Investigating the Impact of Digital Transformation

Yu Kan

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:
Yong Tan, Chair
Leela Nageswaran
Mingwen Yang

Program Authorized to Offer Degree:
School of Business

University of Washington

Abstract

Investigating the Impact of Digital Transformation

Yu Kan

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

Digital transformation is fundamentally changing the way companies operate and engage with their customers, leading to significant shifts in various aspects of society. This thesis offers a thorough analysis of the multifaceted impact of digital transformation on businesses and society by focusing on three distinct yet interconnected perspectives: the role of artificial intelligence (AI) in customer care, the influence of user-generated content (UGC) on informational accessibility and inclusivity, and the effect of crowdsourcing on operational efficiency. First, I study how the identity cues of AI agents influence customers' strategic expression of emotions and satisfaction in customer care interactions. By emphasizing the significance of transparency and the potential challenges of non-disclosure in AI-mediated customer care, this research contributes to the growing literature on the social and behavioral implications of AI in business settings. Second, the thesis examines the underrepresentation of plus-sized users in online reviews and its consequences for informational asymmetry and purchasing behavior. By highlighting the role of privacy concerns arising from societal stigma and the need for platforms to actively promote inclusivity, this research sheds light on the crucial issue of informational accessibility and inclusivity in the digital era. Third, the thesis explores how leveraging crowdvoting in assortment planning can enhance a subscription-based platform's operational performance and user engagement. By demonstrating the transformative potential of crowdsourcing in retail operations, this research contributes to the literature on

the impact of digital transformation on business processes and customer engagement. By synthesizing insights from these three perspectives, this thesis contributes to the complex and multifaceted impact of digital transformation on businesses and society. It underscores the importance of balancing technological innovation with social responsibility, efficiency with equity, and progress with inclusivity. Ultimately, this thesis serves as a foundation for future research on the evolving impact of digital transformation, emphasizing the need for ongoing, interdisciplinary investigations that keep pace with the rapid advancements in technology and their far-reaching implications for business and society.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Foremost, I want to express my deepest gratitude to my advisor, Professor Yong Tan, who has been my role model. His careful guidance throughout my academic journey has been invaluable. Under his mentorship, I have not only learned to be humble and enthusiastic about academia but also how to lead a meaningful and fulfilling life.

Second, I want to extend my sincere thanks to my coauthor and committee member, Professor Uttara Ananthakrishnan. I will never forget her generosity in teaching me skills and sharing her valuable insights. The research opportunities she provided me with have been truly invaluable.

Third, I also want to express my gratitude to my coauthor and committee member, Professor Leela Nageswaran. She taught me one very important lesson: 'rigorous,' which I continue to strive for in my studies.

I would also like to express my sincere gratitude to my other committee members, Professor Mingwen Yang and Professor Jing Tao. Their constructive suggestions have helped me refine my research study to meet a higher quality standard, and I truly appreciate their contributions. I would like to extend my gratitude to Shawna Reimers, Beau Kirkeby, Nuzulita Budhiari, Shannon Goodwin, and other staff at the Foster School of Business for their generous support during my studies.

Lastly, I would like to thank my parents, Minggang Kan and Chaohua Wang, and my in-laws, Zongyi Zhang and Chunhong Li, for their unconditional love and support. Knowing that there is always a home behind me has given me the strength to reach this far. I also want to thank my classmate, friend, and husband, Mingrui Zhang. We have both discovered and will continue to discover many business stories, but you are my great love story.

Chapter 1

INTRODUCTION

The rapid advancement and widespread adoption of digital technologies are fundamentally reshaping the business landscape and society at large, ushering in an era of unprecedented change and innovation. At the forefront of this digital transformation are cutting-edge technologies such as artificial intelligence (AI), user-generated content (UGC) platforms, and crowdsourcing. These technologies offer immense potential for enhancing efficiency, personalization, and value creation, promising to revolutionize the way businesses operate and interact with customers. However, they also introduce new complexities and considerations that warrant careful examination, as the impact of digital transformation extends far beyond the realm of technology, encompassing social, ethical, and behavioral dimensions that shape its outcomes.

This thesis aims to investigate the impact of digital transformation from three distinct yet interconnected perspectives, shedding light on the challenges and opportunities that arise as businesses and society navigate the digital age. By examining the role of AI in customer care, the influence of UGC on informational accessibility and inclusivity, and the effect of crowdsourcing on operational efficiency, this thesis seeks to provide a comprehensive understanding of the multifaceted nature of digital transformation and its implications for businesses and society.

The first perspective focuses on the transformative role of AI in customer care. As organizations increasingly turn to AI-powered agents for handling customer interactions, understanding the dynamics between AI and human agents becomes crucial. In Chapter 3, I focus on how different identity cues associated with AI agents influence customers' strategic expression of emotions and ultimate satisfaction. By examining the interplay between AI

identity cues, customer behavior, and satisfaction, this chapter highlights the importance of transparency and the potential challenges of non-disclosure in AI-mediated customer care. It underscores the need for responsible AI deployment in customer-facing roles, emphasizing the importance of striking a balance between efficiency and trust in AI-human interactions. This chapter contributes to the growing body of research on the social and behavioral implications of AI, providing valuable insights for businesses seeking to leverage AI in customer care while maintaining trust and transparency.

The second perspective examines the impact of UGC on informational accessibility and inclusivity. Chapter 4 addresses the issue of underrepresentation of plus-sized users in online reviews and its consequences for informational asymmetry. As the fashion industry moves towards greater size inclusivity, the lack of representative information from plus-sized users can hinder informed decision-making and perpetuate exclusionary practices. This chapter explores the mechanisms behind this disparity, focusing on the role of privacy concerns stemming from societal stigma. By shedding light on the factors that contribute to the underrepresentation of marginalized groups on UGC platforms, this chapter highlights the need for platforms to actively foster inclusivity and mitigate informational asymmetry. It underscores the importance of designing inclusive features and policies that encourage participation from diverse user groups, thereby promoting equity and accessibility in the digital realm. This chapter contributes to the discourse on digital inclusion, emphasizing the critical role of UGC platforms in shaping the informational landscape and the need for proactive measures to ensure that the benefits of digital transformation are distributed equitably.

The third perspective investigates the transformative potential of crowdsourcing in retail operations. Chapter 5 examines how adopting crowdvoting impacts a subscription-based platform's operational performance and user engagement. As e-commerce retailers face high uncertainty when making inventory decisions for new products, crowdvoting has emerged as a strategy to incorporate customer preferences and reduce demand uncertainty. By demonstrating how leveraging the wisdom of the crowd can improve efficiency and customer satisfaction in digital retail environments, this chapter highlights the potential of crowdsourcing

to transform traditional business processes. It underscores the importance of engaging customers in the value creation process and harnessing collective intelligence to drive innovation and operational excellence. This chapter contributes to the literature on retail operations and customer engagement, providing valuable insights for businesses seeking to harness the power of crowdsourcing in the digital age and optimize their assortment planning strategies.

By combining insights from these three perspectives, this thesis aims to provide a comprehensive understanding of the impact of digital transformation on businesses and society. It highlights the complex interplay between technological innovations, such as AI and crowdsourcing, and the human factors that shape their outcomes, such as emotional expression, informational inclusivity, and user engagement. By examining the challenges and opportunities presented by digital transformation through these distinct lenses, this thesis contributes to the ongoing discourse on the role of technology in shaping the future of business and society.

The insights gleaned from this thesis have far-reaching implications for businesses, policymakers, and researchers alike. They underscore the importance of approaching digital transformation with a holistic perspective that considers not only the technological aspects but also the social, ethical, and behavioral dimensions. As businesses continue to embrace AI, UGC platforms, and crowdsourcing, it is crucial to develop strategies that prioritize transparency, inclusivity, and customer-centricity alongside efficiency and innovation. Policymakers must grapple with the regulatory challenges posed by these technologies, ensuring that the benefits of digital transformation are distributed equitably and that the rights and privacy of individuals are protected. Researchers, in turn, must continue to investigate the evolving implications of digital transformation, adapting their research agendas to keep pace with the rapid changes in the technological landscape.

This thesis also highlights the importance of interdisciplinary research in understanding the impact of digital transformation. By drawing insights from fields such as information systems, marketing, operations management, and social sciences, this thesis demonstrates the value of a multidisciplinary approach in capturing the complex and multifaceted na-

ture of digital transformation. It underscores the need for cross-disciplinary collaboration and knowledge exchange in order to develop a more comprehensive understanding of the challenges and opportunities presented by the digital age.

As digital transformation continues to reshape the business landscape and society at large, it is imperative for businesses, policymakers, and researchers to stay attuned to its evolving implications. This thesis underscores the importance of proactive engagement with the challenges and opportunities presented by digital transformation, emphasizing the need for continuous learning, adaptation, and innovation. By fostering a culture of experimentation, collaboration, and customer-centricity, businesses can navigate the complexities of the digital age and harness the power of technology for competitive advantage. Policymakers, in turn, must strive to create a regulatory environment that promotes innovation while safeguarding the rights and interests of individuals and society as a whole. Researchers must continue to push the boundaries of knowledge, exploring new frontiers in digital transformation and contributing to the development of evidence-based strategies for businesses and policymakers.

Chapter 2

LITERATURE REVIEW

In this chapter, I review the existing literature related to the three contexts investigated in this dissertation. Section 2.1 begins by examining the literature on emotional AI in customer care, focusing on the emerging role of AI-powered agents in handling customer interactions and the implications of AI identity cues on customer behavior and satisfaction. I then review the theoretical foundations of strategic behavior and emotional misrepresentation in customer-agent interactions, as well as the empirical studies investigating the impact of agent identity on customer perceptions and outcomes.

Section 2.2 explores the literature on informational accessibility and inclusivity in online platforms, with a specific focus on user-generated content (UGC) and its role in shaping the informational landscape. I first summarize the theories on information asymmetry and its consequences for consumer decision-making, particularly in the context of online reviews. Then, I review the empirical studies examining the factors that contribute to the underrepresentation of marginalized groups, such as plus-sized users, in UGC and the implications for informational inclusivity. I also discuss the research gap in understanding the role of privacy concerns and social stigma in driving participation disparities on online platforms.

In Section 2.3, I delve into the literature on crowdsourcing and its applications in retail operations, focusing on the emerging practice of crowdvoting in assortment planning. I begin by reviewing the theoretical foundations of crowdsourcing, including the wisdom of crowds and collective intelligence, and their potential to transform traditional business processes. I then summarize the empirical studies investigating the impact of crowdsourcing on operational efficiency, demand forecasting, and customer engagement in various retail settings. I also highlight the research gap in understanding the mechanisms through which crowdvoting

influences assortment decisions and user behavior on subscription-based platforms.

By reviewing the relevant literature across these three contexts, this chapter aims to provide a comprehensive theoretical and empirical foundation for the essays presented in this dissertation. The insights gleaned from this review will inform the research questions, methodological approaches, and interpretations of the findings in each essay, contributing to a nuanced understanding of the multifaceted impact of digital transformation on businesses and society.

2.1 Emotional AI

2.1.1 Emotion AI in Customer Care

The use of AI in customer care is gaining increasing attention in both marketing practice and research. Huang and Rust (2023) propose an AI-enabled customer care journey that leverages the emerging capabilities of emotion AI, exemplified by interactive generative AI (GenAI) models, to recognize, understand, manage, and connect with customer emotions. They argue that emotion AI has the potential to provide emotionally intelligent customer care at scale, improving both customer emotional well-being and customer lifetime value.

Previous research has examined various aspects of AI applications in service settings. For example, Caruelle et al. (2022), Huang and Rust (2018), Kopalle et al. (2022), Liu-Thompkins et al. (2022), and Rust and Huang (2014) discuss how information technology can facilitate interactive communication between service providers and customers. Importantly, Huang and Rust (2018) propose that AI can possess mechanical, thinking, and feeling intelligences that roughly map to doing, thinking, and feeling tasks. In the context of customer care specifically, AI has traditionally played a limited, supportive role to human agents due to constraints in its ability to detect and respond to emotions (Huang and Rust 2018, Wilson and Daugherty 2018). However, the rapid progress in affective computing and the rise of GenAI models have opened up new possibilities for AI to handle emotionally complex customer interactions more effectively (Huang and Rust 2023).

The importance of emotions in customer care is well-established. For example, Berry (2000) highlights that relationship building relies on emotional connections, which give customers a feeling of belonging and being understood. Research by Rust et al. (2004) and Magids et al. (2015) demonstrates that customers with strong emotional ties to a company tend to be more loyal and valuable. Herhausen et al. (2023) underscore the importance of managing customer emotions for effective service recovery following failures. Despite this, there have been doubts about AI's ability to effectively handle customer emotions compared to human agents (Rust et al. 2021).

Huang and Rust (2023) proposed AI-enabled customer care journey provides a novel framing of how state-of-the-art emotion AI like GenAI can potentially bridge this gap. GenAI might detect customer emotions from various signals, display empathy, offer tailored emotion management suggestions, and ultimately foster emotional connections between customers and brands. In summary, while AI has traditionally played a limited role in customer care, the rapid development of emotion AI exemplified by GenAI is opening up new possibilities for AI to provide emotionally intelligent customer care at scale. Huang and Rust (2023)'s framework provides a foundation for future research to systematically examine the strengths, limitations, and implementation of emotion AI in customer care contexts to simultaneously enhance customer emotional well-being and firm performance. Further empirical work is needed to validate and build upon their propositions.

2.1.2 Emotional Misrepresentation in Negotiation

Within the context of after-sales customer care, negotiation has increasingly become a critical issue. Two primary stakeholders, namely the agents and the customers, engage in discussions with potentially conflicting objectives. While agents operate under company-imposed constraints concerning resource allocation, customers aim to maximize their own utility derived from these allocations. Consequently, to attain this objective, customers employ strategic approaches. Prior research has highlighted that some customers engage in deceptive behaviors during negotiations. For instance, a study by Murnighan et al. (1999) reveals that

over one-third of "experienced" negotiators participated in both active deception and passive misrepresentation in a single interaction. As noted by Fulmer et al. (2009), these tactical behaviors can be categorized into two main types: *informational* or *emotional* tactics of deception. Most research has predominantly centered on the former tactic, which entails the management and manipulation of information during the negotiation process (Weingart et al. 1990, Lewicki and Robinson 1998).

Researchers have increasingly centered on emotional misrepresentation, which Campagna et al. (2016) defines as "the deliberate expression of an emotion that is different from the one genuinely felt by the negotiator." As highlighted by Thompson et al. (1999), emotions play two different roles in the context of negotiations: firstly, there's the inherent emotion experienced internally by the negotiators. Secondly, there's the expressed emotion, which negotiators can strategically choose to display in their interactions, potentially influencing the dynamics of the negotiation. These two facets are interconnected and can coexist within an individual. The negotiator can simultaneously engage with their genuine emotions while also managing how they present these emotions to their counterpart. The deviation of expressed emotion from experienced emotion leads to emotional misrepresentation. A series of research indicates that expressing more intense emotions can lead to larger concessions from the recipient, thus encouraging behaviors of emotional misrepresentation. Van Kleef et al. (2004) show through lab experiments that participants yielded more to angry opponents than to happy or neutral emotional states. Van Dijk et al. (2008) and Wang et al. (2012) further testify that expressing anger in negotiations can lead to an apparent concession. They also propose such effectiveness is contingent on the bargaining situation and has hidden costs. As summarized by Hutson (2015), the repeated observation underscores the notion that employing emotional misrepresentation, like feigning anger, can aid a negotiator in securing better deals. In their paper, Campagna et al. (2016) argue that emotional misrepresentation can also backfire, triggering more negative effects from the counterparts. In the existing literature on emotional misrepresentation, the primary emphasis has been on the consequences of such behavior, with relatively fewer studies exploring the trigger of it.

One possible explanation for the limited number of research in this area is the inherent difficulty in capturing and measuring strategic emotional behavior. While the expressed level of emotion in user-generated text can be easily identified, discerning the underlying emotional state proves to be a more challenging task.

2.1.3 Social Identity Theory and Identity Ambiguity

Social Identity Theory (SIT) is first proposed by Tajfel (1974). The theory introduces essential concepts like social identity, ingroups (groups to which an individual feels they belong), and outgroups (those they perceive as different from themselves). It suggests that individuals derive part of their self-concept from the groups they belong to, leading to ingroup favoritism and outgroup discrimination. This categorization process is fundamental for developing a social identity.

As an extension of SIT, the Self-Categorization Theory by Turner et al. (1987) points out that it helps individuals define themselves in relation to others. While they do not explicitly discuss the negative effects of ambiguity in identity, it implies that less distinct group identities may weaken ingroup cohesion and outgroup differentiation. When social categories are broad or ambiguous, they may lack the clear boundaries that are necessary for individuals to identify strongly with a group. This ambiguity can blur the distinction between ingroups and outgroups, which is a key component of SIT. Abrams and Hogg (2006) further prove that unclear or overlapping group boundaries due to broad or ambiguous categories can lead to confusion and reduced group cohesion. Without clear categorization, the typical processes of in-group favoritism and outgroup discrimination, central to SIT, may become disrupted. Roccas and Brewer (2002) point out in their paper that ambiguity in social categorization can also lead to identity conflict and cognitive dissonance, as individuals struggle to navigate their place within complex social structures. Drawing upon the insights of SIT and social categorization theories Zuckerman (1999) also observe a key trend: As the complexity of an individual's identity increases, it becomes more challenging for outsiders to clearly understand his or her identity. This complexity can lead to what is termed as an "illegitimacy

discount." The concept of illegitimacy discount, as outlined by Zuckerman (1999), involves a decrease in attention and support from others when they find the volunteer's identity to be unclear or confusing. Therefore, in our context, non-disclosure agents have a higher ambiguity in their identity. Such broad identity cues may potentially lead to feelings of uncertainty and identity confusion among the audience. Hence, users may behave more negatively when interacting with non-disclosure agents compared to agents with clearer identities.

2.2 User-Generated Content

2.2.1 The Role of UGC in Reducing Information Asymmetry

Information asymmetry refers to a situation where one party in a transaction possesses more or superior information compared to the other. It originates from the relationship between a seller and a buyer, as discussed in the context of the used car market by (George et al. 1970), where the seller has more information about the car's condition. Similarly, in the labor market, a job seeker might have incentives to misrepresent their skills to an employer, who has limited information on the candidate's actual capabilities (Spence 1978, Stiglitz 1975). Information asymmetry can lead to adverse selection, where the party with less information risks selecting inferior products because the quality is misrepresented. Prior studies in this research area find a negative impact on market efficiency and equality due to information asymmetry in various contexts, spanning from financial markets (Aboody and Lev 2000, Armstrong et al. 2011, Reuer and Koza 2000, etc.) to health care (Rochaix 1989, Blomqvist and Léger 2005, Khurana et al. 2019, etc.).

In the context of online retail, buyers often cannot ascertain the quality of a product firsthand and thus suffer from information asymmetry. Online retailers have adopted several techniques over the last two decades to mitigate information asymmetry. This includes providing elaborate descriptions of the products, photos, and an extensive system of online reviews. UGC in the form of online reviews plays a major role in mitigating information asymmetry, enabling consumers to make more informed decisions based on the experiences

of others. There is a rich body of literature that has studied the economic value of online reviews (Chatterjee 2001, Chen and Xie 2005, Chevalier and Mayzlin 2006, Forman et al. 2008, Moe and Schweidel 2012, Gao et al. 2015, Hu et al. 2006, Archak et al. 2011, Zervas et al. 2021, Hu et al. 2009b). UGC not only helps consumers reduce information asymmetry but also aids firms in improving their product quality using information from the consumers (Ananthakrishnan et al. 2023).

Despite the importance and the ubiquitous presence of UGC, there is still significant information asymmetry in online retail. Studies in the IS literature show that online reviews suffer from acquisition bias (a consumer who buys a product is more likely to review the product) and underreporting bias (only consumers with extreme opinions are more likely to leave a review) leading to a "J-shaped" review curve (Hu et al. 2006, 2017, 2009a, Gao et al. 2015, Zervas et al. 2021).

Information asymmetry poses significant economic challenges both for online retailers and customers since most platforms allow customers to return products online [1]. Consumers purchasing apparel online often cite fit and sizing among their top reasons to return the products (Nestler et al. 2021). Returns have become a major issue for online apparel retailers. The National Retail Federation estimates that customers in the U.S. returned over 800 billion U.S. dollars worth of goods during 2022.[2] Surveys from Invesp, Forrester group and KPMG estimate the return rate to be between 30% and 50% of all online retail.[3]

Therefore, despite significant efforts from online retailers to incorporate UGC on their platforms, the reduction in information symmetry depends on the quantity and quality of reviews. This, in turn, is closely tied to customers' motivations to review products or services.

---

[1] https://www.wsj.com/articles/those-new-online-returns-fees-are-driving-away-shoppers-2d3aa5fb, last accessed April 25, 2024.

[2] https://nrf.com/media-center/press-releases/2022-retail-returns-rate-remains-flat-816-billion, last accessed April 25, 2024.

[3] https://www.invespcro.com/blog/ecommerce-product-return-rate-statistics/, https://www.theguardian.com/money/2022/jan/08/uk-surge-in-post-christmas-returns-reveals-dark-side-of-online-shopping-boom, last accessed April 25, 2024.

2.2.2 Motivations to contribute on UGC platforms

The motivations of individuals who choose to contribute to UGC on online platforms have been widely studied in the last decade (for a comprehensive literature review, please refer to Naab and Sehl (2017)). Users who write reviews share their experiences, both positive and negative, with other customers trying to evaluate the quality of a product before making purchase decisions. Individuals who contribute to UGC devote significant time and effort to the process. The motivation to contribute to UGC can be intrinsic or extrinsic. Intrinsic incentives are utilities related to creative expression, gaining attention, audience, or social recognition (Toubia and Stephen 2013, Zhang and Zhu 2011). Extrinsic incentives are factors that are stimulated by platform interventions, including monetary incentives (Burtch et al. 2018b, Khern-am nuai et al. 2018, Cabral and Li 2015) and non-monetary incentives (Burtch et al. 2018b, Goes et al. 2016, Gallus 2017, Huang et al. 2019). In the context where users are allowed to post images, users can further gain image-related utility from posting images (Toubia and Stephen 2013).

Intrinsic motivation can further be attributed to altruism (Hennig-Thurau et al. 2004, Zhang and Zhu 2011, Qiao et al. 2020). Altruistic incentives require personal resources to craft a review, recall personal experiences, and take the time to express thoughts. They benefit other potential consumers with essential information about specific products and caution others about items of subpar quality (Hennig-Thurau et al. 2004, Zhang and Zhu 2011, Qiao et al. 2020).

While altruism motivates users to contribute to online reviews for the benefit of others, self-enhancement (Wien and Olsen 2014) is a compelling internal drive that leads individuals to engage in behaviors aimed at improving their social image and self-worth. Unlike altruism, which focuses more on the collective benefit, self-enhancement is tied to personal decisions users make regarding what they choose to reveal about themselves. This can be prominently observed when users decide to include photos in their reviews. The act of sharing personal photos or images related to their experiences with the product serves as an expressive tool for

users to present themselves in a certain light, enhancing their online personas. In this way, users can strategically utilize their reviews to demonstrate their tastes, lifestyles, or expertise, seeking validation and recognition from the online community. This personal image curation and its ensuing rewards underscore the self-enhancement aspect of user motivations in online review participation.

The act of posting personal-related information may expose reviewers to potential privacy risks. Users might feel that their personal experiences, tastes, or even identities could be revealed, analyzed, and used without their consent (Bélanger and Crossler 2011, Lahlou 2008). In particular, sharing images might be considered as an explicit disclosure of personal information, as images can provide a wealth of details about the reviewer, including their appearance, surroundings, lifestyle, and more.

In our study, we take the dimensions of altruism, self-enhancement, and privacy concerns into account when examining users' motivation to participate in online review platforms. We consider the distinct differences among these three types of motivations that influence the users' decision to post reviews and photos. In doing so, we address an important managerial question on what platforms should do to incentivize users to contribute on review platforms, given significant privacy concerns.

