine Learn.Res.*, vol. 3, pp. 993–1022, 2003.

[44] J. Cao, T. Xia, J. Li, Y. Zhang, and S. Tang, "A density-based method for adaptive lda model selection," *Neurocomputing*, vol. 72, no. 7-9, pp. 1775–1781, 2009.

[45] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National academy of Sciences*, vol. 101, pp. 5228–5235, 2004.

[46] W. Zhao, J. J. Chen, R. Perkins, Z. Liu, W. Ge, Y. Ding, and W. Zou, "A heuristic approach to determine an appropriate number of topics in topic modeling," *BMC bioinformatics*, vol. 16, no. 13, p. 58, 2015.

[47] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, pp. 1–135, 2008.

[48] C. J. Hutto and E. Gilbert, “Vader: A parsimonious rule-based model for sentiment analysis of social media text,” in *Eighth international AAAI conference on weblogs and social media*, 2014.

[49] J. Chan and A. Ghose, “Internet’s dirty secret: Assessing the impact of online intermediaries on hiv transmission,” *Mis Quarterly*, vol. 38, no. 4, pp. 955–976, 2013.

[50] D. Qiao, S.-Y. Lee, A. Whinston, and Q. Wei, “Incentive provision and pro-social behaviors,” in *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

[51] K. Xu, J. Chan, A. Ghose, and S. P. Han, “Battle of the channels: The impact of tablets on digital commerce,” *Management Science*, vol. 63, no. 5, pp. 1469–1492, 2016.

[52] M. Bertrand, E. Duflo, and S. Mullainathan, “How much should we trust differences-in-differences estimates?” *The Quarterly journal of economics*, vol. 119, no. 1, pp. 249–275, 2004.

[53] D. H. Autor, “Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing,” *Journal of labor economics*, vol. 21, no. 1, pp. 1–42, 2003.

[54] R. A. Moffitt *et al.*, “Policy interventions, low-level equilibria, and social interactions,” *Social dynamics*, vol. 4, no. 45-82, pp. 6–17, 2001.

[55] M. Geetha, P. Singha, and S. Sinha, “Relationship between customer sentiment and online customer ratings for hotels-an empirical analysis,” *Tourism Management*, vol. 61, pp. 43–54, 2017.

[56] K. Hosanagar, D. Fleder, D. Lee, and A. Buja, “Will the global village fracture into tribes? recommender systems and their effects on consumer fragmentation,” *Management Science*, vol. 60, no. 4, pp. 805–823, 2014.

[57] C. F. Manski, “Identification of endogenous social effects: The reflection problem,” *The review of economic studies*, vol. 60, no. 3, pp. 531–542, 1993.

[58] H. S. Nair, P. Manchanda, and T. Bhatia, “Asymmetric social interactions in physician prescription behavior: The role of opinion leaders,” *Journal of Marketing Research*, vol. 47, no. 5, pp. 883–895, 2010.

[59] J. Berger, T. S. Alan, and J. R. Scott, “Positive effects of negative publicity: When negative reviews increase sales,” *Marketing Science*, vol. 29, no. 5, pp. 815–27, 2010.

[60] Y. Jiang, Y.-C. Ho, X. Yan, and Y. Tan, "When online lending meets real estate: Examining investment decisions in lending-based real estate crowdfunding," *Information Systems Research*, vol. 31, no. 3, pp. 715–730, 2020.

[61] L. Diestre, B. Barber IV, and J. Santaló, "The friday effect: Firm lobbying, the timing of drug safety alerts, and drug side effects," *Management Science*, vol. 66, no. 8, pp. 3677–3698, 2020.

[62] M. Iansiti and K. R. Lakhani, "Competing in the age of ai: Strategy and leadership when algorithms and networks run the world.," *Harvard Business Press*, 2020.

[63] C. C. Bennett and K. Hauser, "Artificial intelligence framework for simulating clinical decision-making: A markov decision process approach," *Artificial intelligence in medicine*, vol. 57, no. 1, pp. 9–19, 2013.