2.2.3 Anonymity

Anonymity is defined as a state of nonidentifiability or "noncoordinatability" of traits. Anonymity hides identifiable characteristics of an individual, ensures privacy, and safeguards against unwanted identification (Wallace 1999). Anonymity is achieved when it becomes impossible to associate a known trait with other identifying characteristics. For instance, within an e-commerce context, a user who leaves a product review without disclosing their name retains anonymity, as the only known information is about their experience with the product or service. When names and/or profiles are linked to each review, anonymity is compromised as other consumers of the platform can tie the consumption of a specific product and the experience with a specific user.

Anonymity while contributing to online reviews is particularly salient when the disclosure of consumption could lead to some negative outcome, such as stigmatization or discrimination for the person who wrote the review. For example, non-anonymous reviews of healthcare providers could potentially disclose the medical conditions that the reviewers sought care for, including conditions that could lead to social stigma or workplace discrimination. This could lead to self-censoring among reviewers for specific services or products. For example, there could be significant self-censoring among users on non-anonymous review platforms such as Yelp or Google reviews when they review mental health providers or prenatal healthcare providers since this information could lead to workplace discrimination. Further, non-anonymity on review platforms could lead to unpleasant interactions with the service providers in the responses. Media reports suggest that some reviewers even get sued over negative reviews by the providers [4]

Online platforms, particularly UGC platforms, face privacy-related concerns, which can hinder user engagement. Self-censoring due to privacy concerns over social stigma or fear could lead to significant loss of information for other potential customers of the platform. For example, non-anonymity could lead to self-censoring only negative reviews, which could disproportionately bias the reviews towards more positive reviews (Huang et al. 2017).

However, providing anonymity on platforms also has significant negative consequences, especially in diminishing social presence and credibility (Burtch et al. 2015, 2016, Jiang et al. 2022). This highlights a significant academic and managerial gap in creating privacy policies on UGC platforms. On the one hand, providing anonymity could help users in increasing the credibility of UGC. On the other hand, non-anonymity could lead to self-censoring, particularly among users who routinely experience social stigma. This is the subject of our research.

---

[4] https://www.cnbc.com/2019/10/10/can-you-get-sued-over-a-negative-yelp-review.html, last accessed April 25, 2024.

2.2.4 Social Stigma

Finally, we contribute to the literature on social stigma and how stigma moderates users' behavior on UGC platforms. Goffman and Goffman (1963) defines stigma as an attribute that deeply discredits an individual's social identity. Building on this framework, Dudley (2000) characterizes stigma as the stereotypes or negative perceptions assigned to an individual or group when their traits or behaviors are perceived as different from social norms. Campbell and Deacon (2006) further outlines stigma into physical deformities, deviations in personal traits, and tribal affiliations, highlighting how these deviations from societal norms fuel stigma. Research indicates that stigma can have profound effects on individuals, including reduced access to resources (Clement et al. 2015), social isolation (Livingston and Boyd 2010), psychological distress (Corrigan and Watson 2002), and poor health outcomes (Hatzenbuehler et al. 2013).

The concept of body inequality primarily refers to the societal discrepancies and prejudices that exist about body size and shape. According to the categorization of stigma by Campbell and Deacon (2006), body inequality is classified as "overt or external deformities". Research has widely demonstrated that cultural and media norms have historically favored slim, "ideal" body shapes, particularly for women, and such standards can lead to negative body image, low self-esteem, and even health issues like eating disorders (Grabe et al. 2008, Spitzer et al. 1999). For instance, Grabe et al. (2008) illustrates how media exposure propagates thin-ideal internalization, body dissatisfaction, and disordered eating in women. Moreover, studies suggest that these body ideals can manifest in negative attitudes, stereotypes, and discrimination toward individuals who do not conform to these societal standards, particularly those who are overweight or obese (Pomeranz 2008, Brewis et al. 2011, Carels et al. 2018). Discrimination faced by plus-sized people has been well-studied. For example, O'Brien et al. (2013) demonstrate how anti-fat prejudice and affects the perception of leadership potential, starting salary, and overall employability. Plus-sized people are routinely stereotyped by educators (Neumark-Sztainer et al. 1999, Lynagh et al. 2015), at work and

during hiring (O'Brien et al. 2007, Flint et al. 2016, Roehling 1999, Sartore and Cunningham 2007, O'Brien et al. 2013), by health care providers (Sabin et al. 2012, Schwartz et al. 2003, Pausé 2014, Lee and Pausé 2016, Teachman and Brownell 2001) and in social settings (Bartels and Nordstrom 2013, Crandall 1995, Carels et al. 2018). This discrimination and stereotyping also extends in the digital space. Stigmatized individuals exhibit reluctance to engage in online communities due to fear of judgment or discrimination (Lin and Reid 2009, Puhl and Heuer 2009). Anecdotal evidence suggests that body-shaming plus-sized celebrities and influencers on social media platforms are routinely subjected to harassment and unpleasant comments about their weight on digital media platforms.[5]

Privacy concerns associated with UGC platforms are likely to be amplified among plus-sized users who face widespread social stigma in their everyday lives. Therefore, plus-sized users might choose to self-censor their opinions on clothes, lest they receive negative feedback about their appearance instead of the product they are reviewing. This decision could be even more salient in reviews accompanied by photos.

Our study is the first in the literature to study the interplay between privacy, social stigma, and the decision to participate on UGC platforms. We investigate whether and why there is an underrepresentation of plus-sized users in reviews and photos on online platforms from the perspective of privacy concerns and societal biases against fuller body types. Thus, we contribute to the emerging literature on the design of inclusive platforms that foster engagement and interactions from all users by promoting a safe and welcoming digital environment (Holland and Tiggemann 2016, Mitkina et al. 2022).

2.3 Crowdsourcing

In this section, we summarize the studies of crowdsourcing, where a firm outsources tasks to a large network of people to test the market prior to a new product launch (Araman

[5] https://www.businessinsider.com/remi-bader-influencer-health-body-shaming-response-2023-9, https://www.allure.com/story/callie-thorpe-response-body-shaming-nike-campaign, https://fortune.com/2015/08/27/plus-size-retailers/, last accessed April 25, 2024.

and Caldentey 2016). Papers in this field show that crowdsourcing generates new ideas (Terwiesch and Xu 2008, Ales et al. 2017, Hwang et al. 2019), facilitates learning through peer feedback (Huang et al. 2014) and from superstar competitors (Zhang et al. 2019), and serves as a source of demand information and funding (Roma et al. 2018). By combining Kickstarter crowdfunding data and expert opinions, Mollick and Nanda (2016) find that there is agreement between the funding decisions of crowds and experts. Using field experiments in the food and consumer electronics product categories, Nishikawa et al. (2017) find that labeling products as "customer-ideated" improved the product's performance by up to 20%, driven by the perception that such products address consumer needs more effectively. In a similar vein, Keppler et al. (2022) show that teacher-led crowdfunding projects improve schools' educational outcomes. Prior studies in the user-generated content literature have also demonstrated its ability to ameliorate uncertainty in product quality (Forman et al. 2008, Dellarocas et al. 2007) and improve the quality of the firm (Ananthakrishnan et al. 2023).

Our paper deals with a specific type of crowdsourcing, referred to as crowdvoting, where potential consumers vote on a product, and firms use the votes to make operational decisions. Within crowdvoting, Marinesi and Girotra (2013) study how offering a price discount for voters can be an effective method to estimate customers' valuations for a product. By explicitly modeling the duration of the voting period, Araman and Caldentey (2016) investigate the resulting tradeoff between the quality of the demand forecast obtained and the potential revenue loss from waiting for votes. Feng et al. (2022) explore how the display of products to voting users – either the entire menu of products or only two products at a time – affects a platform's ability to identify consumers' most preferred product. Different from these studies, we explore how the adoption of crowdvoting impacts a retailer's purchasing decisions and operational performance.

This work is also related to advance selling. This is a practice where a firm can make an advance "pre-order" decision and another ordering decision after observing demand information. The earliest work on this topic is Fisher and Raman (1996) who study the benefits of

a quick-response supply chain and early orders in estimating the demand distribution. Gallego and Özer (2001) obtain the optimal inventory policy in the presence of advance demand information, and they find that as more advance demand information becomes available, the average inventory level in their multi-period setting goes down. By considering pricing and strategic consumers, Li and Zhang (2013) show that advance demand information can hurt a seller's profit. Crowdvoting differs from these settings in several ways: Whereas pre-orders provide information about actual demand realizations, crowd votes may be an unreliable, noisy proxy for demand. This is aggravated by the fact that voters do not receive a reward for voting and the question of alignment between the tastes of the voter base and the consumer base. Marinesi and Girotra (2013) further caution that votes may not supply accurate information to the firm. Our study complements this stream of work by advancing our understanding of how crowdvoting impacts operational decisions and consumer behavior.

The recent surge in subscription-based platforms where users pay a flat fee to access a wide variety of products or services has led to a body of work. Wang et al. (2022) empirically study the impact of email engagement on a subscription platform and find that it improved subscriber retention. Several papers consider the platform's pricing decision between offering a subscription plan versus a per-use plan in various settings, for example, congestion-prone services (Cachon and Feldman 2011) and online grocery retail (Belavina et al. 2017). Choi et al. (2023) investigate how a platform's prediction capability and product curation affects pricing decisions in a ship-then-shop subscription model (e.g., Amazon's Try Before You Buy). Given that consumers on subscription-based platforms value variety, Gal-Or and Shi (2022) study how differentiation in vendors affects a platform's market entry decision and Kan et al. (2022) show how personalized recommendations can increase sales diversity. We add to this body of work by developing a fundamental understanding of how product co-creation through crowdvoting affects user behavior and product assortment for a subscription-based platform.

In summary, our paper contributes to the literature by studying a new business practice of using crowdvoting in retailing, a phenomenon that is growing in importance on e-commerce

platforms. We present the first empirical evidence of the business value of this service. We provide specific insight into how it shapes a retailer's purchasing strategy, operational performance, and customer satisfaction. More broadly, these insights can also be valuable to subscription-based platforms who are interested in utilizing product co-creation strategies to increase user engagement and loyalty.

Chapter 3

EMOTION AI IN DISGUISE SPURS STRATEGIC BEHAVIOR IN CUSTOMER CARE

3.1 Introduction

Customer care, which focuses on building and strengthening long-term customer relationships, is increasingly recognized as a critical component of sustainable business success (Huang and Rust 2023). Unlike mere informational provision or problem resolution typically associated with customer service, customer care fosters customer engagement by constructing a profound emotional connection with the company. This perspective of service transcends conventional transactional interactions (Schanke et al. 2021, Luo et al. 2019) and demands a customer service approach that is empathetic.

However, customer care faces two significant challenges. Firstly, the strategic behavior known as "emotional misrepresentation," where the expressed emotion level differs from the actual emotion level, is a commonly used strategy in negotiation and bargaining (Campagna et al. 2016). Such strategic behavior can lead to a loss for the firm due to its inability to accurately allocate resources in alignment with the true states of its customers, referred to as misallocation loss (Frankel and Kartik 2019, Yu et al. 2022). Neglecting such strategic interaction can further influence consumer satisfaction and loyalty. Hence, the existence of such strategic behavior can undermine the effectiveness of customer care and increase costs for businesses (Campagna et al. 2016, Frankel and Kartik 2019, Yu et al. 2022).

Secondly, the choice of agent for delivering customer care, whether human or AI, remains a complex issue. This is particularly true with the evolution of emotion AI, which involves AI systems that can interpret and interact with humans based on an understanding of human emotions. The advent of sophisticated Large Language Models (LLMs) like ChatGPT has

shown significant promise in enhancing customer care (Huang and Rust 2018). However, the impact of AI identity cues on emotional misrepresentation remains unclear, despite their relevance to this context. In particular, it is uncertain which specific identity cues might trigger such strategic behavior from customers. Prior literature has shown that AI's identity cues can significantly influence the customer decision-making process. A previous study by Luo et al. (2019) demonstrates that while chatbots can match the performance of skilled human workers in promoting customer purchases, disclosing a chatbot's identity before the conversation can significantly diminish its effectiveness, specifically resulting in a steep decrease in purchase rates.

The choice of an agent's identity can impact such strategic behavior; however, previous papers have not adequately distinguished between the content effect and the identity effect of AI. Both the response generated by AI and the identity of AI itself can influence the interaction between customers and AI. Previous studies have examined these two effects together but have not clearly distinguished between them. In our research, we focus on the identity effect. The reason is that different identity cues can significantly influence customers' decisions, even when the content remains the same. Moreover, the identity effect has both theoretical and practical tensions.

From an academic perspective, skepticism towards disclosing AI identity persists across scholarly discussions. The practice of employing AI or revealing its identity is often criticized in comparison to utilizing human agents or maintaining anonymity, as delineated in Table 3.1. The majority of prior studies converge on the notion that AI, in general, leads to adverse outcomes, whereas human or anonymous identity cues are associated with more favorable results. Researchers have discovered that AI negatively affects various metrics, including purchase conversion, persuasion outcomes, and service evaluation. Hodge et al. (2021) further suggest that the disclosure of AI's identity, coupled with attempts at humanization, can exacerbate these negative impacts. An exception to this trend is noted by Gnewuch et al. (2023), who examine employee workload and logically infer that chatbot integration reduces such burdens.

Table 3.1: Literature Summary: Effect of Agent Identity on Customer Service Outcomes

Study	(1) Outcome	(2) AI	(3) Human	(4) Non-disclosure
Luo et al. (2019)	Purchase Conversion	X	✓	✓
Shi et al. (2020)	Persuasion Outcome	X	✓	-
Schanke et al. (2021)	Transaction Conversion	X	-	-
Han et al. (2022)	Service Evaluations	X	✓	-
Cui et al. (2022)	Price Quote	X	✓	-
Gao et al. (2023)	Engagement; Resolution	-	✓	-
Gnewuch et al. (2023)	Interaction with Agents; Workload	✓	X	✓

X: agents with this identity negatively impact the outcome. ✓: agents with this identity positively impact the outcome. -: agents with this identity are not studied.

From a practical perspective, the ongoing concerns and ambiguities regarding the impact of content and the impact of identity cues significantly hinder AI adoption in online commerce. Furthermore, the situation is intensified by regulatory measures in several jurisdictions that mandate the disclosure of deploying AI. For instance, the California Bot Disclosure Law[1], enacted in July 2019, requires companies to announce their use of bots, potentially affecting business owners' interests adversely, as highlighted by Luo et al. (2019). The recently passed EU AI Act[2] also demands transparency when deploying AI. This regulatory push for transparency may create a misalignment in incentives between regulators and companies, as businesses may perceive these disclosure requirements as detrimental to their interests.

Therefore, this study aims to investigate the identity effect of AI in the customer care context. We (1) examine the effects of varying identity cues on consumer strategic behavior, (2) assess whether the disclosure of AI identity cues negatively influences this dynamic, potentially leading to heightened strategic behavior in interactions with AI, (3) explore the reasons behind the diverse behavioral patterns exhibited by customers when interacting with agents presenting different identity cues, and (4) propose strategies for businesses on the optimal assignment of customers to different types of agents. Our findings reveal that disclosing a bot's identity does not invariably lead to negative consequences, particularly concerning strategic behavior. Conversely, concealing a bot's identity may provoke more strategic behavior among consumers, potentially leading to lower satisfaction in scenarios involving online complaints.

Our study presents a timely and much-needed inquiry into emotion AI in customer care. We are among the first to empirically explore strategic behavior in this context, marking a novel direction in this domain. The scarcity of research in this area is surprising given the ever-increasing prevalence of chatbots in customer service roles, further emphasizing the relevance and importance of our research in this field. This gap in existing research

[1] https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001

[2] https://artificialintelligenceact.eu/

is due to the intricate nature of strategic behaviors and the challenges associated with the detection and measurement of them. This aspect of consumer behavior, particularly during interactions with chatbots that can have varying identities (AI, Human, GPT-based AI, etc.), is multi-dimensional and requires a more nuanced and detailed approach for a comprehensive investigation. Hence, our research aims to bridge this gap by providing empirical insights into the intricate dynamics of customer strategic behavior in response to emotion AI.

Motivated by the apparent paradox and research gap, our study endeavors to investigate the following questions within a post-sale context, where the primary customer focus is seeking emotional care rather than gathering information as often seen in pre-sale or in-sale situations. Firstly, we examine how disclosure or non-disclosure of an agent's identity impacts customers' strategic behavior. Secondly, we seek to determine which situation results in detrimental effects on customer service evaluations. Lastly, we aim to discern the underlying mechanisms driving such influences. Our investigation not only provides valuable insights into the nuances of customer-agent interactions but also has the potential to inform effective customer service strategies in an increasingly AI-driven world.

3.2 Hypotheses Development

AI, particularly in emotion detection, is inherently flawed due to the subjective nature of emotions. There is almost always noise in emotion AI detection results and such challenges are well-documented in previous literature (Purdy et al. 2019, Yu et al. 2023). For example, as Purdy et al. (2019) points out, AI usually has limited abilities to recognize and interpret emotional expressions across diverse cultures, which makes it harder to draw precise conclusions. Given this inherently imperfect capability of AI in emotion detection, people tend to have stereotypes about AI, perceiving it as mechanical and lacking emotional depth (Sundar and Kim 2019).

Such stereotypes about AI significantly influence customers' emotional misrepresentation behavior in the context of customer service. The effectiveness of the strategy heavily relies on the customer's perception that the service agents are empathetic and understand their

emotional states (Reis et al. 2017). However, there is a prevalent belief among customers that AI agents, powered by algorithms, are incapable of human emotions, leading to responses that are perceived as preprogrammed. From the perspective of customers, expressing negative emotions is costly (Gross 1998). If AI is perceived to lack precise emotion detection capabilities, customers may not be motivated to act strategically by exaggerating their emotions. As a result, AI agents regulate the emotional misrepresentation behavior of customers. Consequently, we hypothesize:

Hypothesis 1a. *Agents with non-disclosure identity cues will trigger greater strategic behavior than agents with AI identity cues.*

The influence of non-disclosed agent identities stems from the fundamental understanding that the transparency of agent identities significantly impacts customer behavior. The academic literature broadly concurs that transparency fosters trust and diminishes uncertainty, thereby promoting more cooperative customer behavior (Stevens et al. 2018, Tapscott and Ticoll 2003). As mentioned previously, the ambiguous nature embedded in non-disclosure identity cues will trigger illegitimacy discount compared to human identity cues. People have doubts or concerns when the identity is unclear. Additionally, when agent identity is not revealed, it might directly encourage customers to act more strategically, taking advantage of the lack of consequences due to anonymity. This behavior may take various forms, including tougher bargaining, information withholding, or manipulation of the interaction to achieve personal objectives. In the digital interaction realm, research suggests that perceived anonymity can encourage more self-centered behaviors due to a lack of personal accountability (Joinson 2001). Additionally, such information asymmetry can lead one party (the platform) to bear the risk of the other party (the customer) behaving in self-interest. Previous studies, such as those conducted by Holmström (1979), have shown that when there is information asymmetry, it often leads to a moral hazard situation, where one party is more likely to behave strategically at the expense of the other party. Consequently, in our specific context, we hypothesize that individuals display more strategic behavior when confronted

with non-disclosure identity cues than human identity cues.

Hypothesis 1b. *Agents with non-disclosure identity cues will trigger greater strategic behavior than agents with human identity cues.*

The concept of psychological distance, as explored in the Construal Level Theory (CLT) by Trope and Liberman (2010), provides a framework to understand how individuals perceive and react to various stimuli or entities that are distant in time, space, social dimensions, and hypothetically. The theory posits that the greater the distance, the more abstractly people think about the entity, event, or person in question. This abstraction level influences their reactions and behaviors towards it. In the context of customer service, this theory can be applied to explain the phenomenon of emotion misrepresentation, particularly how consumers exaggerate their negative emotions more when interacting with a customer service agent identified as human compared to an agent with a non-disclosure identity cue.

When consumers interact with customer service agents who have a non-disclosure identity cue, they perceive a greater psychological distance. This is because the lack of personal or "human" identifiers makes the agent seem more abstract and less relatable, leading to a depersonalized perception of the interaction. Studies have shown that psychological distance can affect various aspects of cognition and behavior, including decision-making, empathy, and moral judgment (Eyal et al. 2008). In the context of customer service, this suggests that the perceived distance between consumers and agents can significantly impact how consumers choose to express their emotions. The increased psychological distance may lead to more strategic behavior, such as emotional misrepresentation. When consumers perceive a human identity cue, the psychological distance is reduced. The presence of human elements in the interaction leads to a more concrete construal of the situation, where social norms and empathy play a larger role in shaping behavior. Consumers are less likely to exaggerate their emotions, as they view the agent as more relatable and deserving of honest communication. Henderson et al. (2006) demonstrates that an increase in temporal distance, a dimension of psychological distance, leads to more strategic behavior. Individuals become more willing to

compromise on lower-priority issues in exchange for concessions on matters of higher priority. Based on the arguments above, we hypothesis the following

Hypothesis 2. *Agents with non-disclosure identity cues have longer psychological distance than agents with human identity cues, which in turn, leads to greater strategic behavior.*

If Hypothesis 1a and 1b are true, we then want to answer whether such effect exists in a more realistic *multi-round* negotiation in the after-sale complaint setting and what are the mechanisms that drive such effects. The resulting behavior could stem from two potential pathways. Firstly, the non-disclosed agent might naturally attract a subset of users who are inherently more strategic to engage in multi-round negotiation. The ambiguity about the agent's identity might be particularly appealing to these individuals, leading to adverse selection (Akerlof 1970). As a result, we may observe overall increases in strategic behavior. Alternatively, the non-disclosed identity of the agent could directly incentivize strategic behavior among all customers, as hypothesized above. The existence of adverse selection has been found in various managerial contexts, such as insurance markets (Stiglitz et al. 2020), online auctions (Lewis 2009), and corporate strategy (Hamrouni et al. 2015), where asymmetric information or uncertainty can lead to adverse selection effects. Hence we propose the following hypothesis.

Hypothesis 3. *Participants with lower morality, reduced altruism, and higher complaint frequency are more likely to engage in multi-round negotiation with agents with non-disclosure identity cues.*

Transparency and trust have long been considered essential factors for customer satisfaction (Klein et al. 2016). When an agent's identity is not disclosed, it may lead to feelings of uncertainty and mistrust, ultimately reducing customer satisfaction. This is corroborated by studies like the one conducted by Djan and Adawiyyah (2020), which found that customer satisfaction is directly affected by trust. In line with these findings, non-disclosed agent identities, which may diminish trust, will further impact customer satisfaction negatively. Hence, we propose the following two hypotheses.

Hypothesis 4. *Agents with non-disclosure identity cues can lead to reduced customer satisfaction. Such a negative effect is mediated by a reduction in both trust and perceived quality of service.*

3.3 Study 1

3.3.1 Experiment Setup

In Study 1, we aim to empirically validate the existence of emotional misrepresentation within the context of consumer complaints. Furthermore, we hypothesize that such strategic behavior is influenced by the identity cues of customer service agents. We conduct an online experiment to simulate a post-purchase complaint scenario in the context of food delivery. Our sample comprises 2,025 participants, collected within a span of two days. We recruit through multiple major social media platforms in China, including WeChat, Weibo, and Xiaohongshu.[3] The experimental procedures are executed via a web-based interface..

Upon arrival, participants receive instructions indicating that they are to engage in two distinct tasks, which involve negotiation with a virtual customer service agent and interactive decision-making. We employ uniform chatbots as customer service agents but display different identities. Participants are informed that their monetary reward is performance-based and contingent upon the outcomes of the two tasks. This incentivizes consequential decision-making on their part. We inform the participants that the potential monetary reward for each task is capped at 15 CNY, yielding a total possible compensation of 30 CNY. The cover story used in the experiment is presented in Appendix B.1.

In both tasks, participants are randomly assigned to one of three customer service agents with different identity cues: AI, human, or non-discloses. Participants are informed that they will negotiate compensation with the same agent across both tasks. Following Andrade and Ho (2009), we employ modified versions of the traditional Dictator Game (DG) and Ultimatum Game (UG). Andrade and Ho (2009) adapt the DG and UG to study emotional

[3]We chose to recruit participants from major media platforms to reflect the demographic of individuals who frequently order food online.

misrepresentation in human-to-human negotiation. In the standard DG, the game involves a proposer (e.g., a service agent) and a receiver (e.g., an unhappy customer); the proposer is endowed with a sum of money and must decide how to allocate it between themselves and the receiver, who has no choice but to accept the proposed division. In the traditional UG, the proposer is once again endowed with a sum of money and responsible for suggesting a division of the money between themselves and the receiver. Contrary to the DG, the receiver has the option to either accept or reject the offer. A rejection results in neither party receiving any monetary gain.

In the study by Andrade and Ho (2009), the authors measure the receiver's anger level, denoted as *Anger1*, in the DG after the receivers observe an unfair offer by the proposer.[4] Subsequently, in the UG, participants are randomly divided into a control group and an emotion-disclosed (treatment) group. In both groups, the level of anger, denoted as *Anger2*, is measured again before the proposer makes an offer. However, the treatment group is informed that their measured level of anger will be disclosed to the proposer, whereas participants in the control group are told that the measurement of their anger level is solely for recording purposes.

We further adapt the approach used by Andrade and Ho (2009) to investigate strategic behavior in interactions, not with human counterparts, but with customer service agents displaying different identity cues. In our modified DG, customer service agents act as proposers, while participants play the role of the receiver. The agents have the authority to offer compensation up to a limit of 15 CNY, a detail also communicated to participants at the start of the game. Serving as the representative of the platform, the agents aim to resolve the complaint while minimizing the compensation payout, whereas participants seek to maximize their compensation. Hence, it is similar to the traditional DG, where the proposer decides how to divide a sum of money between themselves (the company) and the customers.

[4]In DG, the proposer has the freedom to distribute money, and the receiver lacks any decision-making power. It is commonly observed that proposers allocate a larger portion of money to themselves than to the receiver.

In the UG, participants are informed that they will engage in a second consumer complaint task with the same agent. The roles remain the same, with participants as receivers and the customer service agents as proposers. However, unlike the traditional division of a monetary reward, the proposer decides the compensation amount from the limit of 15 CNY. Additionally, following Andrade and Ho (2009), we measure *Anger1* and *Anger2* to identify emotional strategic behavior among participants. The detailed process and transcript used in the experiment can be found in Appendix A.3.2 Figure A.1 and Appendix A.3.3 Figure A.2.

3.3.2 Identity Cues and Emotional Strategy

We standardize the negotiation process while varying the identity cues of customer service agents, with distinctions only in the titles and profile photos presented.[5] For the three groups with identity cues of non-disclosure, human, and AI, the displayed titles are "Agent Xiaomei,""Human Agent Xiaomei," and "AI Agent Xiaomei," respectively. [6] Each title is accompanied by a corresponding profile photo, as illustrated in Figure 3.1. This approach ensures a controlled environment to assess the impact of identity cues on the interaction while maintaining consistency across other variables.

As demonstrated in the experimental procedure in Appendix A.3.3, participants are initially randomized to one of the three customer service agents with distinct identity cues. Subsequently, they engage in the DG with the designated agent. Within the DG, agents propose an unfair offer, functioning as an emotional trigger to induce a specific degree of anger in participants. Immediately following this, participants report their current anger level on a 101-point scale (from 0 to 100), noted as *Anger1*.

The experiment then transitions to the UG. Participants are informed they will interact with the same customer service agent. Before the agents make the second compensation

[5]The dialogue, responses, and compensation strategies are standardized across all three groups, differing only by the identity cues presented.

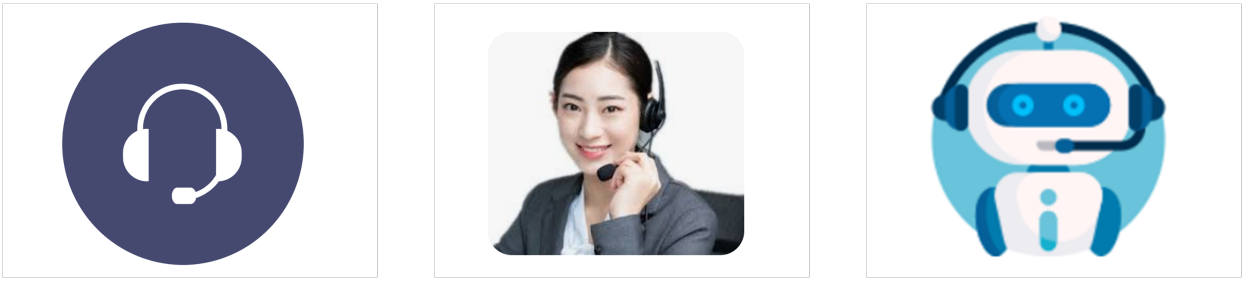
[6]The name "Xiaomei" is commonly used for customer service roles within the Chinese context.

decision, participants are required to report their current anger level once more, noted as *Anger2*. Participants are then randomly assigned to either the emotion-disclosed group or the control group. In the emotion-disclosed group, participants are informed that their latest anger assessment will be shared with the customer service agent before the compensation decision is made. Conversely, participants in the control group are simply informed that their updated anger level will be utilized solely for research purposes by the experiment designers.

According to Andrade and Ho (2009), customers are likely to strategically amplify their expressed emotions to influence proposers within traditional social interaction contexts like DG and UG. Our study extends this framework to a post-purchase customer care context, which is characterized not by a simple two-party interaction but by a negotiation between customers and customer service agents. In our case, the agents can be AI with varied identity cues, which is the key difference between this experiment and those in Andrade and Ho (2009). These varied identity cues serve as our treatments of interest. Consequently, we hypothesize that participants in the emotion-disclosed group will similarly misrepresent their emotional states, exaggerating their level of anger to influence customer service agents' compensation decisions, knowing that these agents will take their emotional state into account. Specifically, we anticipate that the second reported anger level *Anger2* will be greater than the first *Anger1*. Secondly, we hypothesize that such strategic behavior is contingent upon the identity cues associated with the agents. In our context, these identity cues are the title and profile photos of the agent displayed during the interaction. Existing research indicates that such cues can influence perceived humanization (Oh et al. 2018, Gao et al. 2023). Enhanced humanization can potentially alter participants' strategic behavior by influencing psychological distance. Given that participants typically maintain varying degrees of psychological distance towards AI compared to human agents, we anticipate that these differences will modulate the strategic behaviors exhibited by participants.

The study utilizes an online experimental design with a sample of 2,025 participants recruited through popular Chinese social media platforms. The experiment is structured as a $3 \times 2 \times 2$ mixed design. This design includes three between-subjects factors for agent

Figure 3.1: Agents' Profile Photos and Titles Displayed in Study 1



(a) Agent Xiaomei (b) Human Agent Xiaomei (c) AI Agent Xiaomei

identity cues (human, AI, or non-disclosed), two between-subjects factors for the disclosure of emotional information (control group and emotion-disclosed group), and two within-subjects factors capturing the temporal evolution of self-reported anger levels (*Anger1* and *Anger2*).

As delineated in the flow chart presented in Appendix A.3.2, upon arrival at the experiment webpage, participants are informed that their involvement will consist of two tasks related to the context of after-sales service complaints. Specifically, they will engage in negotiations with customer service agents to determine the amount of monetary compensation, up to 15 CNY, which the agents are authorized to offer. It is emphasized that the participants' payment for participating in the study is directly contingent upon the outcomes of these negotiations (to ensure incentive compatibility). Following the initial introduction and training phase, participants are randomly assigned to interact with an agent displaying one of the three distinct identity cues: human, AI, and non-disclosed. Each type of agent is identified by the corresponding title and profile photo. Participants are notified that they will engage in negotiations with the same agent across both tasks. [7]

[7]We conduct training on the different identity cues, followed by a qualifying test to ensure that participants can differentiate among these cues. Those who do not pass the test are excluded from further participation in the experiment. Note that disclosing three agent types in the training process mimics the reality that customers are aware of the fact that different agent types exist in contemporary digital business contexts.

Task 1 is a modified version of the traditional DG. In a conventional DG setting, a proposer is allocated a sum of money that they can distribute between themselves and a receiver. In our adapted version, participants are informed that the customer service agents have a maximum compensation limit of 15 CNY. To systematically manipulate both the offers and the consequent emotional responses from participants, we standardized the responses from all chatbots, regardless of different identity cues. Specifically, each chatbot is programmed to propose an unfair offer of 5 CNY out of a 15 CNY limit. Importantly, participants are not aware of this uniform offer strategy. The design of the online interface minimizes the likelihood of participant collusion or discovery of the experimental conditions. After receiving the unfair offer, participants are asked to report *Anger1*.

Following the practice of Andrade and Ho (2009), we insert a filler task consisting of 23 demographic questions (e.g., age, gender) unrelated to the emotional measure between the DG and the UG. Given the time delay brought about by the filler task, participants' anger levels will naturally mitigate over time. Hence, without any intervention, we should observe a drop in anger levels. This design gives us more confidence to conclude that any observed increase in reported anger is strategic.

Participants then proceed to Task 2, which is also a modified version of a traditional UG. In Task 2, participants are reminded that they will continue to interact with the same customer service agent. The task involves a second negotiation regarding monetary compensation, constrained by the same 15 CNY limit as in Task 1. Before beginning the negotiation, participants are asked to report their level of anger, *Anger2*. Within each identity cue category, participants are randomly assigned to either the emotion-disclosed group or the control group. Participants in the emotion-disclosed group are informed that the anger levels they report during this task will be disclosed to the agents. On the other hand, those in the control group are informed that their reported levels of anger will be used exclusively for research purposes by the experiment's designers. [8]

[8] A experimental manipulation is introduced here, focusing on the affective information in a between-subjects design.

3.3.3 Experiment Results

Within the cohort of 2,025 participants, 1,224 successfully complete the training test on identity cues, two manipulation tests concerning the identities of customer service agents, and five attention check questions. These 1,224 data points are included in the subsequent analysis, and the detailed subjects count in each group are presented in Table A.3 in Appendix A.3.1. A paired-sample t-test is conducted to examine the fluctuations in anger as a within-subjects factor influenced by varying identity cues. To capture the within anger level change, we use *Anger2* minus *Anger1*.

When interacting with AI agents, there is no statistically significant difference in the reported anger levels for the emotion-disclosed group ($M = 2.04, SD = 1.33, t = 1.52, p = 0.128$) nor for the control group ($M = -2.13, SD = 1.33, t = -1.60, p = 0.111$). Conversely, engagement with human agents leads to a significant escalation in anger within the emotion-disclosed group ($M = 4.20, SD = 1.32, t = 3.17, p = 0.002$), and a marginal reduction in the control group ($M = -1.90, SD = 1.10, t = -1.73, p = 0.086$). Notably, in scenarios featuring non-disclosed agent identities, the emotion-disclosed group exhibits a significant rise in anger levels ($M = 2.23, SD = 0.88, t = 2.52, p = 0.013$), while the control group exhibits a considerable decline ($M = -5.00, SD = 1.52, t = -3.29, p = 0.001$).

These findings provide preliminary evidence that participants' strategic expression of anger is influenced by the perceived identity of the customer service agent. The greatest display of strategic behavior is observed with non-disclosed identity cues. In the control group, we observe a significant decrease in anger, which could be attributed to the temporal gap introduced by the filler task conducted between tasks 1 and 2. When participants are aware that their expressed anger will be revealed to the agents, they report significantly higher anger levels. This phenomenon is only partially observed with human agents, where an increase in anger is noted for the emotion-disclosed group, but no corresponding significant decrease is observed within the control group. This pattern somewhat diminishes the assertion that the elevation in anger levels is a result of emotion manipulation for strategic

purposes with human agents. Lastly, the participants' interaction with AI agents demonstrated no strategic misrepresentation of anger, as evidenced by the lack of significant change in anger levels for both the emotion-disclosed and control groups. This outcome suggests that participants are less likely to engage in the strategic display of emotions when they interact with agents with AI identity cues.

Our analysis further includes a one-way ANOVA to compare the anger levels reported in the emotion-disclosed (ED) group with those in the control (C) group under different identity cues. Consistent with the premise that initial emotional states are comparable across conditions, no significant differences are found in the anger levels (*Anger1*) during Task 1 for any identity cues. Specifically, for the AI identity, we observe ($M_{ed} = 55.28$ vs. $M_c = 57.17$, $SDs = 24.67$ and 27.00, $F(403) = 0.538$, $p = .464$), for human identity ($M_{ed} = 50.81$ vs. $M_c = 56.22$, $SDs = 30.55$ and 30.56, $F(403) = 3.146$, $p = .077$), for non-disclosure identity, ($M_{ed} = 60.10$ vs. $M_c = 56.79$, $SDs = 22.27$ and 27.79, $F(415) = 1.793$, $p = .181$).

In Task 2 of the experiment, we anticipated a strategic display of emotions in response to the manipulation of affective information within the emotion-disclosed group. For non-disclosure identity cues, we find a pronounced divergence in anger levels (*Anger2*) between the emotion-disclosed and control groups ($M_{ed} = 62.33$ vs. $M_c = 51.79$, $SDs = 21.25$ and 28.55, $F(415) = 18.229$, $p < .001$). However, the data show no strategic behavior when interacting with AI and human identity cues. For AI identity cues, the mean anger levels between the emotion-disclosed and control groups show minimal disparity ($M_{ed} = 57.32$ vs. $M_c = 55.05$, $SDs = 25.77$ and 26.59, $F(403) = 0.758$, $p = .384$), and a similar pattern emerges for human identity cues ($M_{ed} = 55.01$ vs. $M_c = 54.32$, $SDs = 29.02$ and 31.31, $F(403) = 0.054$, $p = .816$). This finding echoes the previous hypothesis of significant anger level change for non-disclosure identity cues. It indicates strategic behavior from the participants in the emotion-disclosed group, leveraging the disclosure of their emotional state as a mechanism to influence the outcome of the negotiation. This result confirms Hypothesis 1a and 1b, suggesting that emotional misrepresentation occurs in the customer service context only when the agents' identity cues are non-disclosed.

The lack of significant differences between the emotion-disclosed and control groups in the AI and human conditions could indicate that participants do not perceive a benefit in strategically using their emotional expressions, or they may believe that such expressions are ineffective. Conversely, the notable difference in the non-disclosure condition suggests a scenario where participants feel that revealing their emotions might influence the decision of an indistinct agent. This insight offers a valuable avenue for future research to further investigate the role of AI identity disclosure and its effect on negotiation processes.

3.3.4 Mediation

We have provided evidence that the presence of non-disclosure identity cues in agents significantly influences the strategic behavior of participants. The subsequent analysis aims to uncover the mechanisms underlying the observed escalation in strategic behavior in response to non-disclosure identity cues. Initially, we must determine the extent of strategic misrepresentation in the reported anger levels in the emotion-disclosed groups. We cannot simply use *Anger2* as a direct indicator of strategic behavior, owing to the confounding influence of both genuine anger levels and strategic motives. However, the control group participants, whose responses are unaffected by strategic considerations, are presumed to convey their authentic levels of anger. Thus, the anger levels reported by the control group can act as a baseline for the emotion-disclosed group's true anger levels.

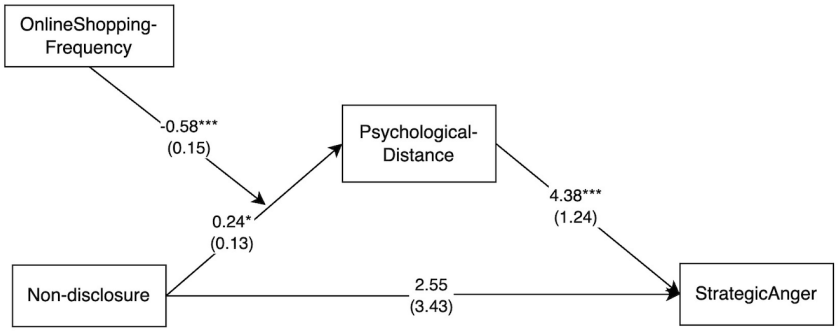
For the purpose of quantifying the strategic emotional expression, we use matching between participants from the emotion-disclosed group and the control group. This matching procedure is executed independently across the three distinct categories of agent identity cues. We compute a propensity score for each participant predicated upon demographic variables listed in Table A.1 in Appendix A.2. Subsequently, each participant in the emotion-disclosed group is paired with a control group counterpart exhibiting the nearest propensity score. This methodology provides an estimated actual anger level for the participants in the emotion-disclosed group. The disparity between the self-reported anger level and the estimated actual anger level is considered to be the strategic expression of emotion, herein

referred to as *StrategicAnger*. To assess the efficacy of our matching procedure, we also conduct a balance check on the covariates within the matched dataset, as presented in Table A.2. The results confirm that our matching protocol has successfully facilitated the formation of comparably balanced groups.

We hypothesize that participants' psychological distance towards customer service agents mediates the effect of identity cues on individuals' strategic behavior, and such effect is moderated by participants' online shopping frequency. As previously explained in Section 3.2, a longer psychological distance leads to increased strategic behavior. Additionally, we also argue that such impacts are more prominent when individuals are less experienced in online shopping. Customers with extensive online shopping experience are likely more accustomed to the impersonal nature of digital platforms. This familiarity could reduce the psychological distance they feel toward customer service agents when those agents' identities are not disclosed. These customers might have developed a certain level of comfort and trust in online interactions, making the lack of personal identity less significant in their service evaluation.

PsychologicalDistance is measured using three items from Li and Sung (2021) on a seven-point scale (e.g., "I am familiar with the agent."). Additionally, data on *OnlineShoppingFrequency* are collected during the filler tasks. To test the moderated mediation effect, we implemented a bootstrapping analysis utilizing the R PROCESS Model 7 (moderated mediation analysis), accompanied by 10,000 bootstrap resamples (Hayes 2015). In our analysis contrasting non-disclosure and human identity cues, the study sample comprises 412 participants, including 211 in the non-disclosure group and 201 in the human group. Similarly, when comparing non-disclosure and AI identity cues, the total number of participants was 415, with 211 in the non-disclosure group and 204 in the AI group. All continuous predictors have been mean-centered prior to the analysis to enhance interpretability.

Figure 3.2: Moderated Mediation Analysis Between Non-disclosure and Human



Note. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Moderated Mediation Analysis Between Non-disclosure and Human

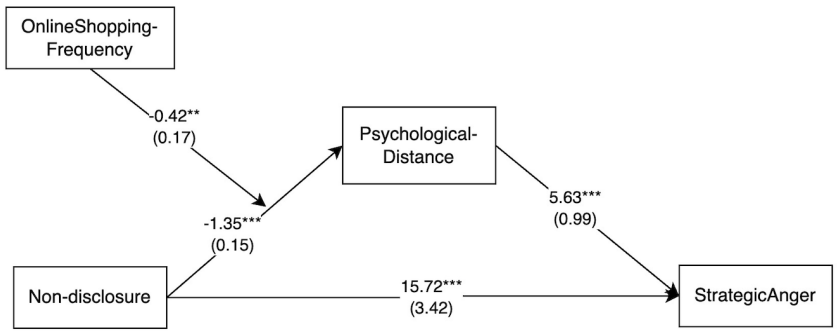
Firstly, we conduct the analysis comparing non-disclosure and human identity cues. We employ a binary independent variable, Non-disclosure, assigned a value of 1 for nondisclosure identity cues and 0 for human identity cues. The dependent variable of interest is *StrategicAnger*, which we have quantified through the aforementioned matching technique. Compared to human agents, non-disclosure agents trigger greater psychological distance among individuals. This increased psychological distance, in turn, leads to heightened strategic behavior.

To investigate the nature of this effect, we conduct a simple slope analysis and evaluate the marginal effect of the non-disclosure identity cue versus the human identity cue at one standard deviation above and below the mean of *OnlineShoppingFrequency*. Our findings reveal a significant moderation effect of *OnlineShoppingFrequency* ($\beta = -0.58, SE = 0.15, t(408) = -3.905, p < 0.001$). Furthermore, there is a notable mediation effect of *PsychologicalDistance* for individuals with less experience in online shopping (*OnlineShoppingFrequency* = 2.29, 1 SD below the mean), with ($\beta = 3.37, 95\%$ CI $= [1.12, 6.10]$). This effect is not observed among more experienced online shoppers, suggesting that the mediation effect of *PsychologicalDistance* is indeed contingent on the frequency of online shopping. Individuals exhibit a

greater *PsychologicalDistance* toward non-disclosed agents, which leads to increased strategic behavior, thus confirming Hypothesis 2.

Moderated Mediation Analysis Between Non-disclosure and AI

Figure 3.3: Moderated Mediation Analysis Between Non-disclosure and AI



Note. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

We further conduct a moderated mediation analysis to explore the differential impacts of non-disclosure versus AI identity cues on StrategicAnger. Our findings indicate that AI agents elicit a higher level of *PsychologicalDistance* among individuals than non-disclosure agents. Consequently, AI identity cues are found to trigger greater strategic behavior via the mediating role of *PsychologicalDistance*. This relationship is also subject to moderation by *OnlineShoppingFrequency*, as previously discussed. Notably, experienced online shoppers tend to experience reduced *PsychologicalDistance* toward agents with non-disclosed identities. In this analysis, we adopt a model similar to the one detailed in section 3.3.4, but we code non-disclosure as 1 for non-disclosed identity cues and 0 for AI cues. The result confirms a significant moderation effect of *OnlineShoppingFrequency* ($\beta = -0.42, SE = 0.17, t(411) = -2.490, p = .013$) as in Figure 3.3.

Similarly, we examine the mediation effects of *PsychologicalDistance*, with *OnlineShoppingFrequency* serving as a moderator. By assessing the marginal impact of non-disclosure

versus AI identity cues at one standard deviation above and below the mean of *OnlineShoppingFrequency*, our bootstrapped analysis reveals a consistently significant negative indirect effect across all levels of *OnlineShoppingFrequency*, though with varying magnitudes. For less experienced online shoppers (*OnlineShoppingFrequency* = 2.34, 1 SD below the mean), we observe an effect of ($\beta = -5.44, 95\%$ CI $= [-9.03, -2.52]$). For those more accustomed to online shopping (*OnlineShoppingFrequency* = 4.14, 1 SD above the mean), the effect intensifies to ($\beta = -9.73, 95\%$ CI $= [-13.67, -6.21]$). This pattern of mediation via *PsychologicalDistance* is consistent with our earlier observations in Section 3.3.4 and lends further support to Hypothesis 2.

Additionally, it is also important to notice that there exists a significant positive direct effect of non-disclosure on *StrategicAnger* ($\beta = 15.72, SE = 3.42, t(412) = 4.601, p < .000$). A potential rationale for this finding could be the perceived limitations of an AI agent's emotion detection capabilities, as discussed previously. AI is often considered less adept at interpreting emotions, thereby influencing individuals to refrain from engaging in emotional strategic behavior. This positive direct effect is more substantial than the negative indirect effect mediated through *PsychologicalDistance*, suggesting that the direct effect predominates. Despite AI leading to greater psychological distance, the overall impact is dominated by the direct effects of AI identity cues. Consequently, we observe reduced strategic behavior in interactions with AI agents.

3.4 Study 2

3.4.1 Experiment Setup

Study 1 demonstrates the presence of emotional misrepresentation in an idealized and simplified customer service context and shows that this effect depends on the agent's identity cues. However, real-life customer complaints often extend beyond simple, one-round negotiations, typically involving multiple rounds of interaction between agents and customers. To capture this complexity, our second experimental design simulates a real-life after-sale

scenario in the food delivery industry, which consists of a two-round negotiation between the agents and customers. Moving away from the simplified customer service framework in Study 1, we design a dialog-based interaction procedure that engages participants with customer service agents. In this process, participants choose their responses to the agents, who in turn provide corresponding replies. As in Study 1, the process was standardized to ensure consistency in replies and compensation plans across different identity cues groups. We posit that customers have encountered an issue with the quality of food and filed a complaint with the platform. Agents then interact with the customers, assisting them to solve the problem. The cover story we used is presented in Appendix B.1. To more closely mimic real-life scenarios, where customers often undergo multiple rounds of negotiation with agents to reach an agreement, our experiment unfolds over two rounds. The experimental process is illustrated in Figure A.1 in Appendix A.3.2.