[64] E. Brynjolfsson and T. Mitchell, "What can machine learning do? workforce implications," *Science*, vol. 358, no. 6370, pp. 1530–1534, 2017.

[65] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

[66] P. J. Ferraro, J. J. Miranda, and M. K. Price, "The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment," *American Economic Review*, vol. 101, no. 3, pp. 318–22, 2011.

[67] M. H. Huang and R. T. Rust, "Artificial intelligence in service," *Journal of Service Research*, vol. 21, no. 2, pp. 155–172, 2018.

[68] X. Luo, S. Tong, Z. Fang, and Z. Qu, "Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases," *Marketing Science*, vol. 38, no. 6, pp. 937–947, 2019.

[69] X. Luo, M. S. Qin, Z. Fang, and Z. Qu, "Artificial intelligence coaches for sales agents: Caveats and solutions," *Journal of Marketing*, vol. 85, no. 2, pp. 14–32, 2021.

[70] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv preprint*, vol. arXiv:1606, no. 05718, 2016.

[71] R. E. Frank and W. F. Massy, "Shelf position and space effects on sales," *Journal of Marketing Research*, vol. 7, no. 1, pp. 59–66, 1970.

[72] X. Dreze, S. J. Hoch, and M. E. Purk, "Shelf management and space elasticity," *Journal of retailing*, vol. 70, no. 4, pp. 301–326, 1994.

[73] V. Nierop, D. E. Fok, and P. H. Franses, "Interaction between shelf layout and marketing effectiveness and its impact on optimizing shelf arrangements," *Marketing Science*, vol. 27, no. 6, pp. 1065–1082, 2008.

[74] B. Copple, "Shelf-determination," *Forbes*, 2002. DOI: https://www.forbes.com/forbes/2002/0415/130.html#24201d595309.

[75] M. Gómez and N. Rubio, "Shelf management of store brands: Analysis of manufacturers' perceptions," *International Journal of Retail and Distribution Management*, vol. 36, no. 1, pp. 50–70, 2008.

[76] B. R. Staats, H. Dai, D. Hofmann, and K. L. Milkman, "Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare," *Management Science*, vol. 63, no. 5, pp. 1563–1585, 2016.

[77] C. Sheu, H. Rebecca Yen, and B. Chae, "Determinants of supplier-retailer collaboration: Evidence from an international study," *International Journal of Operations and Production Management*, vol. 26, no. 1, pp. 24–49, 2006.

[78] A. Brinkhoff and J. Großpietsch, "Secrets to successful manufacturer-retailer collaboration," *Mckinsey*, 2014. DOI: https://www.mckinsey.com/industries/retail/our-insights/secrets-to-successful-manufacturer-retailer-collaboration#.

[79] D. J. Bowersox and P. J. Daugherty, "Logistics paradigms: The impact of information technology," *Journal of Business logistics*, vol. 16, no. 1, p. 65, 1995.

[80] R. E. Sabath, C. W. Autry, and P. J. Daugherty, "Automatic replenishment programs: The impact of organizational structure," *Journal of Business Logistics*, vol. 22, no. 1, pp. 91–105, 2001.

[81] W. Oh and H. C. Lucas Jr, "Information technology and pricing decisions: Price adjustments in online computer markets," *Mis Quarterly*, vol. 30, no. 3, pp. 755–775, 2006.

[82] S. A. Leachman and G. Merlino, "Medicine: The final frontier in cancer diagnosis," *Nature*, vol. 542, no. 7639, p. 36, 2017.

[83] A. Ranganathan and A. Benson, "A stitch in time: Work complexity and the divergent effects of employee monitoring on productivity," *Working Paper*, 2017.

[84] T. N. Hubbard, "The demand for monitoring technologies: The case of trucking," *The Quarterly Journal of Economics*, vol. 115, no. 2, pp. 533–560, 2000.