In Study 2, we retain the three identity cues for customer service agents used in Study 1, namely non-disclosure, AI, and human. Furthermore, we introduce a fourth identity cue: GPT-based AI, an advanced AI chatbot powered by a large language model. As in Study 1, a crucial aspect of analyzing the emotional strategic behavior involves the dependent variable of *StrategicAnger*, ideally represented by the discrepancy between expressed anger levels and true anger levels of the same individual. However, it is challenging to measure both expressed and true anger accurately from the same participants. This is because such a design implicitly prompts participants to acknowledge their strategic behavior by indicating different levels of expressed versus felt anger, which could mutually influence the responses to these questions and lead to biased outcomes. Therefore, following the approach in Study 1, we randomly divide participants under each identity cue into two groups: an emotion-express group and an emotion-feel group. Participants in the emotion-express groups are instructed to select the response that best represents the level of anger they intend to express, labeled as $ExpressedAnger$. In contrast, participants in the emotion-feel groups are asked to report the level of anger they genuinely feel, denoted as $FeltAnger$.[9] Thus, Study 2 employs a 4×2

[9]Different from Study 1, where we utilize a numerical scale from 0 to 100 to measure participants' anger,

mixed design.

For the four different identity cues, we use the title and profile photos depicted in Figure 3.4, with dialogue or wording that remains consistent across all identities.[10] We first introduce the participants to a food delivery complaint context, where the participants encounter food quality issues and file a complaint to the food delivery platform for compensation. Then, participants will be randomly assigned a customer service agent with one of the four identity cues: AI, human, non-disclosure or GPT-based AI. Subsequently, participants are randomly placed into either the emotion-express or emotion-feel group. Then the participants will be introduced to their agents. After the agents' identities are disclosed, participants are prompted to indicate their level of anger, with emotion-express group specifying their *ExpressedAnger* and emotion-feel group reporting *FeltAnger*. Subsequently, the agent offers a compensatory solution of 5 CNY. Participants must then decide whether to accept this offer. Those who accept will receive the 5 CNY and conclude the experiment. Participants who decline the initial offer proceed to the second round (hereafter referred to as R2, with the first round being R1), where they are asked to indicate their expressed or true anger level again. In this round, the agent proposes a higher compensation of 8 CNY. Participants receive the 8 CNY and then exit the experiment.

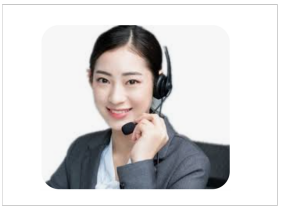
Our study encompasses 1,426 participants, all of whom participate in the experimental game in exchange for monetary rewards that reflect their negotiation outcomes with the agents. Out of this initial cohort, 149 participants fail the manipulation test, while another 67 do not pass the attention check. This leads us to focus on the remaining 1,210 valid participants. The final results consist of 595 participants in the emotion-express group and 615 in the emotion-feel group. Table A.4 and A.5 in Appendix A.3.1 provide a summary of participant numbers engaged in both R1 and R2, categorized by different groups.

Study 2 adopts a conversational context. Participants choose from nine expressions that increase in levels of expressed anger, detailed in Appendix A.4.1.

[10]The chat scripts used in Study 2 are detailed in Appendix A.3.3 Figure A.3.

Figure 3.4: Agents' Profile Photos and Titles Displayed in Study 2

(a) Agent Xiaomei



(b) Human Agent Xiaomei

(c) AI Agent Xiaomei

(d) GPT-based AI Agent Xiaomei

3.4.2 Model Free Evidence

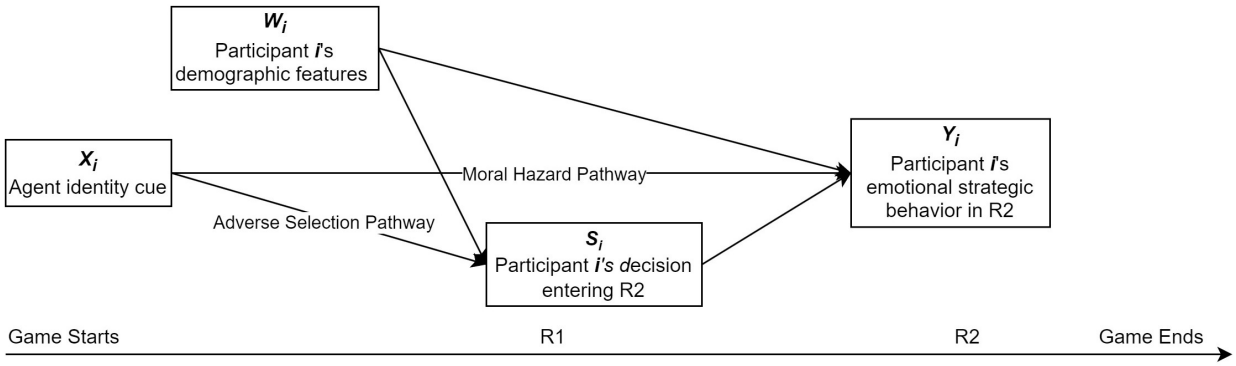
To explore the impact of an agent's identity on participants' strategic behavior, we compared their *ExpressedAnger* and *FeltAnger* in the emotion-express and emotion-feel groups across the four identity conditions. In R1, no significant strategic behavior was observed. Interestingly, R2 revealed pronounced strategic behavior among participants in the non-disclosure group. Figure A.4 in Appendix A.5.1 displays the anger levels of participants in R2. No significant difference between *ExpressedAnger* and *FeltAnger* is observed when the agent's identities are human, AI, or GPT-based AI. Nevertheless, in cases where the agent's identity is undisclosed, participants' expressed anger is noticeably higher than what they genuinely felt. This preliminary result is consistent with our findings in Study 1 and confirms Hypothesis 1a and 1b, demonstrating that such emotional strategic behavior under non-disclosure identity cues still exists in a real-life multi-round setting. In the subsequent sections, we will focus on identifying these effects and unveiling the underlying mechanisms that drive them.

3.4.3 Identification

Due to this multi-round setting, we postulate two potential pathways as shown in Figure 3.5. The first is the moral hazard pathway. Here, the non-disclosure of an agent's identity can trigger moral hazard, which might incentivize participants to behave immorally in pursuit

of higher rewards. This ambiguous identity cue makes the identity categorization process hard for participants, which in turn, triggers greater negative effects. The second pathway is adverse selection. We hypothesize that non-disclosure tends to select participants with lower moral standards, lower altruism, and more complaint experience for R2. The culmination of these effects leads to the strategic behaviors we observe in R2.

Figure 3.5: Causal Graph and Identification



Adverse Selection Pathway

In this section, we aim to investigate the adverse selection pathway first. We hypothesize that participants' moral level, altruism and familiarity with online complaints will influence their strategic behavior. Therefore, we measure participants' *Altruism*, *Amorality*, and *ComplaintFrequency* at the end of the experiment. *Altruism* is measured using the proportion of money that participants are willing to donate if they win a lottery (Falk et al. 2023). *Amorality* is measured using five items following Dahling et al. (2009) (e.g., "I am willing to be unethical if I believe it will help me succeed."). Then we also collect participants' complaint frequency in the last 6 months. All measurements are listed in Table A.6 Appendix A.4.2. The variables *Amorality* and *Altruism* offer direct or indirect insights into participants' moral standards. Typically, individuals with lower moral standards are associated with lower levels of altruism and higher amorality. *ComplaintFrequency* serves as an

indicator of the participant's level of experience, with a higher frequency suggesting greater experience with the complaint process. Our Hypothesis 3 posits that non-disclosure tends to select participants with higher *Amorality*, lower *Altruism*, and a higher *ComplaintFrequency* to R2. Additionally, we suggest that participants exhibiting such traits are more inclined towards strategic behavior.

We compare participants who opt to continue to R2 with those who choose to exit the game after R1 across four emotion-expression groups (hereafter referred to as R2 participants and R1 participants, respectively). As depicted in Figure A.5 in Appendix A.5.2, we compare the aforementioned three variables between R1 participants and R2 participants. Not surprisingly, our observations suggest that the non-disclosure condition significantly influences participant behavior through an adverse selection effect.

To further testify the observations from Figure A.5, we conduct a one-way ANOVA using the R1/R2 participants as the between-subjects factor. The results indicate that non-disclosure identity cue leads to the selection of participants for R2 who have higher *ComplaintFrequency* ($M_{R1} = 1.84$ vs. $M_{R2} = 2.39$, $SDs = 1.05$ and 1.12, $F(134) = 8.166$, $p = .005$), higher *Amorality* ($M_{R1} = 2.81$ vs. $M_{R2} = 3.50$, $SDs = 1.40$ and 1.53, $F(134) = 7.086$, $p = .009$) and lower *Altruism* ($M_{R1} = 0.26$ vs. $M_{R2} = 0.17$, $SDs = 0.23$ and 0.14, $F(134) = 8.385$, $p = .004$).

As a comparison, we do not observe significant differences in *Amorality* in other three groups with AI agent ($M_{R1} = 3.54$ vs. $M_{R2} = 3.09$, $SDs = 1.66$ and 1.64, $F(157) = 2.873$, $p = .092$), GPT-based AI agent ($M_{R1} = 3.50$ vs. $M_{R2} = 3.21$, $SDs = 1.72$ and 1.47, $F(151) = 1.207$, $p = .273$) and human agent ($M_{R1} = 3.21$ vs. $M_{R2} = 3.24$, $SDs = 1.79$ and 1.67, $F(149) = 0.006$, $p = .936$). Similarly, no significant difference in *ComplaintFrequency* is found for the AI ($M_{R1} = 2.15$ vs. $M_{R2} = 2.34$, $SDs = 1.12$ and 1.06, $F(157) = 1.192$, $p = .277$), GPT-based AI agent ($M_{R1} = 2.15$ vs. $M_{R2} = 2.14$, $SDs = 1.07$ and 0.89, $F(151) = 0.014$, $p = .905$) and human agent groups ($M_{R1} = 2.05$ vs. $M_{R2} = 2.36$, $SDs = 1.09$ and 1.13, $F(149) = 2.997$, $p = .085$). In terms of *Altruism*, we still find no significant difference among AI agent ($M_{R1} = 0.24$ vs. $M_{R2} = 0.20$, $SDs = 0.18$ and 0.17, $F(157) = 1.904$, $p = .170$) and

GPT-based AI agent ($M_{R1} = 0.24$ vs. $M_{R2} = 0.20$, $SDs = 0.22$ and 0.18, $F(151) = 1.797$, $p = .182$), but we observe significant divergence in human agent group ($M_{R1} = 0.34$ vs. $M_{R2} = 0.21$, $SDs = 0.28$ and 0.19, $F(149) = 11.200$, $p = .001$). In conclusion, we find that non-disclosure agents tend to adversely select participants with higher amorality, more complaint experience, and lower altruism for R2. Hence, these results confirm Hypothesis 3. This subset of participants is more prone to strategic behavior, thereby influencing the overall results of the experiment.

Moral Hazard Pathway

In the previous discussion, we considered the adverse selection pathway, where there exists an adverse selection under the non-disclosure identity cue. Beyond this pathway, it's plausible that non-disclosure may also introduce a moral hazard, either directly or through mediators, prompting strategic behaviors in participants. To identify this pathway, we must overcome two challenges. First, similar to Study 1, it's crucial to determine the degree of strategic misrepresentation, specifically *StrategicAnger*, which serves as our outcome variable. In alignment with Study 1, we use matching techniques and the difference between *ExpressedAnger* and the matched *FeltAnger* is *StrategicAnger*. Secondly, given the two-round structure of our experiment, the strategic behavior in R2 can only be observed if a participant opts to enter R2. Therefore, the estimated effects of agent identity on strategic behaviors are weight-averaged over the subset of participants who continue to R2. As discussed in Section 3.4.3, the identity of the agents can affect the kind of participants who proceed to R2. Ideally, we would measure the strategic behavior in R2 for all participants, but such a counterfactual approach is not feasible. Therefore, to probe the potential moral hazard pathway, we must rectify any skew introduced by this selection. To achieve this, we employ a Heckman-style two-stage model (Heckman 1979), which allows us to estimate the weighted average treatment effects across all participants in our dataset.

To isolate the strategic emotional expression *StrategicAnger*, we match R2 participants from the emotion-express group with those from the emotion-feel group. This matching

is performed separately for each of the four different agent identity groups. We calculate a propensity score for each user based on the demographic variables displayed in Table A.2 in Appendix A.2. We then pair each participant in the emotion-express group with the participant who has the closest propensity score in the emotion-feel group. From this process, we obtain an estimated *FeltAnger* for the participants in the emotion-express group. The difference between *ExpressedAnger* and *FeltAnger* is what we define as *StrategicAnger*. In order to verify the quality of our matching, we perform a balance check on the covariates in the matched data, as displayed in Table A.2 in Appendix A.2. Similar to Study 1, the standardized mean differences are all close to zero and the variance ratios are approximately one, suggesting that our matching has effectively created well-balanced groups.

Using *StrategicAnger* as our outcome variable, we specify our Heckman-style two-stage model as follows:

$$\text{Selection Equation: } S_i = \mathbf{1}_{S_i^*>0}, where\ S_i^* = \gamma W_{2i} + \eta_i, \tag{3.1}$$

$$\text{Outcome Equation: } Y_i = \beta_0 + \beta_1 T_i + \beta_2 W_{1i} + \epsilon_i, \tag{3.2}$$

$$\begin{pmatrix} \eta_i \\ \epsilon_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix} \right).$$

In the first stage of the model, we use a probit model to estimate the selection equation, which models the probability of participants continuing to R2 (S_i) as a function of W_{2i}. In the second stage, we employ a linear model to estimate the outcome equation here. The linear model constructs the strategic behaviors as a function of the agent identities, controlling for other covariates, as well as the inverse Mills ratio (ρ) obtained from the first-stage estimation. The inverse Mills ratio serves as a control for self-selection and helps us to correct the bias brought by selection.

Table 3.2 presents our results. In column (1), *StrategicAnger* is the outcome variable and non-disclosure serves as the baseline group. All three comparison groups have negative and significant coefficients compared to the non-disclosure identity cue. Specifically,

Table 3.2: Heckman Selection Model: Strategic Expressed Emotion and Satisfaction

	(1) StrategicAnger	(2) Satisfaction
GPT-based AI	-1.45***	0.52***
	(0.47)	(0.18)
AI	-1.97***	0.19
	(0.44)	(0.18)
Human	-1.08**	0.60***
	(0.42)	(0.18)
ρ	-0.64	-1.07***
	(0.70)	(0.24)
Amorality	-0.92	-0.11
	(0.60)	(0.25)
Alruism	-0.66	0.90***
	(1.13)	(0.32)
ComplaintFrequency	0.51	-0.35
	(0.70)	(0.27)
Constant	2.50***	5.23***
	(0.88)	(0.31)
Observations	328	595
R^2	0.060	0.120

Robust standard errors are reported in parentheses.

* p<0.10, ** p<0.05, *** p<0.01

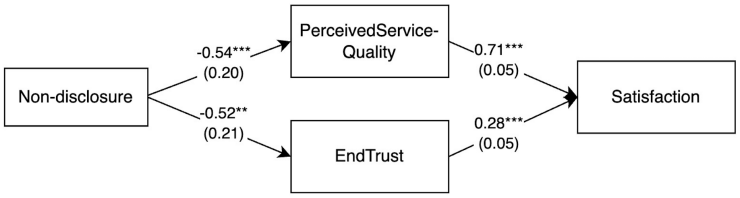
This table reports the treatment effects of agent identity cues on customers' strategic emotion expression in R2 and customers' satisfaction with the agents. The baseline group is Non-disclosure.

StrategicAnger decreases by 1.08 when participants interact with a human agent and by 1.97 when engaging with an AI agent. The effect of the GPT-based AI identity cue, which reduces strategic behavior by 1.45, falls between that of AI and human agents. GPT-based AI agents use advanced text models with superior emotion understanding compared to the average AI agents. However, their understanding is not as nuanced as that of a human, making it logical to observe its intermediate effect. Taken together, these findings demonstrate that all three comparison groups introduce less strategic behavior than the non-disclosure group. These outcomes provide further support to Hypotheses 1b and 1a, reinforcing the notion that the non-disclosure identity cue catalyzes greater strategic behavior in customers both through adverse selection and moral hazard pathways.

However, strategic behavior is not the only factor of interest to business owners; the satisfaction rate is also of great importance. As Vasquez Sampere (2015) points out, many companies emphasize customer service, in return for higher customer satisfaction. Therefore, we construct a similar Heckman-style model as Equation 3.1 and 3.2, with the outcome variable Y_i in Equation 3.2 switched to *Satisfaction*. We follow Han et al. (2022) to measure the *Satisfaction*, using three items rated on a seven-point scale (e.g., "Overall, how satisfied or dissatisfied did your experience with the agent leave you feeling?"). The results are presented in Table 3.2 column (2), with non-disclosure remaining the reference group. Interestingly, there is no significant difference between the AI and non-disclosure groups, as indicated by AI's insignificant coefficient. As a comparison, GPT-based AI enhances the *Satisfaction* by 0.52, while the human identity cue increases it by 0.60. Being a more advanced language model, it outperforms both the AI and non-disclosure groups, nearing the performance of human agents.

This result advances the current understanding by showing that disclosing an AI's identity does not necessarily lead to poorer performance, especially in a customer care context. The non-disclosure of an AI's identity may even trigger more strategic behavior from users. Additionally, in our study, non-disclosure receives similarly low satisfaction evaluations as those for AI identities that are disclosed. Such underperformance of AI can be mitigated by

Figure 3.6: Mediation effect: Perceived Service Quality



employing more advanced algorithms, such as GPT-based models.

3.4.4 Mediation

In this section, we delve into the proposed mediation effect of trust and perceived quality of service on satisfaction proposed in Hypothesis 4. Firstly, as shown in Figure 3.6, we explore the mediation effect of *PerceivedServiceQuality* and *EndTrust* on satisfaction. We measure participants' trust in the agent twice during the experiment: initially after the disclosure of the agent's identity, denoted as *StartTrust*, and subsequently after the full negotiation process, referred to as *EndTrust*. *PerceivedServiceQuality* is measured using three items following Han et al. (2022) (e.g., "poor/excellent"). *Trust* is measured using four items following Sirdeshmukh et al. (2002) (e.g., "very undependable/very dependable"). We employ PROCESS Model 4 (parallel mediation model) with ρ, *Altruism*, *Amorality*, and *ComplaintFrequency* as covariates and a bootstrapped sample of 10,000 (Hayes 2017). Results revealed that the total effect is $(\beta = -.50, t(322) = -2.39, p = .017)$. We observe no significant direct effect $(\beta = .03, t(322) = .46, p = .649)$. These results indicate that satisfaction is entirely mediated by *PerceivedServiceQuality* $(\beta = -.39, SE = .16, 95\%, CI = [-.70, -.08])$ and *EndTrust* on agent $(\beta = -.15, SE = .07, 95\%, CI = [-.30, -.03])$. As illustrated in Figure 3.6, the use of non-disclosure identity cue decreases participants' trust and perceived service quality in customer service agents, which subsequently reduces their satisfaction with the agents.

3.4.5 Policy Tree

In the preceding section, we showed that the identity of customer service agents significantly influences customers' strategic behavior and satisfaction levels, with substantial implications for business-customer interactions. In practice, companies often face constraints like budget limits, regulatory compliance, and the need for streamlined operations. Furthermore, companies typically gather observational data from users, which can then be utilized to learn treatment assignment rules. Consequently, it is important to develop an effective approach for agent assignment that takes into consideration these constraints and the available observed data. Such a methodology will aim to minimize strategic behavior and enhance customer satisfaction.

We adopt a semiparametrically efficient policy learning method, the Policy Tree (Athey and Wager 2021, Zhou et al. 2023). The main focus is to utilize observational data to develop a policy, represented by π, within a given policy class Π specified by practitioners, which encompasses various problem-specific constraints. The learned policy aims to make assignment decisions based on a participant's characteristics denoted by $X_i \in \mathcal{X}$. This method involves fitting a depth k tree through exhaustive search to identify the optimal tree, which maximizes the sum of rewards. The optimal treatment assignment rule is determined by solving the following:

$$\pi^* = \arg\max \left\{ \frac{1}{n} \sum_{i=1}^{n} \Gamma_i(\pi(X_i)) : \pi \in \Pi \right\}.$$

Here, $\Gamma_i \in \mathbb{R}^a$ represents a vector of rewards for each action $a \in A$, where in our case, the actions correspond to the identities of the four different agents. Athey and Wager (2021) propose to use doubly robust estimator of average treatment effect. By doing so, they establish guarantees regarding the utilitarian regret of deploying a policy, denoted as:

$$R(\pi) = \max \{\mathbb{E}[Y_i(\tilde{\pi}(X_i)] : \tilde{\pi} \in \Pi\} - \mathbb{E}[Y_i(\pi(X_i)].$$

This regret metric quantifies the disparity between the expected utility achieved by deploying the learned policy π over a target population and the maximum attainable utility that can

be achieved by deploying any policy within the class Π over the same population. The Policy Tree method is proven to satisfy regret bounds that diminish at a rate of $1/\sqrt{n}$ and to estimate the globally optimal policy assignment efficiently.

We construct the following Policy Trees, visualized in Figure 3.7, with the goal of minimizing strategic emotion expression. To determine the appropriate tree depth and evaluate the accuracy of the policy learning procedure, we employ cross-validation. The data is randomly divided into K folds $S_k, k = 1, \ldots, K$, and for each fold, we train a policy $\hat{\pi}^{(-k)}$ using all data except for the instances in S_k. Here, we set the number of folds K to 10. Next, we proceed to estimate the rewards using the following formula:

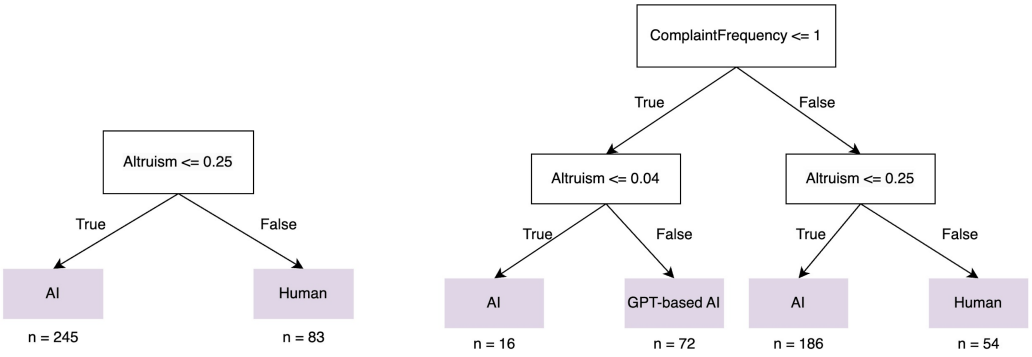
$$U = \frac{1}{n} \sum_{k=1}^{K} \sum_{i \in S_k} \hat{\Gamma}_i(\hat{\pi}^{(-k)}(X_i)).$$

The estimated reward for a depth-1 tree is $-0.873(\pm 0.338)$, and for a depth-2 tree, it is $-1.000(\pm 0.387)$.[11] Notably, we observe that the depth-1 tree achieves a higher reward. The policy assignment aligns with insights derived from prior analyses, suggesting that to mitigate strategic behavior among participants, they should not be assigned to agents with non-disclosure identity cues. Both the depth-1 and depth-2 decision trees identify the same *Altruism* threshold for participants who have filed at least two complaints in the past 6 months before the experiment. For participants with fewer than two complaints, the depth-2 tree specifies a distinct *Altruism* threshold. Nonetheless, the recommendation consistently suggests that individuals with lower Altruism scores should interact with AI, while those with higher scores should be assisted by human agents. The only exception is in the depth-2 decision tree for individuals with less experience in filing complaints: those with high *Altruism* scores benefit more from interacting with GPT-based AI than with human agents. This observation is logical, considering GPT-based AI's advanced emotion detection capabilities, which approach human levels. For less experienced users, GPT-based AI might not be perceived strictly as a chatbot, enabling it to seamlessly substitute for human agents.

[11] For computational simplicity, we consider the negative value of *StrategicAnger* as the outcome, transforming the problem from minimizing to maximizing the outcome. The increase in reward corresponds to a reduction in strategic emotion expression. Standard errors are indicated in parentheses

However, in a realistic setting, it is argued that such features as *Altruism* are hard to observe. Therefore, we also constructed Policy Trees using the demographic features we collected. The results are shown in Appendix A.5.3. For both depth-1 and depth-2 trees, the results indicate that individuals should not be served by non-disclosure agents. We observe that gender is an important division variable, suggesting that males should be served by human agents, while females should be served by AI agents. Furthermore, in our depth-2 tree, we find that if a female individual has a service-related job, she should be served by a human agent. This finding aligns with previous research on gender differences in emotional expression and regulation. Studies have shown that women tend to be more emotionally expressive and are more likely to engage in emotional disclosure compared to men Brody and Hall (2008). By assigning female customers to AI agents, it could potentially reduce emotional misrepresentation, as women may be less likely to exaggerate their emotions when interacting with an AI agent that is perceived to have limited emotional understanding.

Figure 3.7: Agents Assignment Using Policy Tree



(a) Depth-1 Policy Tree (b) Depth-2 Policy Tree

ComplaintFrequency can take values from 1 to 5, representing the number of complaints in the past 6 months. Altruism can take values from 0 to 1, with an average of 0.20.

3.5 Conclusion

Our study provides several intriguing insights and discussions surrounding the role of non-disclosure identity cues in strategic customer behavior and satisfaction. Contrary to the proposition made by Luo et al. (2019), Shi et al. (2020), Schanke et al. (2021), Han et al. (2022), Cui et al. (2022), our findings reveal that non-disclosure of agents' identities results in inferior performance compared to both human agents and AI agents in the after-sale context by triggering larger strategic behavior from customers. In our first experiment, in which we adapt the DG and UG design, we demonstrate the moderated mediation impact of psychological distance on strategic behavior, influenced by the varying frequencies of participants' online shopping. This effect persists in a multi-round context that more closely mirrors real-life scenarios. In our second study, we further demonstrate the existence of strategic behavior under non-disclosure identity cues. Then we identify and provide evidence for two distinct mechanisms that lead to this phenomenon. We reveal that, beyond the direct pathway of moral hazard, there exists a pathway of adverse selection within the context of repeated interactions.

In our second experiment, we also investigate the impact of varying identity cues on customer satisfaction levels. Our findings indicate that non-disclosure and chatbot identity cues lead to lower satisfaction when compared to human or GPT-based AI identity cues. We find the mediation effect of trust and perceived service quality on satisfaction rating. This raises important considerations for businesses and organizations in how they present their agents, especially in online contexts. This study alleviates, to an extent, the concerns surrounding the legal mandate to disclose AI agents' identity enforced in certain regions. Moreover, our findings also present a practical solution for businesses trying to balance automation needs and customer satisfaction - by using advanced chatbots like ChatGPT. These advanced chatbots, with their superior language processing abilities, can meet both the demands of the law and high-quality customer interactions, thus ensuring high levels of customer satisfaction.

3.5.1 Theoretical Contributions

We contribute to the growing literature on emotion AI in customer care by examining the effect of identity cues on customer behavior and satisfaction (Huang and Rust 2023). Previous research in pre-sales customer service settings has found that disclosing an agent's AI identity can negatively impact metrics like purchase conversion. However, we show that in the post-sales customer care context, where the focus is on emotional care rather than information gathering, non-disclosure of agent identity can actually lead to greater strategic behavior by customers compared to AI identity disclosure. This advances our understanding of how the impact of AI transparency can vary based on the stage and purpose of the customer service interaction.

Furthermore, we are among the first to empirically demonstrate the nuanced effects of identity cues on spurring emotional misrepresentation, a form of strategic behavior that has primarily been studied analytically in the literature (Frankel and Kartik 2019, Yu et al. 2022, Wang et al. 2023). By examining actual customer behavior across interactions with agents presenting four distinct identity cues (AI, GPT-based AI, human, non-disclosed), we show that ambiguous agent identity can encourage customers to exaggerate their emotions for strategic gain. This builds upon prior research on emotional gaming behavior in human-human negotiations (Andrade and Ho 2009).

Additionally, we bridge social identity theory with the customer care literature by highlighting how identity ambiguity can lead to an "illegitimacy discount" (Zuckerman 1999). When agent identity is undisclosed, customers perceive greater psychological distance, enabling them to behave more strategically without the constraints of clear social norms. In contrast, a human identity cue reduces this distance and encourages more authentic emotional expression.

By uncovering the mechanisms through which non-disclosure impacts strategic behavior, both directly and indirectly via adverse selection, our findings challenge assumptions that AI transparency is always detrimental to interaction outcomes (Luo et al. 2019, Shi et al.

2020, Schanke et al. 2021, Han et al. 2022, Cui et al. 2022). We also provide valuable insights for the emotion AI field regarding how even highly sophisticated language models like GPT may be perceived as having imperfect emotional understanding, tempering strategic behavior compared to ambiguous agent identity.

In summary, we advance knowledge on the role of identity cues and emotional misrepresentation in AI-mediated customer care. Our empirical investigation of how agent identity ambiguity can intensify strategic behavior and diminish consumer satisfaction offers important theoretical and practical contributions to the study of AI-human interactions in service contexts.

3.5.2 Practical Contributions

This study provides several implications for policymakers, companies, and researchers, highlighting the interplay between AI identity disclosure, strategic consumer behavior, and satisfaction in AI-human interactions. From a policymaker perspective, the evidence supports the need for transparency in the deployment of AI in customer service. This can guide the formulation of regulations that mandate the disclosure of AI's role, potentially enhancing customer trust and overall satisfaction with AI-enabled services. The results presented in our paper can alleviate the concerns about regulations on AI usage disclosure. Knowing that AI disclosure will greatly reduce strategic behavior will help companies gain confidence and make sure of the smooth launch of such regulations.

For companies, our study provides important guidelines for personalizing agents to customers based on customer characteristics and past behavior. Utilizing Policy Tree analysis, companies can optimize agent assignment to mitigate strategic behavior and enhance satisfaction, leveraging AI effectively while managing resource allocation. In light of these findings, businesses should consider developing data-driven strategies that utilize customer history, preferences, and behaviors to tailor interactions. Customization and personalization in AI deployment can significantly improve service outcomes and mitigate customers' strategic behavior, marking a critical area for future development and research. Moreover,

the distinction between traditional AI and GPT-based AI agents in terms of impacting customer satisfaction signals the potential benefits of investing in advanced AI models. Such technologies, including large language models, could significantly improve the empathy and emotion detection capabilities of AI agents, bridging the performance gap between AI and human agents. The mediation analysis further highlights trust and perceived service quality as critical factors influencing customer satisfaction. This insight indicates that enhancing these factors should be a strategic priority, potentially through improving the reliability of AI interactions and transparent communication about AI capabilities and limitations.

Additionally, our research paves the way for further investigation, especially into how the capabilities of advanced AI can enhance customer service interactions. The impact of customer service on strategic behavior may be divided into two categories: the identity effect and the content effect. Our study primarily examines the identity effect of customer service identity cues. Future research should delve into the content of customer service—that is, how to design advanced AI systems to minimize such strategic behavior. Exploring the specific attributes of advanced AI that result in reduced strategic behavior and lead to increased satisfaction and trust could offer critical guidance for AI development. Moreover, this study encourages further examination of the behavioral economics of AI-human interactions, focusing on the roles of adverse selection and moral hazard in various contexts and service domains.

In summary, this study contributes to the growing body of literature on AI-human interactions by providing empirical evidence on the consequences of non-disclosure in AI and human agent's identity and its impact on strategic consumer behavior and satisfaction. These findings inevitably have profound implications for the role of AI in customer service. While AI has undoubtedly transformed the customer service landscape, our results emphasize that a mere technological solution may not suffice. AI should be paired with thoughtful strategies about transparency to ensure customers' trust and satisfaction. In conclusion, the practical implications underscore the importance of transparency, the potential of advanced AI, and the need for personalized service strategies in enhancing customer experience and business

outcomes.

Chapter 4

ONE SIZE FITS ALL? INFORMATIONAL ACCESSIBILITY AND INCLUSIVITY IN ONLINE PLATFORMS

4.1 Introduction

In the last decade, popular media, ad campaigns, and activism have played a crucial role in bringing size inclusion (and the lack thereof) to the forefront of e-commerce. The fashion industry has long been criticized for insufficient inclusivity in terms of body sizes (Lewis 2019). According to the market research firm NPD, while the sale of plus-sized women's apparel was over 21 billion USD in 2016 and contributed to 6% of the overall apparel market, plus-sized brands represented only 2.3% of the total number of brands carried by the top multi-brand retailers [1]. This is not due to the lack of demand—according to a report by the Centers for Disease Control and Prevention National Center for Health Statistics, the average woman wears a size 14 or greater.[2] To address the lack of diversity in body size, an increasing number of brands are now expanding their product lines to be more inclusive.[3]

The increase in offerings from online retailers is not complemented by an increase in available information for plus-sized consumers online. On the one hand, most of the photos of models provided by retailers reflect the fit for small-sized users, perpetuating an industry standard that does not reflect the diversity of its customer base. According to Maguire et al. (2023), of the 219 fashion shows in 2023 around the world, only 0.6% of the models were plus-sized. Along similar lines, user-generated content, a crucial resource for many con-

[1] https://www.cbsnews.com/news/the-fashion-industry-thinks-big-about-plus-sizes/

[2] https://www.cdc.gov/nchs/fastats/body-measurements.htm, https://www.cdc.gov/nchs/data/series/sr_03/sr03-046-508.pdf, last accessed April 25, 2024.

[3] https://www.pressherald.com/2021/08/18/old-navy-tries-to-normalize-plus-size-apparel-in-growth-push/, last accessed April 25, 2024.

sumers, is similarly lacking in its representation of plus-sized users. Despite the enormous purchasing power wielded by plus-sized consumers, they find themselves navigating a digital landscape marked by informational scarcity. The lack of plus-sized user-generated reviews exacerbates information asymmetry, making it exceedingly challenging for plus-sized consumers to make informed purchasing decisions. This asymmetry is particularly problematic, considering the importance of online reviews and user-generated content in shaping consumer choices. Therefore, our study seeks to probe these disparities, exploring the impact of the lack of representative information on consumer behavior and the mechanisms that drive the information asymmetry.

Information asymmetry on e-commerce platforms can significantly decrease customer satisfaction, increase returns, and create other detrimental outcomes for platforms. Online reviews play a crucial role in alleviating this information asymmetry, helping consumers make better decisions (Chatterjee 2001, Chen and Xie 2005, Chevalier and Mayzlin 2006, Forman et al. 2008, Moe and Schweidel 2012, Gao et al. 2015, Hu et al. 2006, Archak et al. 2011, Zervas et al. 2021, Hu et al. 2009b). However, studies in IS literature show that review platforms suffer from significant bias in online reviews because of acquisition bias (a consumer who buys a product is more likely to review the product) and underreporting bias (only consumers with extreme opinions are more likely to leave a review), leading to a "J-shaped" review curve (Hu et al. 2006, 2017, 2009a, Gao et al. 2015, Zervas et al. 2021). Interestingly, this literature has entirely focused on the *experiential* aspect of consumption and the decision to write reviews, but not on the user-level aspects that might prevent them from writing reviews.

Writing reviews is a significant act of altruism—one where consumers strive to inform other consumers about product quality by expending considerable energy and time (Hennig-Thurau et al. 2004, Zhang and Zhu 2011, Qiao et al. 2020). When consumers write reviews, they cede some level of privacy by expounding on their experience, both positive and negative. A significant imbalance in the reviews can appear when privacy costs are high, an aspect unstudied in the current literature. In particular, privacy costs might be significantly high

among consumers of certain body types due to societal stigma and lived experiences of anti-fat bias (Lin and Reid 2009, Bissell and Hays 2010, Teachman et al. 2003, Wang et al. 2004). This leads to a situation where plus-sized consumers face more informational uncertainty when purchasing apparel because other plus-sized users face a higher privacy cost in writing reviews, exacerbating the cycle of informational asymmetry.

In this study, we first empirically document the informational asymmetry in UGC—how plus-sized users disproportionately post fewer photos in their reviews compared to other users. Next, we show how reviews, particularly reviews with photos from similar-sized users, play a key role in influencing purchase decisions. We find that the impact of photos from similar-sized users is salient among plus-sized users. This is because they already experience significant informational asymmetry through models that are drastically different and UGC that has a significant skew. We delve into the mechanisms that might influence the imbalance in UGC.

To do so, we leverage large-scale, novel, proprietary consumer-rental-level data from an apparel rental platform to answer these questions empirically. This platform operates on a subscription model, and users can rent clothes for a flat fee, similar to streaming platforms like Netflix. The platform requires the customers to provide star ratings for a product after the rental, but the decision to write reviews or add photos is completely optional, similar to other e-commerce platforms.

Users of this platform are subscribers and, thus repeat customers. Unlike other papers in the UGC literature, we are uniquely positioned to observe the content *generation*, *pre-consumption*, and *post-consumption* processes. Further, the platform collects information about the users' size and preferences when they sign up. Observational data obtained from e-commerce platforms usually have an issue of selection bias, where researchers cannot observe the consumers who *do not* post reviews. However, unlike other papers in this literature, we can precisely identify consumers who rent a product and choose not to review the product. We also observe the demand data at the rental-level. This helps us to understand the impact of reviews even for users who do not rent the product. Our data allows us to delve deeper

to discover the mechanism of why consumers write reviews and the conditions that can lead to a significant imbalance in reviews along size lines. We develop a two-stage Probit model to uncover the potential factors that can affect users' decisions to post review photos. In our model, we build a hierarchical decision structure to capture the decision-making process of a user in posting review photos in the post-consumption process. Specifically, we identify the privacy cost by examining how a user's decision to disclose identity-related information varies when posting a review photo.

Our findings reveal that there is a significant under-representation of reviews from plus-sized users. Interestingly, when plus-sized users leave reviews, they are less likely to leave reviews with photos (henceforth referred to as photo reviews). However, we find photo reviews from plus-sized users play a significant role in increasing demand from other similar-sized users, utilizing a Difference-in-Differences design. Specifically, a review photo posted by a plus-sized user can increase the demand for a product by 10% among other plus-sized users. This finding is further confirmed by a click-stream analysis, which shows that photo reviews from similar-sized users significantly boost conversions from their peers, with the most pronounced impact observed among plus-sized users.

Results from our hierarchical model suggest that reviews from other users tend to discourage focal users from posting their review photos, as they reduce the altruistic motivation of a user. In other words, users are less likely post review photos to provide information if this information is already available. Conversely, the presence of review photos, particularly those from users of similar sizes, appears to encourage others to share review photos. This suggests that review photos, unlike text-only reviews, do not diminish the altruistic motivation to contribute. This is likely because when a group of users faces social stigma in everyday life, seeing other users from the same group on a platform creates a feeling of solidarity. Crucially, the model identifies that plus-sized users face the highest privacy costs, which deters their participation in online communities. This contributes to the issue of underrepresenting plus-sized groups in these spaces.

The implications of our study are far-reaching. We are the first in the literature to demon-

strate how consumers who routinely face social stigma can be less motivated to contribute to the review ecosystem. We find that consumers who face social prejudice have increased privacy costs when deciding to write reviews. We find that reviews from similar-sized users are particularly important in the context of apparel, especially for plus-sized users. The increase in privacy cost for plus-sized users to write reviews can, in turn, lead to significant negative externality for other plus-sized consumers on the platform — they have much less information to go by when making purchase decisions, especially since models provided by the platform are vastly different from the focal user.

Most platforms aim to increase user-generated content by trying to appeal to the altruism of the users (Burtch et al. 2018b, Goes et al. 2016, Gallus 2017, Huang et al. 2019, Khern-am nuai et al. 2018, Cabral and Li 2015). However, altruism is only part of the motivation to write reviews. Our results highlight that privacy can significantly contribute to the review ecosystem, an important and yet understudied aspect of UGC. By understanding the barriers inhibiting users with high privacy costs from contributing to photo reviews, digital platforms can design features to ameliorate privacy concerns. Platforms can also aim to increase information that is not user-generated. If a subset of users makes suboptimal choices due to reduced information, platforms should be explicitly inclusive in product descriptions. This can be accomplished by increasing the diversity of models. In domains with high privacy costs, increasing the diversity of models is not just an act of virtue signaling. Instead, it also empowers consumers with crucial information that can influence purchasing behavior, especially when such information is rare in UGC due to privacy concerns.

Our results also highlight the different motivations behind the users' choice of the modality of reviews on a UGC platform. We find that users who are more likely to face social stigma about their appearance in real life are less likely to participate in UGC with images. However, we find that the motivation to participate in textual reviews is dependent on the existing informational void. In the case of textual reviews, the decision to review is determined by the informational void that the users seek to fill altruistically. For example, in our context, we find users are less likely to review a product when there are other textual reviews

about the product from users of the same size. This applies to users of all sizes. However, this result is significantly different when it comes to photo reviews. Our results show that plus-sized users are more likely to post photos only if there are photos of other plus-sized users. This highlights how the modality plays a key role in the decision to participate in UGC.

UGC platforms built around identity-revealing modalities, such as images and videos, are less likely to include users who face routine stigmatization, such as plus-sized users. Interestingly, our model reveals that the participation of other plus-sized users induces more participation from the focal user. When significant privacy costs are involved, platforms should seek to provide features that can preserve the anonymity of the users. Platforms can engage plus-sized influencers, who then are likely to inspire other plus-sized users to create image or video UGC.

The remainder of the paper is structured as follows. We develop our hypotheses in Section 4.2, then we describe the research context, data, and variables in Section 4.3. Section 4.4 describes our model specifications and presents the main results. We explore mechanisms in Section 4.5 and provide several robustness checks in Section 4.6. Section 4.7 concludes the paper and discuss implications.

4.2 Hypotheses Development

Plus-sized individuals may experience marginalization due to prevailing beauty standards that glorify specific body types and vilify others. Link and Phelan (2001) elaborate on such stigma, emphasizing its role in creating inequalities through processes of labeling, stereotyping, separation, status loss, and discrimination. Applying this to the realm of UGC, plus-sized users may be more hesitant to contribute user-generated photos due to fear of negative evaluation, social rejection, or online harassment (Puhl and King 2013). Fardouly et al. (2015) and Tiggemann and Slater (2013) have explored the impact of social media on body image concerns, suggesting that engagement with image-centric platforms can exacerbate self-objectification, body surveillance, and the internalization of thin ideals. For plus-sized

users, the heightened visibility and focus on physical appearance inherent in UGC, especially photographs, may deter their active participation due to the anticipated or experienced body inequality within these platforms.

We integrate the concepts of social stigma and body inequality to explain the disparity of plus-sized users in UGC contributions. This disparity is not merely a matter of individual choice or preference but is deeply embedded in the structural and societal dynamics that govern digital spaces. Plus-sized individuals may navigate these spaces with an acute awareness of the potential for stigma and discrimination, which in turn influences their willingness to contribute content, especially content that is personally revealing, such as photos. Based on this, we hypothesize the following,

Hypothesis 5. *Plus-sized users contribute less to UGC, especially photo reviews, compared to medium- or small-sized users.*

Information asymmetry can lead to market inefficiencies, where consumers, suffering from making a poor decision, may be hesitant to engage in transactions (George et al. 1970). User-generated content (UGC), especially user-generated photos, can significantly reduce information asymmetry by providing potential buyers with more accurate and relatable product representations. This is particularly relevant for the fashion industry, where fit, size, and appearance can vary widely and are crucial factors in the purchasing decision. User-generated photos offer a form of social proof and a realistic portrayal of how products look on different body types, thus mitigating the risk perceived by consumers (Chevalier and Mayzlin 2006).

In the context of plus-sized products, the availability of user-generated photos from plus-sized users can be particularly impactful. Traditional marketing and product imagery often do not adequately represent the plus-sized demographic, leading to a higher level of information asymmetry and uncertainty for these consumers. By showcasing how products look on bodies similar to theirs, plus-sized users' photos can significantly enhance trust and reduce the perceived risk, making potential buyers more comfortable and confident in their decision to rent or purchase. The reduction of information asymmetry through user-generated pho-

tos can directly influence consumer behavior, leading to increased sales. This is supported by the theory of perceived risk, which suggests that consumers' willingness to purchase is inversely related to the perceived risk associated with that purchase (Mitchell 1999). By providing clearer, more relatable, and trustworthy information, UGC can lower perceived risk, thereby encouraging transactions. In the case of products targeted at plus-sized demographics, user-generated photos from plus-sized users align closely with the needs and concerns of the target audience, making them more effective in reducing information asymmetry and enhancing sales or rentals.

Hypothesis 6. *Photo reviews from plus-sized users increase rentals of products targeted at plus-sized demographics.*

Anonymity in online platforms can encourage free expression and reduce the fear of judgement or retaliation, as users feel more secure when their real identities are not disclosed (Joinson 2001). Especially when it comes to sharing personal content, such as photo reviews, anonymity shields users from direct personal critique and social stigma. Acquisti and Gross (2006) argue that privacy concerns are also a significant factor influencing online behavior, including the willingness to share personal information. The disclosure of personal identifiers, such as names, can exacerbate these concerns, especially when sharing content that might be subject to scrutiny or judgement, like photographs depicting one's body. Users may fear negative feedback, which can be more personally targeting when their real identity is known, leading to a reluctance to post such content.

Hypothesis 7a. *The disclosure of users' identities, specifically through revealing their names, discourages users of all size groups from posting photo reviews.*

The negative effects of online identity disclosure are particularly pronounced for plus-sized users. While concerns regarding privacy are common across all size groups, plus-sized individuals face an additional layer of vulnerability due to the heightened social stigma associated with body size, as discussed previously. For plus-sized individuals, revealing personal

identifiers such as names and faces exposes them to potential body shaming and targeted harassment (Puhl and Heuer 2009). Hence, it may significantly reduce their participation in activities like posting photo reviews. Anonymity serves as a crucial protective mechanism against such stigma, offering a shield against direct personal critique and discrimination. Consequently, revealing their identities is likely to influence plus-sized users the most compared to their medium- and small-sized counterparts. Therefore, we propose the following:

Hypothesis 7b. *The disclosure of users' identities, specifically through revealing their names, most significantly deters plus-sized users from posting photo reviews.*

Hypothesis 7c. *The disclosure of users' identities, specifically through revealing their faces, most significantly deters plus-sized users from posting photo reviews.*

4.3 Data and Variables

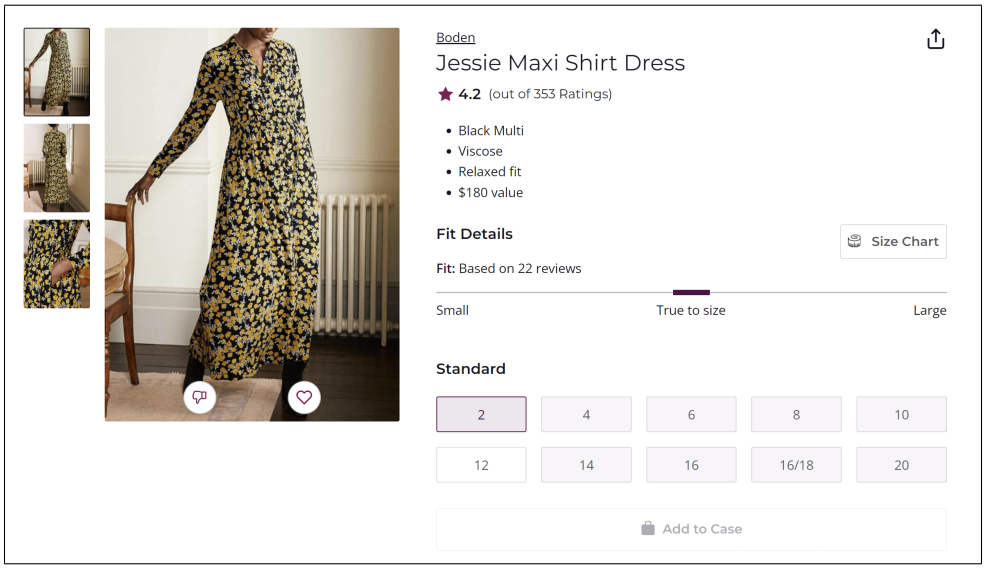
4.3.1 Platform and Context

We partnered with a US-based clothing rental platform that operates on a subscription model, catering specifically to women. This platform offers customers the opportunity to rent apparel for a fixed monthly charge. Our partner firm allows customers to choose a certain number of items to rent each month, depending on their subscription tier. The selected items are shipped to the customer as a "case", which they can wear for up to one month before sending the case back to exchange for a new set. After each rental period, the items are cleaned at the processing center and re-listed on the platform. Customers are encouraged to leave reviews for the items they rent, either in the form of text or photo reviews. If customers are unsatisfied with specific pieces in their case, they have the option to return only those items, and the firm will ship replacements to fill the gap in the case. This is significantly different from digital platforms, where consumers can select other items in the assortment if they are dissatisfied with a TV show or a movie, and there is not much transactional cost for a potential mismatch between customers and a digital product. However, in clothing rental platforms, when customers don't like a product, they experience the hassle of sending back

a specific piece. The company faces a significant cost in sending out a replacement piece to the case if there is customer dissatisfaction. Further, poor experiences can also potentially lead to customers unsubscribing from the platform.