[85] D. S. Nagin, J. B. Rebitzer, S. Sanders, and L. J. Taylor, “Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment,” *American Economic Review*, vol. 92, no. 4, pp. 850–873, 2002.

[86] E. Duflo, R. Hanna, and S. P. Ryan, “Incentives work: Getting teachers to come to school,” *American Economic Review*, vol. 120, no. 4, pp. 1241–78, 2012.

[87] L. Pierce, D. C. Snow, and A. McAfee, “Cleaning house: The impact of information technology monitoring on employee theft and productivity,” *Management Science*, vol. 61, no. 10, pp. 2299–2319, 2015.

[88] M. Ibanez and M. Toffel, “How scheduling can bias quality assessment: Evidence from food-safety inspections,” *Management Science*, 2019.

[89] H. S. Heese, “Inventory record inaccuracy, double marginalization, and rfid adoption,” *Production and Operations Management*, vol. 16, no. 5, pp. 542–553, 2007.

[90] E. Frey and T. Rogers, “Persistence: How treatment effects persist after interventions stop,” *Policy Insights from the Behavioral and Brain Sciences*, vol. 1, no. 1, pp. 172–179, 2014.

[91] G. Charness and U. Gneezy, “Incentives to exercise,” *Econometrica*, vol. 77, no. 3, pp. 909–931, 2009.

[92] H. Allcott and T. Rogers, “The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation,” *American Economic Review*, vol. 104, no. 10, pp. 3003–37, 2014.

[93] U. Gneezy and J. A. List, “Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments,” *Econometrica*, vol. 74, no. 5, pp. 1365–1384, 2006.

[94] E. E. Anderson and H. N. Amato, “A mathematical model for simultaneously determining the optimal brand-collection and display-area allocation,” *Operations Research*, vol. 22, no. 1, pp. 13–21, 1974.

[95] P. Hansen and H. Heinsbroek, “Product selection and space allocation in supermarkets,” *European journal of operational research*, vol. 3, no. 6, pp. 474–484, 1979.

[96] M. Corstjens and P. Doyle, “A model for optimizing retail space allocations,” *Management Science*, vol. 27, no. 7, pp. 822–833, 1981.

[97] A. Bultez and P. Naert, “Sh. arp: Shelf allocation for retailers’ profit,” *Marketing science*, vol. 7, no. 3, pp. 211–231, 1988.

[98] A. Lim, B. Rodrigues, and X. Zhang, "Metaheuristics with local search techniques for retail shelf-space optimization," *Management Science*, vol. 50, no. 1, pp. 117–131, 2004.

[99] N. Amrouche and G. Zaccour, "Shelf-space allocation of national and private brands," *European Journal of Operational Research*, vol. 180, no. 2, pp. 648–663, 2007.

[100] C. W. Kuo and S. J. S. Yang, "The role of store brand positioning for appropriating supply chain profit under shelf space allocation," *European Journal of Operational Research*, vol. 231, no. 1, pp. 88–97, 2013.

[101] H. H. C. Chuang, R. Oliva, and S. Liu, "On-shelf availability, retail performance, and external audits: A field experiment," *Production and Operations Management*, vol. 25, no. 5, pp. 935–951, 2016.

[102] A. J. Mersereau, "Information-sensitive replenishment when inventory records are inaccurate," *Production and Operations Management*, vol. 22, no. 4, pp. 792–810, 2013.

[103] K. K. Cox, "The effect of shelf space upon sales of branded products," *Journal of Marketing Research*, vol. 7, no. 1, pp. 55–58, 1970.

[104] R. C. Curhan, "The relationship between shelf space and unit sales in supermarkets," *Journal of Marketing Research*, vol. 9, no. 4, pp. 406–412, 1972.

[105] T. Dunning, "Improving causal inference: Strengths and limitations of natural experiments," *Political Research Quarterly*, vol. 61, no. 2, pp. 282–293, 2008.

[106] V. Venkatesh, S. A. Brown, and H. Bala, "Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems," *MIS quarterly*, vol. 37, no. 1, pp. 21–54, 2013.