To ensure customer satisfaction, the platform strives to provide accurate information that facilitates informed decision-making, as shown in Figure 4.1. One of the most crucial aspects of the platform is the incorporation of text and photo reviews from other customers who rented the item. These reviews serve as valuable resources for potential renters, allowing them to gain insights into the fit, quality, and overall experience with specific items. By relying on the collective feedback of the community, customers can make more confident rental choices, reducing the likelihood of dissatisfaction and the need for returns.

Figure 4.1: Example of Product Page



4.3.2 Data

Through our partnership, we acquired a proprietary longitudinal dataset that encompasses the platform's product offerings and their rentals over time, as well as user-level information,

including demographic information, rental patterns, product considerations, and product reviews.

Our data covers a two-year period from January 1, 2021 to December 31, 2022[4]. In total, we collect data on 11,445 active users[5] and 66,543 pieces of rented items. During this time, there are 251,720 rental records, 92,166 online reviews, and 17,421 photo reviews. Each rental record in our dataset is detailed, including

- Product information: This includes the style, color, occasion, price, and the date the product became available on the platform.

- User information: This includes the user's age, body size, and the date they joined the platform.

- Rental information: This includes the timestamp of the rental, the rating given by the user, and any associated reviews and/or photo reviews. Note that there could be only at most one review but multiple photo reviews associated with a rental by a user.

4.3.3 Variables and Summary Statistics

Item-week level panel data

For the item-week level panel data, we first select only plus-sized items to focus on the demand patterns of plus-sized consumers. Next, we aggregate the rental-level data at the item-week level, and in total, we have 149,509 item-week observations. Furthermore, we define the weekly demand for each item on the platform as the ratio of the number of days the product is rented out to the number of days it is available for rent. For instance, if an item is returned by a customer and is unavailable for two days due to cleaning, its availability that

[4]During this period, the platform did not implement any policy change that is relevant to our research context. In later sections, we performed a study of policy intervention happening in April 2023. For that specific analysis, we used the data from January 2022 to November 2023.

[5]We define active users as those who have at least one rental activity in our observation period.

week is just 5 days instead of 7 days. If the item is rented out after being available for one day, the demand for that week is calculated as 0.8 (or 80%). Conversely, if a customer keeps the product for an entire week, the demand for that week is expressed as 1.0 (or 100%). This approach to measuring demand reflects the interplay between product availability and rental activity. It parallels demand assessment methodologies used in similar rental platforms like Airbnb, as discussed in Zhang et al. (2022).

The main variable of interest is $\text{AfterPlusPhoto}_{it}$, defined as whether there is a plus-sized photo associated with the product for item i at period t. This panel data allows us to effectively control for item-level, time-invariant characteristics such as categories, styles, and colors. For time-variant characteristics at the item level, we consider influences from three sources:

- Item popularity: We track the total number of clicks TotalClick_{it} received for item i in week t to capture the difference in the popularity of each product in each period.

- Cumulative reviews and ratings: We also track the cumulative impact of user reviews and ratings. We calculate AccReview_{it} for the cumulative number of reviews and AvgRating_{it} for the cumulative average ratings of item i up to week t.

- Seasonality effect: For instance, a dress designed for summer is likely to see increased demand in that season. To account for this, we incorporate a product-quarter level fixed effect, controlling for different seasonal trends for each product.

Table 5.1 summarizes the definitions and summary statistics of all item-week observations.

Review posting data

We next look into the review posting data. We consider three possible choices a customer can make following each rental: no review, posting a review without a photo, or posting a review with a photo. Therefore, we construct two binary variables related to each rental r: Review_r

Table 4.1: Summary Statistics of Item-Week Panel Data

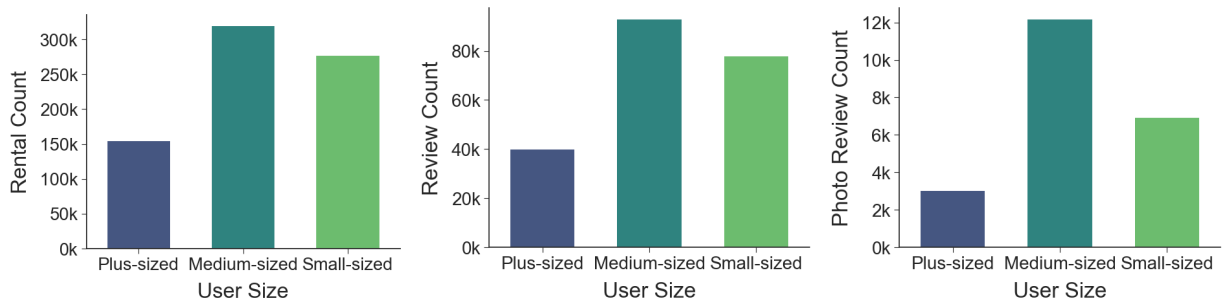
	Variable Definition	Mean	Std. Dev.	Min.	Max.
Outcome Variables:					
Demand	The ratio of the number of days the product is rented out to the number of days it is available for rent for the item in the week	0.47	0.48	0	1
Independent Variable:					
AfterPlusPhoto	Binary variable, indicating whether there is a plus-sized photo associated with the product for the item until the week, 0 otherwise	0.05	0.20	0	1
TotalClick	Total number clicks received by the item in the week	6.21	10.85	0	228
AccReview	Cumulative number of reviews received by the product associated with the item until the week	7.48	9.84	0	140
RatingAvg	Cumulative average rating by the product associated with the item until the week	3.30	1.38	0	5

and Photo_r, indicating whether the user posts a review and a photo review, respectively, after the rental. The main effect of interest is how the body size of each individual affects the review/photo posting behavior.

To study the differences in review and photo posting behavior among users of different size groups, we first categorize all users into three groups based on their sizes: small-sized group (size below 6), medium-sized group (size from 6 to 12), and plus-sized group (size above 12). The distribution of rentals, reviews, and photo reviews across different size groups is displayed in Figure 4.2a, 4.2b, and 4.2c, respectively. We observe that the rental distribution is comparable to the review distribution based on Figure 4.2a and 4.2b. This indicates that a similar percentage of rentals from each group lead to reviews. However, in Figure 4.2c, there

is a notable decrease in the number of photo reviews for the plus-sized group in comparison to the other two groups. This finding provides initial evidence that plus-sized users are underrepresented in online photo reviews.

Figure 4.2: Distribution of Reviews and Photo Reviews Across User Size Groups



(a) Rental Count (b) Review Count (c) Photo Review Count

In our study, we also explore various other factors that could affect a user's decision to write a review or post a photo along with the review. This includes the user's experience of a rental, captured by UserRating_r, and the influence of existing feedback, as indicated by the number of accumulated reviews (TotalFeedback_r) and photos (TotalPhoto_r) at the time of the rental r. To specifically understand the impact of body size in the context of reviews and photos, we also construct the variables SimSizeReview_r and SimSizePhoto_r, which are the percentage of reviews and photo reviews, respectively, coming from users who are of the similar size as the focal user. This allows us to differentiate the effects of reviews and photos between similar-size and different-size users. Additionally, we consider the average ratings (AvgRating_r) of items at the time of the rental, which might impact the user's perception of quality. Item-specific attributes such as the price (Price_r) and duration on the platform (ItemAge_r) are also taken into account, as they could influence user interactions with the item. The tenure of the user on the platform (UserTenure_r), their age (UserAge_r), the number of reviews ($\text{UserTotalFeedback}_r$) and photo reviews ($\text{UserTotalPhoto}_r$) posted by the user

Table 4.2: Summary Statistics of Review Posting Data

	Variable Definition	Mean	Std. Dev.	Min.	Max.
Outcome Variables:					
Review	Binary variable, indicating whether the user posts a review after the rental	0.37	0.48	0	1
PhotoReview	Binary variable, indicating whether the user posts one or more photo reviews after the rental	0.03	0.18	0	1
FaceRevealing	Binary variable, indicating whether the user posts one or more photo reviews with face revealing after the rental	0.03	0.17	0	1
Independent Variables:					
SmallSize	Binary variable, indicating whether the user associated with the rental is small-sized	0.35	0.48	0	1
PlusSize	Binary variable, indicating whether the user associated with the rental is plus-sized	0.21	0.41	0	1
UserAge	The age of the user associated with the rental	43.06	9.42	18	80
Price	The price paid by the platform for the item associated with the rental	70.80	44.11	0	525
UserRating	The user's rating of the rented item	4.02	1.09	1	5
UserTenure	The tenure of the user since the date that the user joined the platform	1.66	1.48	0.00	6.04
UserTotalFeedback	The total number of reviews from the user prior to this rental	206.86	244.77	0	1306
UserTotalPhoto	The total number of photo reviews from the user prior to this rental	1.22	11.17	0	311
AvgRating	The average rating of the item prior to this rental	3.95	0.46	1	5
TotalFeedback	The total number of reviews received by the item prior to this rental	17.58	20.19	0	259
TotalPhoto	The total number of photos received by the item prior to this rental	1.64	2.29	0	24
SimSizeFeedback	The percentage of existing reviews from similar-sized users prior to this rental	0.44	0.28	0	1
SimSizePhoto	The percentage of existing photos from similar-sized users prior to this rental	0.25	0.38	0	1
PhotoAesthetic	The aesthetic score[6] of the photo posted by the user, if any	0.14	0.77	0	6.06
ItemAge	The number of years since the first day that the item is available on the platform	1.69	1.19	0	6.04

Note: The statistics for Price are reported after rescaling due to the nondisclosure agreement.

prior to the rental are other variables included in our analysis. Table 4.2 summarizes the definitions and summary statistics of variables for the review data.

4.4 Empirical Evidence

In this section, we seek to verify Hypothesis 5 and 6. First, we establish that there is a marked difference in the way plus-sized women contribute to UGC, which can in turn exacerbate the informational asymmetry. Second, we demonstrate how this informational asymmetry affects users' decision-making. Our rich data allows us to observe rental decisions before and after a photo review is provided. We provide evidence of how photo reviews from similar-sized users are more likely to influence subsequent rental decisions. This effect is particularly salient among plus-sized users.

4.4.1 Photo Review Posting Behavior Across Different User Size Group

To gain insights into different review and photo review posting behavior across user size groups, we formulate the following regression model:

$$\mathrm{Y}_r = \beta_0 + \beta_1 \mathrm{PlusSize}_u + \theta \mathrm{Control}_r + \mu_t + \lambda_i + \lambda_i \times \delta_t + \epsilon_r. \tag{4.1}$$

In Equation 4.1, r represents a rental from user u for item i in week t. We consider two binary dependent variables (Y_r): whether user u posts a review and whether the user posts a review photo for rental r. $\mathrm{PlusSize}_u$ is a binary variable indicating whether the user u is from the plus-sized group. In addition to each user's size, we also include the user's age ($\mathrm{UserAge}_r$) and the duration of their membership on the platform ($\mathrm{UserTenure}_r$) as control variables. We also include the user's rating of this rental ($\mathrm{UserRating}_r$) to control for the user's experience with the rental. We further include $\mathrm{Log(AccReviews)}_r$ to account for the number of reviews of item i in rental week t, and $\mathrm{RatingAvg}_r$ to control for the accumulated rating of item i in week t. The variable $\mathrm{ItemAge}_r$ represents the time interval from when item i first became available on the platform until week t. We also include week-level fixed effects

γ_t to control for common calendar shocks, while λ_i accounts for item-specific characteristics. To account for seasonality, we further include the item-quarter level fixed effects.

Table 4.3 demonstrates the variations in online review behavior among different size user groups. Our results show that plus-sized users exhibit a marked decrease in the likelihood of posting reviews or photo reviews after renting, indicated by coefficients of -2% and -1%, respectively, compared to their small or medium-sized counterparts. These results highlight a notable imbalance in review and photo review contribution behaviors across different size groups, emphasizing a particular under-representation of plus-sized users in both reviews and photos. It is also important to note that during the period of this analysis, the reviews were anonymous and the only way the users can reveal their identity is via photos, as demonstrated in Figure 4.3.

Figure 4.3: An Example of a Review

This finding suggests that the likelihood of contributing to UGC is correlated with the users' size, controlling for user and product-related characteristics. Furthermore, the users' size also affects the likelihood of leaving photo reviews on the platform. This confirms our Hypothesis 5 and strengthens the argument that users facing significant social stigma in everyday life might be less likely to participate in online forums, especially in situations that

Table 4.3: Review and Review Photo conversion

	(1)	(2)
	Review	Photo Review
PlusSize	-0.02***	-0.01***
	(0.01)	(0.00)
UserAge	0.00***	-0.00***
	(0.00)	(0.00)
UserTenure	-0.03***	-0.01***
	(0.00)	(0.00)
UserRating	-0.03***	0.01***
	(0.00)	(0.00)
ItemAge	-0.03	-0.00
	(0.04)	(0.01)
Log(AccReview)	-0.12***	-0.01***
	(0.00)	(0.00)
AvgRating	-0.01	-0.00
	(0.01)	(0.00)
Constant	0.74***	0.03
	(0.08)	(0.03)
Item FE	Yes	Yes
Week FE	Yes	Yes
Item*Quarter FE	Yes	Yes
Observations	174958	174958
R^2	0.421	0.415

Standard errors clustered at the item level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is rental (r) level.

decrease their anonymity. Thus, there is a significant impact of users' size on their decision to participate in the supply side of UGC, leading to informational asymmetry. Next, we analyze whether this informational asymmetry can have any ramifications on the demand.

4.4.2 Impact of Plus-sized Photo Reviews on Plus-Sized User Demand

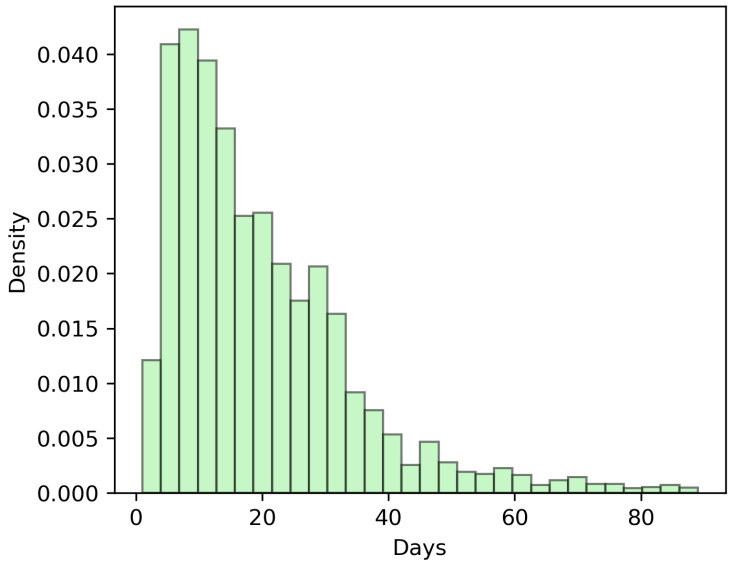
In this section, we study how plus-sized users affect the demand for a product. The product page on the platform is not static — each item accrues reviews over time as users consume the item. In other words, the information available to each customer visiting a product page varies based on the information present in reviews at a particular time t. Thus, a customer visiting an item with no reviews only has the baseline information provided by the platform (description and the model) while the customer visiting an item with textual and or photo reviews has more information. The variation in the level of information available to each customer provides us with the identification, which we exploit in our analysis.

Further, unlike other observational studies in the online review literature (e.g., Chevalier and Mayzlin 2006, Zervas et al. 2021) which mostly rely only on reviews written on a platform, we have a unique opportunity to observe the demand for any item i in week t at a subscriber level as the item gets more reviews. This presents us with a way to identify how reviews impact demand. In this analysis, each item refers to a unique product that is available for rental on the platform. Each item i is available in multiple sizes.[7] If a plus-sized user rents an item i in her size, she can choose to write a textual review and/or leave a photo review after her rental. The timing of a photo review from a plus-sized user for an item is exogenous to the demand for the product at that time. This identification assumption is further strengthened by the fact that users can review the item at any point during their use or immediately after sending it back. There is a large variance in the timing of a photo review as shown in Figure 4.4. This variance, in turn, affects the information available to each customer of each item when they visit the product page. For example, user u_1 in week

[7]Given that items without a plus-sized inventory can only have demand = 0, we exclude these items from our analysis.

t_1 might not have seen a photo review, while user u_2 in week t_2 sees a photo review and thus has more information for the same product. We exploit this informational shock to determine how the demand for the item i changed before and after a photo review from a plus-sized user.

Figure 4.4: Distribution of Days from Rental to Photo Review Posting



This presents us with a Difference-in-Differences (DiD) framework. We consider an item to be treated as soon as it receives the first plus-sized photo review. This implies that each item gets treated at a different point in time, which provides us with a staggered DiD framework. In the staggered adoption framework, each untreated item (i.e., the items that have *not yet* received any photos from plus-sized users) serves as a control for the treated category until it gets treated.

We formulate the model as follows:

$$y_{it} = \beta_0 + \beta_1 \text{AfterPlusPhoto}_{it} + \gamma \text{Control}_{it} + \lambda_i + \delta_t \times \lambda_i + \mu_t + \varepsilon_{it}. \tag{4.2}$$

Here, the outcome variable is y_{it}, which represents the demand (measured by rentals) from

plus-sized users for item i in week t. We denote the treatment as AfterPlusPhoto$_{it}$. Since the information available to plus-sized users increases substantially after the item receives the first plus-sized photo, AfterPlusPhoto$_{it}$ remains 1 from the week t when it received the first plus-sized photo (Proserpio and Zervas 2017, Chevalier et al. 2018, Ananthakrishnan et al. 2023).

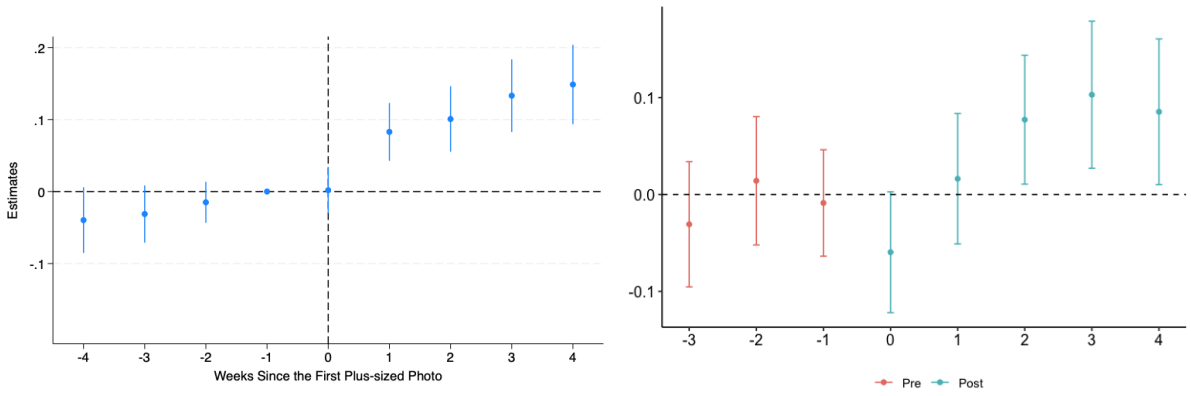
β_1 is the coefficient of interest and provides us with the average treatment effect (ATE), which is the average difference of the differences between the average outcomes of the treated and the control products in each period. We add item-level fixed effects for time-invariant item characteristics, such as brand and style, which can influence the demand. We also include week-level time fixed effects to account for the common calendar shock that could influence demand for all products, following prior literature (Autor 2003, Angrist and Pischke 2009, Greenwood and Wattal 2017, Burtch et al. 2018a). In addition to adding the two-way fixed effects, we include item-quarter interaction fixed effects to control for seasonal effects that affect the demand for specific products. For example, red and party dresses are more likely to be in demand during the holiday season, while flowy or linen dresses are in higher demand during the summer. These changes in seasonal demand tied to each product are captured by the item-quarter fixed effects (Zervas et al. 2017).

Our unique dataset also allows us to add a series of control variables. We include the total number of reviews each product has received as of week t, the cumulative average rating by week t, and the number of clicks received by this item in week t (to control for interest in item) as shown in Section 4.3.3. We estimate this model using a matched-sample procedure (Xu et al. 2017, Chan and Wang 2018). There is a large long tail of items (over 75%) that are never rented on the platform, and we remove these items from our analysis.

Table 4.4 presents the estimation results of our DiD design. Column (1) shows the results without the control variables, while Column (2) includes the control variables. The coefficient of AfterPlusPhoto$_{it}$ is positive and significant, indicating that a review photo from a plus-sized user positively influences product demand. On average, a photo from a plus-sized user is associated with a 10% increase in demand.

To ensure the validity of the DiD analysis, we verify the parallel trend assumption by replacing AfterPlusPhoto$_{it}$ with the weeks since treatment for five periods before and after the treatment. Figure 4.5a shows the event study estimates using the TWFE estimator. Reassuringly, there are no significant differences between the control and treated groups in the pre-period thus verifying the parallel trends assumption. Furthermore, we observe a significant increase in demand among plus-sized users in the periods after the reception of the first plus-sized photo.

Figure 4.5: Parallel Trend Assumption



(a) TWFE Estimator (b) CS Estimator

Note. We use a 95% confidence interval.

Recent studies have raised concerns about negative weights in the staggered DiD estimation using TWFE estimators (De Chaisemartin and d'Haultfoeuille 2020, Sun and Abraham 2021, Goodman-Bacon 2021). To address this issue, we present the main results using the estimator proposed by Callaway and Sant'Anna (2021) (henceforth referred to as the CS estimator). The results remain consistent with our previous findings. Figure 4.5b illustrates the event study estimates obtained from the CS estimator.

Our analysis demonstrates the importance of photo reviews from plus-sized users for other plus-sized users, confirming Hypothesis 6. These findings suggest the presence of significant

Table 4.4: Impact of Photos from Plus-sized Users on Demand

	(1)	(2)
	Demand	Demand
AfterPlusPhoto	0.12***	0.10***
	(0.02)	(0.02)
TotalClick		-0.01***
		(0.00)
Log(AccReview)		-0.02
		(0.05)
RatingAvg		0.02*
		(0.01)
Constant	0.47***	0.48***
	(0.00)	(0.06)
Item FE	Yes	Yes
Week FE	Yes	Yes
Item × Quarter FE	Yes	Yes
Observations	148969	148969
R^2	0.550	0.563

Standard errors clustered at the item level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) - week (t) level.

information asymmetry among plus-sized users and highlight how reducing this asymmetry could benefit the plus-sized community. This observation raises important questions: What are the main drivers of this information asymmetry and what can platforms do to alleviate it? We address these questions in the following section.

4.5 Mechanisms

In this section, we aim to uncover the mechanisms contributing to the imbalance in review photo posting behavior among users of different sizes. In the theoretical section, we explored several reasons that might drive the intent to contribute to UGC. Here, we provide evidence of the potential mechanisms contributing to the imbalance in review photos among users of different sizes and verify the Hypothesis 7a, 7b and 7c, derived from the theoretical section.

4.5.1 The Role of Privacy

Our partner platform initially only displayed the reviews and anonymized the reviewers' names. In April 2023, the platform implemented a major change to its review page by displaying the names of reviewers.[8] Further, other subscribers to the platform could click on the reviewers' profiles and view their rental history on the platform. This significant change in informational control could lead to two potential outcomes. On the one hand, deanonymizing reviewers' names could boost the production of reviews, as reviewers might be incentivized to enhance their status and demonstrate their penchant for style on the platform. On the other hand, deanonymizing reviewers' names could heighten privacy concerns, particularly for photo reviews, where other subscribers could view the photos and also the profiles of the reviewers posting photo reviews (thus connecting the photo with a profile). This privacy concern might be even more salient among plus-sized users. Therefore, this policy change presents an ideal backdrop for studying how privacy concerns influence users' willingness to post reviews and photos, as described in Hypothesis 7a and 7b.

To do so, we estimate the following model:

$$\mathrm{Y}_r = \beta_0 + \beta_1 \mathrm{AfterPolicy}_r + \beta_2 \mathrm{AfterPolicy}_r \times \mathrm{PlusSize}_u + \lambda_i + \delta_t \times \lambda_i + \mu_t + \gamma_u + \varepsilon_r. \tag{4.3}$$

In Equation 4.3, Y_r is the dependent variable. This refers to Review_r or $\mathrm{PhotoReview}_r$, indicating whether a review or photo review is posted following a rental. $\mathrm{AfterPolicy}_r$ is

[8] See Appendix B.1 for examples of a review before and after the policy shock.

a binary variable that indicates if the rental r occurred after the deanonymization policy change. The coefficient of the interaction term between AfterPolicy_r and Plus_u captures the potential differential effects of the policy change among users of different sizes. We include the user-level fixed effect to account for time-invariant user-level characteristics that might influence their decision to write reviews. Similar to the analysis in Section 4.4.1, we also include the UserRating_r as a control as well as the Year-Month and Item-Quarter fixed effects.

We run this model using all the rental data from Jan 2022 to Nov 2023[9]. The results are provided in Table 4.5. Column (1) of Table 4.5 indicates that deanonymizing reviews do not have a significant effect on textual review posting behavior.The policy change did not impose additional privacy costs on plus-sized users in the context of review postings. However, the results differ for photos. Post-policy, there is a noticeable decline in the number of photos users post, specifically among plus-sized users, thus confirming Hypothesis 7a and 7b. Together, these results suggest that there is a significant heterogeneity in privacy concerns between users of different sizes. Thus, privacy concerns can significantly reduce the generation of UGC among plus-sized users. This differential impact underscores the importance of considering varying privacy costs in designing informational controls that can balance privacy concerns and yet, provide incentives for users to contribute to UGC platforms.

4.5.2 The Mechanisms of Photo Review Posting

Finally, we delve into exploring the mechanisms that guide users of different sizes in posting photo reviews. Specifically, to understand the varying privacy concerns across user groups with different body sizes, we examine the reluctance associated with sharing identity-revealing information, particularly through face-revealing photos[10] in photo reviews.

[9]For all other analyses, we used data from Jan 2021 to Dec 2022.

[10]For examples of both face-revealing and non-face-revealing review photos, please see Online Appendix B.2.

Table 4.5: Impact of Privacy Policy Shock on Review and Photo Posting Behavior

	(1)	(2)
	Review	Photo Review
After	0.001	-0.013***
	(0.005)	(0.002)
After × PlusSize	-0.007	-0.008***
	(0.007)	(0.003)
UserRating	-0.039***	0.014***
	(0.001)	(0.000)
Constant	0.562***	0.012***
	(0.005)	(0.002)
User FE	Yes	Yes
Item FE	Yes	Yes
Year-Month FE	Yes	Yes
Item × Quarter FE	Yes	Yes
Observations	240520	240520
R^2	0.748	0.755

Standard errors clustered at the item level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is rental (r) level.

Model Setup

We model the decision-making process of customers posting photo reviews on a platform as a two-stage process. Initially, after each rental, users evaluate their utility to post photo reviews by considering the existing reviews, photo reviews, their overall satisfaction with the product, and their own characteristics. The decision to post a photo review is modeled as

follows:

$$u_{photo,r} = \beta_0^1 + \beta^1 X_r + \varepsilon_r^1$$

$$PhotoReview_r = 1, \text{ if } u_{photo,r} > 0.$$

Here, X_r represents factors affecting the utility of posting a photo review. Specifically, we consider characteristics of existing reviews of this product (including the total number of existing reviews and the total number of existing photos), the characteristics of this product (average rating, price of the product, the length of time the product has been on this platform), the experience of this rental (user i's rating of this product), and the characteristics of this user associated with the rental (user i's tenure on the platform, age, size). To capture the heterogeneous effect of reviews and photo reviews from similar-sized peers versus different-sized peers, we further include the percentage of similar-size reviews and the percentage of similar-size photo reviews.

Upon deciding to post a photo review, the user then considers whether to include identity-revealing information, such as their face, in the photo. To rule out potential mechanisms other than privacy concerns that contribute to the heterogeneous utility of different user groups, we include two sets of control variables. First, we include the aesthetic score of the photo to gauge the face-revealing decision from the perspective of aesthetics. Second, we control for the total number of reviews and review photos posted by the user prior to the rental to account for the influence of self-presentation needs (Goffman 2006). By incorporating these controls, we can more accurately attribute the varied choices in revealing faces among different user groups specifically to privacy concerns.

Specifically, we model the utility of revealing the face as

$$u_{face,r} = \beta_0^2 + \beta_1^2 SmallSize_u + \beta_2^2 MediumSize_u + \beta_3^2 PlusSize_u + \gamma^2 W_r + \varepsilon_r^2$$

$$FaceRevealing_r = 1, \text{ if } u_{face,r} > 0 \quad \text{and} \quad PhotoReview_r = 1.$$

Here, $SmallSize_u$, $MediumSize_u$ and $PlusSize_u$ are binary variables indicating the size of the user associated with the rental. For identification purposes, we normalize $MediumSize_u$

to zero. W_r includes the aesthetic score of the photo and the user's total number of reviews and photo reviews prior to the rental. For the aesthetic score, we utilized the Neural Image Assessment (NIMA) method (Talebi and Milanfar 2018) to estimate the aesthetic score of review photos. Details can be found in Appendix B.3.

In this model setup, we only observe the face-revealing decision of users who decide to post a photo review, and this may impose a selection bias. As a result, estimation of the coefficients could yield inconsistent estimates if the error terms (ε_r^1 and ε_r^2) are correlated (Heckman 1979). Therefore, instead of specifying the error terms to be independent, we assume the error terms follow a bivariate standard normal distribution with a correlation coefficient ρ as

$$(\varepsilon_r^1, \varepsilon_r^2) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right),$$

following Van de Ven and Van Praag (1981).

Model Estimation

Unlike traditional Heckman selection models where the dependent variable in the second stage is a continuous variable, here we only observe a binary choice instead of the continuous value of latent utility ($u_{face,r}$) in the second stage. Therefore, we cannot apply the standard two-step Heckman bias correction method directly. Instead, our estimation is based on the maximum likelihood estimation (MLE).

Given the independence across observations, the likelihood of observed photo review posting behavior of each individual can be expressed as the product of the probabilities across the two stages. Thus, the log-likelihood function L can be expressed as:

$$\begin{aligned}\ln L = &\sum_{PhotoReview_r=1, FaceRevealing_r=1} \ln\left\{\Phi_2\left(\bar{u}_{face,r}, \bar{u}_{photo,r}; \rho\right)\right\}\\ &+ \sum_{PhotoReview_r=1, FaceRevealing_r=0} \ln\left\{\Phi_2\left(-\bar{u}_{face,r}, \bar{u}_{photo,r}; -\rho\right)\right\}\\ &+ \sum_{PhotoReview_r=0} \ln\left\{1-\Phi\left(\bar{u}_{photo,r}\right)\right\},\end{aligned}$$

where $\Phi_2(\cdot)$ is the cumulative bivariate normal distribution function (with zero mean and correlation coefficient ρ), $\Phi(\cdot)$ is the cumulative standard normal distribution function, and $\bar{u}_{photo,r}$ and $\bar{u}_{face,r}$ is the expected utility (without adding the random errors) in the two-stage decisions respectively.

Model Result

Table 4.6 shows the results of the parameter estimates. At the stage of photo review decision, users exhibit a lower likelihood of posting a photo review as the volume of existing reviews increases. This suggests a diminishing altruistic incentive when there is an abundance of reviews. In contrast, the presence of more photos, particularly of similar size, encourages users to share their photo reviews. This underscores the significance of modality in shaping users' participation decisions in UGC platforms. Notably, users with lower ratings exhibit a greater likelihood of posting photos. Interestingly, younger users exhibit a stronger inclination to share photos than their elder counterparts.

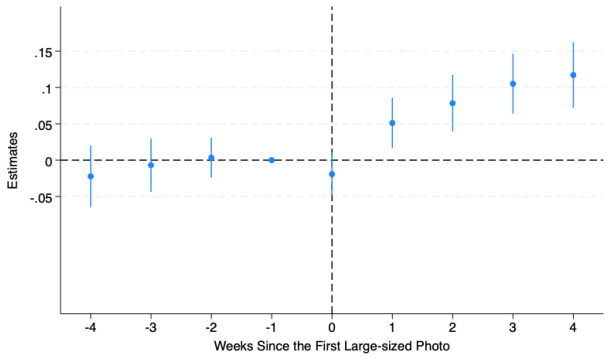
In the decision-making stage of whether to post a photo with a face, we can discern the impact of privacy concerns when posting identity-related information. Our findings reveal that plus-sized users incur the highest privacy costs, which discourage them from sharing photos containing faces. Similarly, small-sized users also contend with relatively elevated privacy costs. In summary, it becomes evident that plus-sized users exhibit a natural inclination to participate less in both review and photo-sharing activities, with privacy considerations acting as an additional deterrent. As alternative mechanisms, we find that when the aesthetic score of the photo is higher and the user has higher needs for social presence, they are more likely to post a review photo with their face revealed.

4.6 Robustness Check

4.6.1 Full Sample without Matching

In our main model, we used a matched sample to address the issue of the unbalance in the treatment assignment. Here, we re-run our main model using the full sample without the propensity score matching method. Figure 4.6 shows the results, which are consistent with our main findings.

Figure 4.6: Robustness Check: The Impact of Photos from Plus-sized Users on Demand using Full Sample



Note. We use a 95% confidence interval.

4.6.2 Alternative Measure of Demand

In our main model, we use the ratio of the number of days that an item is rented to the number of days that the item is available as the measure of the product demand. In this robustness check, we adopt an alternative demand metric: the rental conversion rate derived from click-level data.

We specify our model as follows:

$$\begin{aligned}\text{RentalIndicator}_c = \beta_0 + \beta_1 \text{SimsizePhotos}_c + \beta_2 \text{SimsizePhotos}_c \times \text{PlusSize}_u \\ + \beta_3 \text{Log(AccReviews)}_c + \beta_4 \text{RatingAvg}_c + \lambda_i + \delta_t \times \lambda_i + \gamma_t + \mu_u + \epsilon_c,\end{aligned} \tag{4.4}$$

where c stands for a click from user u on item i at time t. We use all the clicks from Jan 2021 to Dec 2022. Our outcome variable RentalIndicator_c indicates whether user u decides to rent item i after clicking on it, it equals 1 if the conversion happens, 0 otherwise. The variable of interest, $\text{SimsizeAccPhotos}_c$, represents the proportion of accumulated review photos for item i, contributed by users of the similar size as the focal user u across all review photos of item i, at the time of click c. To examine the varying impact of similar-size photos on different size groups, we include the interaction term of $\text{SimsizeAccPhotos}_c$ and PlusSize_u, which equals to 1 if the user is plus-sized. Subsequently, we incorporate several control variables similar to our main model. We include Log(AccReviews)_c to capture the number of reviews of item i at click time t. We also include RatingAvg_c to capture this rating of item i at click time t. We include the item-quarter interaction fixed effect and the running week fixed effect as in our DiD design. Moreover, we include the user-level fixed effect to control any characteristics corresponding to the user.

Our results are presented in Table 4.7. Column (1) shows the result without the interaction term while Column (2) shows the result with the interaction term. The coefficient of the interaction term, that is β_2 in (4.4), captures how the number of similar-size photos impacts rental from the baseline group, specifically plus-sized users. On average, an additional review photo from the plus-sized group is associated with a 1% increase in the probability of rental conversion. These findings indicate that review photos from similar-size users have a positive impact on the rental conversion of a click and this effect is greater for plus-sized users. This further confirms our main finding: photos from plus-sized users significantly increase future demand from plus-sized users.

4.7 Conclusion and Discussion

Our analysis uncovers significant disparities along size lines on e-commerce platforms. Plus-sized users face substantial informational uncertainty due to their under-representation in reviews and review photos. Despite their increasing purchases, we find that plus-sized users are less likely to leave text or photo reviews, which can lead to even less information for subsequent plus-sized users. This information asymmetry is particularly detrimental to sales among plus-sized users, as our DID model and click-level analysis show that photo reviews from plus-sized users significantly increase purchases from the plus-sized population.

By deploying a policy change on the platform, we empirically uncover the mechanisms behind such underrepresentation. In April 2023, the platform began displaying users' names, entailing privacy concerns. For plus-sized users, the social stigma associated with identity revelation could exacerbate privacy concerns. A comparison of user behavior before and after the policy launch indicates that all users reduced their photo postings, with plus-sized users being the most affected.

Beyond privacy concerns, we also identify other factors contributing to the underrepresentation of plus-sized consumers by constructing a hierarchical choice model. We find that review photos from similar-sized users significantly influence the review photo posting behavior. Specifically, a user has more incentive to post a photo review when there are more similar-sized photos posted already. We also find that the privacy cost is higher among plus-sized users, which also contributes to their reticence in sharing face-inclusive reviews.

Our research contributes to the existing literature by highlighting the role of privacy in the reviewing behavior of consumers who routinely face social stigma. This is not unique to clothes. Anecdotal evidence suggests that plus-sized users have significant uncertainty when visiting restaurants or entertainment places. In fact, this has led to the creation of exclusive apps where plus-sized consumers can inform other plus-sized consumers about relevant issues.[11] Reports in media also points to a lack of information in reviews for consumers with

[11] https://www.nytimes.com/2019/03/12/dining/larger-customers-restaurants.html, last

disabilities who might have questions about accessibility.

Our results have important managerial implications for platform design. We show that the modality of UGC is paramount in engagement. If the platform seeks textual input through UGC (such as reviews), the decision to review will be dependent on how much value the focal user is adding to the conversation, especially if there are other users of the same size and with similar experiences. In this case, platforms can choose to prompt users on specific aspects of the product that are missing from other reviews, which might increase participation in textual reviews. Review platforms like Yelp and Google Maps address this issue by providing information about accessibility, suitability for families with children, etc.

If the platform is built around videos or images, there is a high likelihood of excluding users who face stigma over their appearance. In this case, it is incumbent on the platform to allay privacy concerns. For example, platforms could provide anonymity features for customers who write reviews and offer options to auto-blur or conceal faces in the photos they post. Platforms can also create a conducive environment for users to share. Our results reveal that when plus-sized users see photos of other plus-sized users, they are more likely to post photos. Platforms can invite plus-sized models or influencers to post content, which could impede self-censorship caused by body image issues.

Plus-sized customers play a significant role in the apparel market. Platforms should increase the informational value by providing models of different sizes rather than conforming to the pervasive zero-sized model because a zero-sized model provides very little information about the fit of the product for most users of the platform. Platforms can provide filtering and sorting options to discover reviews from users closest to their size. An inclusive and representative environment in their promotional materials and user-generated content is not only beneficial for branding but provides a rather large, understudied, and untapped business value. Apparel returns contribute to billions of losses in the apparel industry, most of them resulting from poor fit. Customers empowered with information are less likely to have a

---

accessed April, 2024.

negative experience and return a product that does not fit. Ultimately, when models are diverse, there is less need for consumers to fill in the void and the informational scarcity is preempted even before the first review is written. Interestingly, technology firms are now beginning to leverage generative AI to create diverse models to improve information for the end-user in the absence of such efforts from apparel vendors.[12]

While our research offers valuable insights into the challenges faced by plus-sized consumers on e-commerce platforms, several limitations should be noted. First, our study focuses specifically on e-commerce platforms, which means that the findings may not be directly applicable to other types of platforms where user-generated content also plays a significant role. Secondly, the demographics of our user sample, possibly limited in gender, age, geography, and socioeconomic status, could introduce biases that may affect the generalizability of our findings. Lastly, the intervention suggestions we put forth, such as anonymity features and auto-blurring faces in review photos, are grounded in theory but have yet to be empirically tested for effectiveness. These limitations provide avenues for future research to delve deeper into these and related areas.

[12] https://blog.google/products/shopping/ai-virtual-try-on-google-shopping/

Table 4.6: Mechanism Analysis

Variable	Estimates
Identity-disclosure Stage	
PlusSize	-0.192***
	(0.064)
SmallSize	-0.078*
	(0.046)
UserTotalFeedback	0.090***
	(0.034)
UserTotalPhoto	0.003
	(0.006)
PhotoAesthetic	0.093***
	(0.012)
Constant	1.786***
	(0.178)
Photo Review Posting Stage	
TotalFeedback	-0.118***
	(0.009)
TotalPhoto	0.094***
	(0.006)
SimSizeFeedback	-0.010
	(0.020)
SimSizePhoto	0.080***
	(0.014)
AvgRating	-0.018***
	(0.005)
Price	0.008
	(0.005)
UserRating	0.277***
	(0.007)
UserTenure	-0.158***
	(0.006)
UserAge	-0.158***
	(0.006)
ItemAge	-0.016
	(0.007)
SmallSize	-0.024**
	(0.012)
PlusSize	-0.221***
	(0.015)
Constant	-1.929***
	(0.012)
ρ	-0.378***
	(0.089)
1st Stage Observations	249,025
2nd Stage Observations	7,683
Log-likelihood	-34331.82

Standard errors are in parentheses.
* p<0.10, ** p<0.05, *** p<0.01.
The unit of analysis is rental (r) level.

Table 4.7: Robustness Check: Impact of Same Size Photo on Rental Conversion

	(1)	(2)
	RentalIndicator	RentalIndicator
SimSizePhoto	0.01***	0.01***
	(0.00)	(0.00)
PlusSize × SimSizePhoto		0.01***
		(0.00)
Log(AccReview)	-0.03***	-0.03***
	(0.00)	(0.00)
AvgRating	0.01***	0.01***
	(0.00)	(0.00)
Constant	0.15***	0.15***
	(0.01)	(0.01)
User FE	Yes	Yes
Item FE	Yes	Yes
Week FE	Yes	Yes
Item × Quarter FE	Yes	Yes
Observations	2161696	2161696
R^2	0.125	0.125

Standard errors clustered at the item level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is click (c) level.

Chapter 5

"BE THE BUYER"–LEVERAGING THE WISDOM OF THE CROWD IN E-COMMERCE ASSORTMENT PLANNING

5.1 Introduction

E-commerce sales in the United States (US) hit a record high of $1 trillion in 2022 (Conley 2023). Of this, roughly 20% is attributed to fashion e-commerce – the buying and selling of fashion-related items, for example, apparel, footwear, and accessories – which is notoriously hard to predict (Pasquali 2022). Retailers in this industry thus make inventory decisions for new products under high uncertainty. To address this fundamental shortcoming, some retailers have adopted a new business practice of "crowdsourced buying" wherein retailers first seek input from their existing customers on the product's desirability and base their purchasing decision on votes from these customers.

This idea of crowdsourcing prior to making assortment and inventory decisions has become quite popular in the last few years. Each of the new ideas submitted by creators on Betabrand, a San-Francisco based apparel company, is voted on by customers before a production decision is made (see Figure 5.1 for an illustration). Threadless, a platform for apparel design, allows users to vote for their favorite T-shirt designs and the best designs are printed and sold (Chen et al. 2020). The furniture retailer, Ikea, launched Co-Create IKEA which allows customers to develop and test new products. Amazon seeks votes on pilots of new TV shows before ordering an entire season of the show. A similar approach is used by Lego Ideas, where users vote to support new Lego designs. This idea of "Open Innovation" is particularly prevalent in industries that have creative and qualitative nuances (Chesbrough 2003).

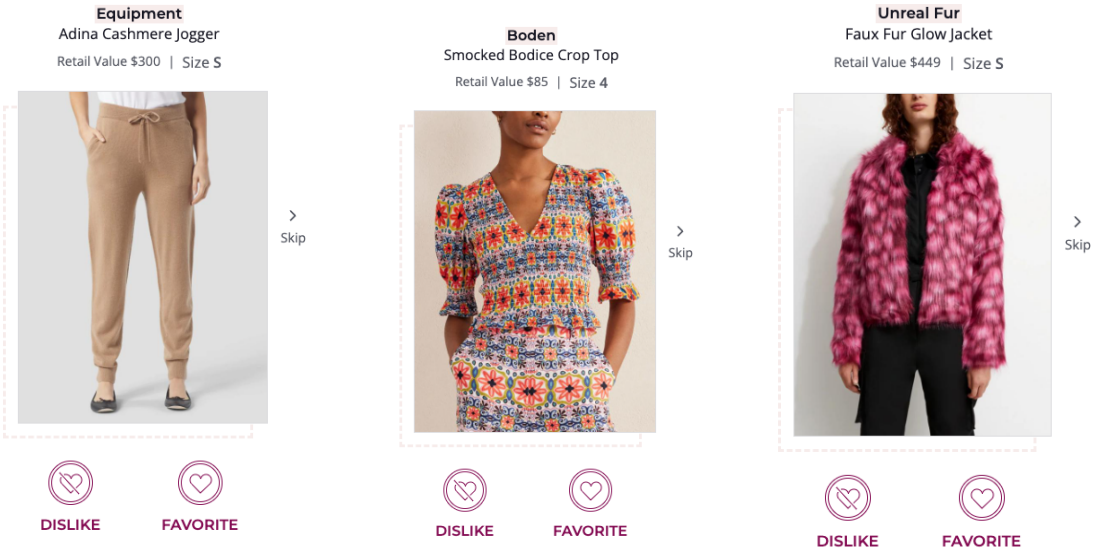
Despite the widespread adoption of crowdvoting in retailing, empirical evidence about its

implications for retailers remains limited. Field experiments revealed that products labeled "customer-ideated" perform better, driven by customers' perception that these products are better designed to match the needs of end-users (Nishikawa et al. 2017). However, it remains unclear whether a firm can improve its operational performance in the absence of such labeling. It is plausible that the voting results can be useful in reducing uncertainty in demand, akin to advance selling (Fisher and Raman 1996), allowing the retailer to make better operational decisions. Crowdvoting, however, involves nuances that are different from advance selling. First, prior work cautions that systems that offer no reward to voters (e.g., via price discounts or first dibs on the crowdvoted products) may not supply accurate information to the firm (Marinesi and Girotra 2013). Second, the voting results can be biased if the tastes of the voter base are vastly different from those of the end users. Lastly, voters can suffer from fatigue, and the novelty can wear off quickly, especially in the absence of a reward. All of these can lead to the retailer making poor decisions.

Figure 5.1: Illustration: Betabrand's crowdvoting system

In this paper, we study this increasingly popular practice of incorporating crowdvoting into the product purchase decision on e-commerce platforms to better understand the implications. To do so, we collaborated with a subscription-based apparel rental platform that launched crowdvoting (hereafter referred to as "the game"). Users can vote on potential

Figure 5.2: Illustration: Partner company's crowdvoting system



(a) (b) (c)

products presented to them; Figure 5.2 illustrates the crowdvoting system used by our partner company. Our primary research question is to identify whether, and if so, by how much, the service improves rental-related performance metrics.

Our setting provides an ideal environment to study this research question. First, rentals are becoming increasingly important as consumers seek commitment-free and sustainable options; even mainstream retailers such as Urban Outfitters have started to offer a rental option (D'Arpizio et al. 2021, US Chamber of Commerce 2019). Second, the right balance between the depth and the breadth of the inventory is critical for subscription-based platforms dealing with non-digital goods such as apparel: Consumers primarily subscribe to these platforms to gain access to a vast assortment of products, but a long wait time for accessing products could lead to customer dissatisfaction.[1] While existing literature has mainly studied the

[1]Unlike subscription-based platforms dealing with digital goods such as Netflix, those dealing with non-

impact of crowdsourcing innovation, we are the first to study the impact of crowdsourcing on the operational efficiency of subscription-based platforms. Third, the subscription aspect allows us to observe rich data on users' repeated interactions with the platform.