[107] C. X. Ou, P. A. Pavlou, and R. Davison, "Swift guanxi in online marketplaces: The role of computer-mediated communication technologies," *MIS quarterly*, vol. 38, no. 1, pp. 209–230, 2014.

[108] J. Chan and A. Ghose, "Internet's dirty secret: Assessing the impact of online intermediaries on hiv transmission," *MIS Quarterly*, vol. 38, no. 4, pp. 955–976, 2013.

[109] P. Rosenbaum and D. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, pp. 41–55, 1983.

[110] E. Duflo, R. Glennerster, and M. Kremer, "Using randomization in development economics research: A toolkit," *Handbook of development economics*, vol. 4, pp. 3895–3962, 2007.

[111] J. McKendrick, "Artificial intelligence will replace tasks, not jobs," *Forbes*, vol. 1, no. 1, pp. 1–3, 2018. DOI: https://www.forbes.com/sites/joemckendrick/2018/08/14/artificial-intelligence-will-replace-tasks-not-jobs/#1a79262ea7fa.

[112] P. Azoulay, T. Stuart, and Y. Wang, "Matthew: Effect or fable?" *Management Science*, vol. 60, no. 1, pp. 92–109, 2014.

[113] G. Burtch, S. Carnahan, and B. N. Greenwood, "Can you gig it? an empirical examination of the gig economy and entrepreneurial activity," *Management Science*, 2018.

[114] G. Burtch, A. Ghose, and S. Wattal, "An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets," *Information Systems Research*, vol. 24, no. 3, pp. 499–519, 2013.

[115] G. Burtch, A. Ghose, and S. Wattal, "Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding," *Information Systems Research*, vol. 27, no. 3, pp. 478–496, 2016.

[116] H. Dai and D. Zhang, "Prosocial goal pursuit in crowdfunding: Evidence from kickstarter.com," *Journal of Marketing Research*, vol. 56, no. 3, pp. 498–517, 2018.

[117] M. Lin, P. R. N., and V. S., "Judging borrowers by the company they keep: Social networks and adverse selection in online peer-to-peer lending," *Management Science*, vol. 59, no. 1, pp. 17–35, 2009.

[118] M. Herzenstein, S. Sonenshein, and U. M. Dholakia, "Tell me a good story and i may lend you money: The role of narratives in peer-to-peer lending decisions," *Journal of Marketing Research*, vol. 48, no. SPL, S138–S149, 2011.

[119] S. Berry, J. Levinsohn, and A. Pakes, "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 841–890, 1995.

[120] S. M. Mudambi and D. Schuff, "What makes a helpful online review? a study of customer reviews on amazon.com," *MIS Quarterly*, vol. 34, no. 1, pp. 185–200, 2010.

[121] S. Xiao, X. Tan, M. Dong, and J. Qi, "How to design your project in the online crowdfunding market? evidence from kickstarter," *Working Paper*, 2014.

[122] X. Tan, Y. Lu, and Y. Tan, "Why should i donate? examining the effects of reputation, peer influence, and popularity on charitable giving over social media platforms," *Working Paper*, 2016.

[123] R. Koning and J. Model, "Experimental study of crowdfunding cascades: When nothing is better than something," *Academy of Management Proceedings*, vol. 2014, no. 1, p. 16 683, 2014.

[124] J. Shang and R. Croson, "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods," *The Economic Journal*, vol. 119, no. 540, pp. 1422–1439, 2009.

[125] M. Xia, Y. Huang, W. Duan, and A. B. Whinston, "Research note—to continue sharing or not to continue sharing? an empirical analysis of user decision in peer-to-peer sharing networks," *Information Systems Research*, vol. 23, no. 1, pp. 247–259, 2012.

[126] M. Spence, "Signaling in retrospect and the informational structure of markets," *American Economic Review*, vol. 92, no. 3, pp. 434–459, 2002.