We obtained a proprietary longitudinal dataset spanning products offered by the platform and their rentals over time, as well as the users on the platform, their rental behavior, and product reviews. The platform launched the game in October 2021, and our dataset includes several months before and after, specifically from April 2021 to July 2022. The platform randomly assigned some portion of their potential products into the game and subsequently based the purchasing decisions for those products on the outcome of the game. The platform also gradually increased the number of items assigned to the game over time. This random assignment helps in estimating the causal impact of the game on the short-term and long-term performance metrics. For those products that were put through the game, we also obtained the voting results from each voting user and the platform's subsequent purchasing decision. Our primary research methodology is a difference-in-differences (DiD) model where the treatment is the launch of an item post-crowdvoting at the product category level, and our outcome of interest is the rental conversion rate. The partner company adopted crowdvoting for different categories in a staggered manner; that is, different categories became treated at different points in time due to the slow rollout of the game over the months since the launch of the game.

Our main contributions are as follows. We find that the rental performance of a product category improves after crowdvoted items become available on the platform; specifically, products are rented 1.84 more times in 14 days and 4.3 more times in 3 months. We performed an item-level analysis, and we found that items purchased after going through the game are rented 1.79 more times in 14 days and 5.05 more times in 3 months, relative to items that are purchased based only on staff input. Furthermore, items that are more favorably rated in the game exhibit even better rental outcomes. These insights offer compelling evidence

---

digital goods such as apparel face unique challenges in assortment planning due to the physical and seasonal nature of these items.

that crowdvoting can enhance a retailer's operational performance. Our findings validate practitioners' recent adoption of such tools and contribute to the academic body of work establishing the impact of crowdvoting in retail.

These results are consistent across several robustness checks. We show that the timing of the game rollout in different categories does not drive the improvement in the rental performance. We verify that the assignment of products into the game was random. We also match items that go through the game and those that do not based on item-level characteristics using propensity score matching; our results on this matched sample show a significant improvement in the number of rentals.

We investigate several mechanisms that drive this improvement in performance. We find that there are significant changes in the way the platform procures inventory after the launch of the game. First, the platform purchases more copies of the same item, increasing the depth of its assortment. Crucially, this improved depth is not at the expense of sacrificing the breadth of the inventory. Second, among crowdvoted items, the inventory depth is higher for items that are more positively voted by the crowd. Further, we find that the items picked by the staff are more likely to be from brands that are already present on the platform. This suggests that crowdvoting plays a crucial role in expanding the inventory by helping the platform stock an otherwise riskier assortment, while the staff input is useful in making decisions regarding more mainstream products. In other words, the two opinions best work in a complementary, rather than substitutable, way. Together, these findings provide evidence that the increase in rentals is attributable to the ability of the retailer to make smarter investments in inventory.

We also explore how users' engagement and satisfaction with the platform change upon participating in the game. Customer satisfaction plays a significant role in subscription-based platforms: a report by McKinsey finds that consumers are more likely to continue subscribing to a platform if they feel "surprised and delighted" and obtain a good value for the money (Chen et al. 2018). We find that users who participate in the game are more likely to engage with the platform by writing more product reviews. This is an important

measure of customer engagement, as writing reviews takes time and effort and is considered purely an altruistic act to inform others. More importantly, we find that the users who participate in the game are less likely to cancel their subscription, and the churn reduces proportionally to their degree of participation in the game. Interestingly, even though the platform does not inform consumers about the outcome of the game, we find evidence that users who are aware of the existence of the products that they voted favorably on are less likely to cancel. This result suggests that crowdvoting can increase consumers' trust in the platform: consumers who believe that their voices were heard find value in continuing to subscribe to the platform.

The rest of the paper is organized as follows. We present our empirical setting and data in §5.2. We present our main econometric models and results in §5.3. We provide robustness checks in §5.4 and study several mechanisms that drive the improvement in the platform's performance in §5.5. We conclude in §5.6.

5.2 Research Setting and Data

Our partner firm is a subscription-based apparel rental platform, where customers can rent items for a flat fee each month. Customers can wear the clothes for up to one month and then exchange them for a new set; the number of items in each set depends on the customer's subscription tier. The returned items get cleaned at the processing center and then re-listed on the platform.

The platform has a dedicated purchasing staff, similar to buyers at a traditional apparel retailer. Based on historical consumption behavior as well as fashion trends, they decide on what items, and how many copies of each item, they should purchase. Merchandising decisions are fairly complicated and error-prone given the idiosyncrasies of fashion (Parker-Strak et al. 2022). Our platform introduced the crowdvoting game to aid them in their purchasing decisions.

Our partner platform presents subscribers with an invitation link to the game on their home page. Users can choose to participate in the game and are shown items as illustrated

in Figure 5.2. Users can choose between "Dislike," "Favorite," or "Skip." The platform collects all the votes from the game and uses this information to make purchasing and inventory decisions. (Figure C.1 in the appendix shows a flowchart of the game.) They use a threshold-based rule – on the total number of positive votes and the fraction of positive votes – to make these decisions.

Our dataset spans over 68 weeks, from April 6, 2021 to July 28, 2022, and includes data before and after the first launch of the game on October 6, 2021. During this period, the platform purchased 5,018 items spanning 52 product categories. Of these items, 2,789 items were purchased after the game was launched. A subset 396 of the items were bought based on users' votes from the game (the rest of the items were purchased by the staff through the traditional purchasing process.) A total 3,192 of items were sent through the game.

The rollout of the game was staggered from October 2021 to July 2022. We consider a product category as treated at the time that the first item in the category is bought after going through the game. Figure 5.3(a) illustrates the rate at which different categories were treated. Out of the 52 product categories (e.g.,'maxi dresses', 'A-line dresses'), 47 categories had at least one item entered into the game, implying that most (but not all) categories eventually got treated. Our conversations with the platform executives revealed that the staggered approach was adopted in order to conduct internal evaluations of the game's impact and efficacy. The decision of which items to include in the game was not strategic, that is, items were randomly chosen to be entered in the game. (We verify this rigorously in §5.4.) We exploit this randomness and the variation in the time at which items were entered into the game in our identification strategy.

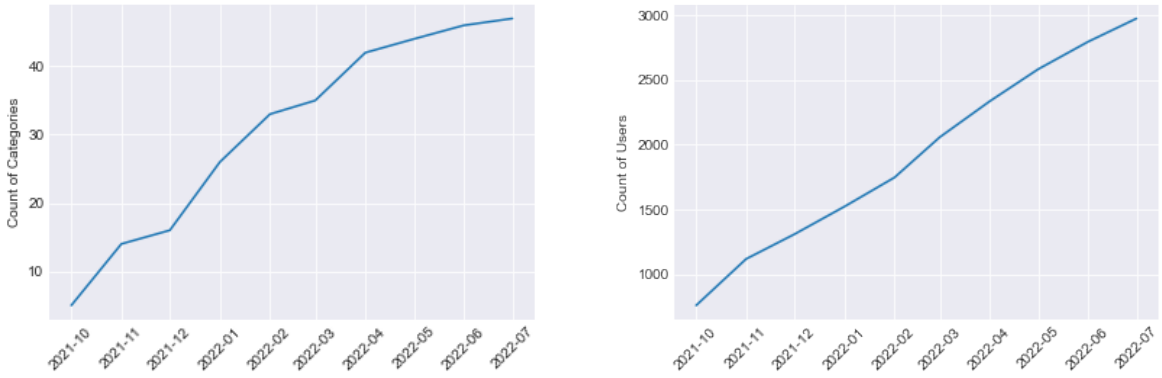
In Table 5.1, we summarize the characteristics of the 2,789 items purchased after the game was launched. Our main outcomes of interest are the average number of rentals for each product category in the first two weeks (14DaysRental) and in the first three months (3MonthsRental), which capture how successful a particular product category is in the short

Table 5.1: Summary Statistics of Items

	Variable Definition	Mean	Std. Dev.	Min.	Max.
Outcome Variables:					
14DaysRental	Count of rentals within the first 14 days item becomes available on the platform	3.90	4.35	0	35
3MonthsRental	Count of rentals within the first 3 months item becomes available on the platform	9.51	10.2	0	87
InventoryLevel	Number of copies of each item	5.37	5.70	1	138
Controls:					
Holiday	Binary variable, indicating whether item is for holiday season	0.62	0.49	0	1
Summer	Binary variable, indicating whether item is for summer season	0.63	0.48	0	1
NeutralColor	Binary variable, indicating whether item is neutral color	0.47	0.50	0	1
ColdColor	Binary variable, indicating whether item is cold color	0.28	0.45	0	1
Relaxed	Binary variable, indicating whether item is relaxed style	0.62	0.49	0	1
Trendy	Binary variable, indicating whether item is trendy style	0.38	0.48	0	1
Dresses	Binary variable, indicating whether item belongs to dresses department	0.36	0.48	0	1
Tops	Binary variable, indicating whether item belongs to tops department	0.31	0.46	0	1
Cotton	Binary variable, indicating whether item is made of cotton	0.36	0.48	0	1
HasPockets	Binary variable, indicating whether item has pockets	0.30	0.46	0	1
LongSleeve	Binary variable, indicating whether item has long sleeves	0.47	0.50	0	1
RetailPrice	Retail price of item	104.8	77.45	13.0	765
TopBrand	Binary variable, indicating whether item is from a popular brand	0.43	0.49	0	1

Note: The statistics for RetailPrice are reported after rescaling due to the nondisclosure agreement.

Figure 5.3: Game Adoption Over Time



(a) At Category Level

(b) At User Level

and long term, respectively.[2] We also observe the purchasing decisions of the platform, including the inventory level of each item.

Our dataset also includes 7,581 unique users who were members of the platform for at least one day after the launch of the game. We observe the entry and exit of users on the platform, as well as all their interactions (including voting, renting, clicking, and reviewing behavior) on the platform. Figure 5.3(b) illustrates the rate at which users start participating in the game. We see that there is significant variation in the adoption of the game, and the number of users who start participating in the game steadily increases throughout our study period. We provide the summary statistics at the user level in Table 5.2. We see that 2,973 of the 7,581 users cast their votes at least once, so the mean user participation rate (denoted by GameParticipation) is 0.39. We also see that the mean number of votes cast by voting users (denoted by GameVotes) is 66. Thus, our rich dataset offers a granular view of the platform and its users, allowing us to study the fundamental impact of crowdvoting.

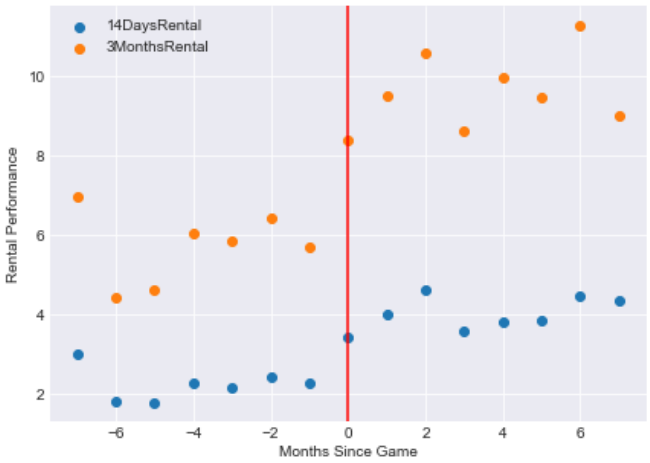
[2]These measures are also used by the platform to evaluate short- and long-term performance.

Table 5.2: Summary Statistics of Users

	Variable Definition	Mean	Std. Dev.	Min.	Max.
Outcome Variables:					
Cancellation	Binary variable, indicating whether the user cancels	0.62	0.49	0	1
ReviewCnt	Count of reviews posted by the user in the month	5.65	6.33	0	50
Treatment:					
GameParticipation	Binary variable, indicating whether user votes in the game	0.39	0.49	0	1
GameVotes	Count of votes by voting users	66.0	239.3	0	2301
GameBuyAware	Binary variable, indicating whether user clicks on the items purchased by the platform that they voted "Favorite" for in the game	0.23	0.42	0	1
Controls:					
Tenure	User's tenure in years	0.64	0.96	0	5.55
RentalCnt	Count of rentals by user	53.2	125.3	0	1247
RatingAvg	User's average rating across all rented items	3.84	0.68	1	5
Age	User's age in years	41.8	9.16	11.8	84.3
Central	Binary variable, indicating whether user is in central area	0.13	0.34	0	1
EastCoast	Binary variable, indicating whether user is in east coast area	0.29	0.45	0	1
WestCoast	Binary variable, indicating whether user is in west coast area	0.39	0.49	0	1
South	Binary variable, indicating whether user is in south area	0.19	0.39	0	1

5.3 *Econometric Model and Main Results*

Figure 5.4: Model-Free Evidence of the Impact of the Game



We start by exploring model-free evidence of the game's impact on rental performance. Figure 5.4 shows the average number of rentals (over 14 days or 3 months) for each item in each product category. We want to see how the rental performance (shown on the Y-axis) changes from the time that a category first includes an item purchased post voting in the game (X-axis). The red line indicates the first month that a category includes such an item. Figure 5.4 indicates a significant increase in the average number of rentals in a category after it acquires items post voting in the game. These findings do not represent causal evidence, but point to a potential improvement in rental performance due to the game. In the rest of this section, we develop rigorous empirical analyses to test for this effect.

5.3.1 Main Specification

We use a difference-in-differences (DiD) model to understand the causal impact of deploying the game on the rental performance. The DiD model has been widely used in the empirical

operations literature (Bell et al. 2018, Levi et al. 2022, Wang et al. 2022). We employ the following specification.

$$\mathrm{Y}_{ct} = \beta_0 + \beta_1 \mathrm{PostGame}_{ct} + \beta_2 \mathrm{ItemCnt}_{ct} + \lambda_t + \mu_c + \delta_t + \mu_c * \delta_t + \epsilon_{ct}, \tag{5.1}$$

where c stands for category and t stands for month. We use category-month panel data to estimate (5.1). We choose to conduct our analysis at the product category level to ensure that we have the appropriate performance measures for the pre-period before the product is launched in our DiD framework.[3] Next, we explain each term in (5.1).

The term Y_{ct} is our main outcome variable, $14\mathrm{DaysRental}_{ct}$ and $3\mathrm{MonthsRental}_{ct}$ corresponding to the short-term and the long-term impacts respectively, calculated as the average over the items launched in product category c in month t. Our treatment variable is denoted by $\mathrm{PostGame}_{ct}$. Specifically, $\mathrm{PostGame}_{ct} = 1$ for all t on or after the month t when a product category c gets its first item via the game. Our coefficient of interest is β_1. In the staggered adoption framework, each untreated product category (the product category that has not obtained any item through the game) serves as a control for the treated category until it gets treated. β_1 thus provides us with the average treatment effect (ATE), which is the average difference of the difference between the average outcomes of the treated and the control product categories in each time period.

Next, we incorporate several additional features in (5.1) to control for observable and unobservable characteristics. The number of rentals could be directly related to the number of distinct items purchased to rent in each category in each month. To account for this, we include $\mathrm{ItemCnt}_{ct}$, which represents the purchased number of items in category c and month t. λ_t is a vector of quarterly fixed effects and μ_c is a vector of category fixed effects. The quarterly fixed effects are used to control for common calendar shocks that affect all categories simultaneously, and the category fixed effects account for time-invariant category level characteristics. δ_t refers to the holiday fixed effects that account for the holiday months,

[3]This is because we cannot observe more granular outcomes such as items' average rental performance before they are bought. Nevertheless, we perform a secondary analyses at the item-level and find that our results are consistent.

namely November and December, during which time the platform usually observes higher demand. We add $\mu_c * \delta_t$ fixed effects to account for the demand shock brought by holidays for specific categories, for example, event/party dresses could experience a higher demand during the holiday months and adding a category-specific time-trend accounts for a flexible decay path in demand for each category. We account for autocorrelation between the errors across time periods by clustering the standard errors at the category level (Bertrand and Mullainathan 2004).

5.3.2 Main Results

We estimate (5.1) for the two outcome variables, 14DaysRental_{ct} and $3\text{MonthsRental}_{ct}$. We present our results in Table 5.3. The coefficient of interest, β_1, is positive and statistically significant for both outcome variables, demonstrating that there is a significant improvement in the average number of rentals after crowdvoted items become available on the platform. On average, the number of rentals increased 1.84 times in two weeks and 4.3 times in three months. Thus, the game leads to improved short- and long-term operational performance for our partner platform. It is noteworthy that this improved performance is not driven simply by more items being available to rent in the treated categories since we controlled for the number of distinct items in each category in (5.1).

We conduct three additional analyses to strengthen these insights. First, the DiD specification requires that there are no measurable differences between the control and the treated groups in the pre-treatment period. In other words, controlling for potential confounders, the difference between the treated and the control categories should be insignificant in the pre-period and significant in the post-period. We test this assumption by modifying specification (5.1), replacing the PostGame_{ct} indicator in (5.1) with a variable T_{cm}. T_{cm} denotes the interval in m months from when the category c first gets treated, and it can range from -7 to 7. 0 is the month at which the category c is treated. We fix the time period corresponding

Table 5.3: Impact of the Game on Rental Performance

	(1)	(2)
	14DaysRental	3MonthsRental
PostGame	1.84***	4.30***
	(0.33)	(0.73)
ItemCnt	0.02	0.02
	(0.01)	(0.03)
Constant	2.08***	5.36***
	(0.16)	(0.40)
Quarter FE	Yes	Yes
Category FE	Yes	Yes
Holiday FE	Yes	Yes
Category*Holiday FE	Yes	Yes
Observations	512	512
R^2	0.369	0.364

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

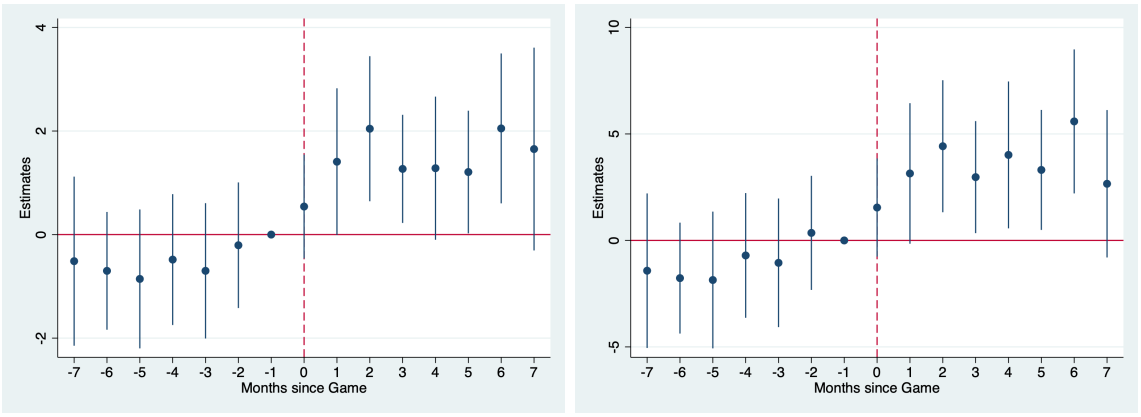
The unit of analysis is category (c) - month (t) level.

to -1 as the baseline in our estimation. This yields:

$$\mathrm{Y}_{ct} = \beta_0 + \sum_{m=-7}^{m=7} \beta_{1m} \mathrm{T}_{cm} + \beta_2 \mathrm{ItemCnt}_{ct} + \lambda_t + \mu_c + \delta_t + \mu_c * \delta_t + \epsilon_{ct}. \tag{5.2}$$

We estimate (5.2) and present the results in Figure 5.5. Specifically, 5.5(a) shows the estimates for the $14\mathrm{DaysRental}_{ct}$ outcome and Figure 5.5(b) shows the same for the $3\mathrm{MonthsRental}_{ct}$ outcome. The coefficients corresponding to the variables in the pre-treatment periods are statistically insignificant. This suggests that prior to getting treated, the rental performance of the treated and control categories are not statistically different from each other. We also observe that the coefficients corresponding to the post-treatment variables are mostly statistically significant and positive. The findings validate the parallel trends assumption crucial for the DiD setting.

Figure 5.5: Evidence of Parallel Trends Assumption for Table 5.3 Results



(a) 14 Days Rental (b) 3 Months Rental

Note. We use 95% confidence interval.

Second, we complement the above analysis with item-level studies. Recall that we estimated the impact of the game at the category-level to explore the change in the rental performance due to the treatment, between the treated and the control product categories. In addition to not observing items' performance before they are bought by the platform, we

also do not observe the performance of the items that went through the game but are not bought by the platform. In other words, we only observe the performance of items that are bought, either post voting in the game or otherwise.

We can, however, leverage the fact that items are randomly entered into the game. Specifically, we can causally measure the impact of the game on the performance of items bought post voting compared to items that are bought without crowdvoting, that is, picked by the staff (purchasing staff who are responsible for procurement decisions). This is essentially a first difference model where we measure the difference in performance between the items bought through the game (treated) and the staff-picked items (control), given that there are no discernible differences between the control group and the treated group due to random assignment. To further establish that there are no differences between the control and treated groups, we perform an analysis in §5.4; the results confirm that the entry into the game was random and that the difference between the items that go through the game and do not go through the game is statistically insignificant. We specify our model as follows:

$$\mathrm{y}_i = \beta_0 + \beta_1 \mathrm{GameBuy}_i + \beta_2 \mathrm{Controls}_i + \mu_c + \lambda_t + \delta_t + \epsilon_i, \tag{5.3}$$

where i stands for item. We use the 2,789 items purchased after the game was launched for this analysis. Similar to that in (5.1), we use the $\mathrm{14DaysRental}_i$ and $\mathrm{3MonthsRental}_i$ outcomes to capture item-level short- and long-term rental performance, respectively. Whereas the outcome measures in (5.1) are the average rental performance measures for the category, the outcome measures in (5.3) are not averaged since they can be measured at the item level. In addition, for each item we define $\mathrm{RentalIndicator}_i$ as a binary dependent variable, capturing whether the item i got rented at least once in the first two weeks of its launch.[4]

$\mathrm{GameBuy}_i$ is the variable of interest, and it is coded as 1 if the item i is purchased after it went through the game and 0 if it is purchased based only on staff input. We also control for 13 features of the items described in Table 5.1. This allows us to ensure that

[4]The partner platform considers this also to be an important measure of success for individual items, in a similar vein to how the opening weekend is often key to a movie's performance.

our comparison is between items purchased in the same time period that have similar styles but are purchased due to input from different sources. Finally, the year-level fixed effects λ_t and month-level fixed effects δ_t account for the common calendar shocks, and μ_c accounts for category-specific characteristics. We cluster standard errors at the category level.

Table 5.4: Impact of the Game on Rental Performance: Item Level Analysis

	(1)	(2)	(3)
	14Days Rental	3Months Rental	RentalIndicator
GameBuy	1.79***	5.05***	0.10***
	(0.30)	(0.73)	(0.02)
Constant	6.44***	14.88***	0.94***
	(0.71)	(1.75)	(0.06)
Controls	Yes	Yes	Yes
Category FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	2789	2789	2789
R^2	0.167	0.170	0.074

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) level.

Our results are presented in Table 5.4. The coefficient of GameBuy$_i$, that is β_1 in (5.3), captures the effect of the game by comparing items purchased from the two different sources (game and staff-picked). On average, items that are purchased by the platform after going through the game are rented 1.79 more times within the first 14 days after they are available on the platform. This effect expands to 5.05 more times within the first 3 months. In other

words, items purchased after going through the game outperform items purchased based only on staff input in the same period. In column (3), we see that this effect persists when focusing on whether an item is rented within the first two weeks: specifically, items purchased after going through the game are 10% more likely to be rented in the first two weeks, further reiterating that crowdvoted items perform better. Thus, our item-level analyses confirm our main findings at the category level.

Third, we can leverage the variation in the number of votes from the voting game and study the magnitude of the rental performance of items bought through the crowdvoting process. In particular, conditional on purchasing an item that goes through the crowdvoting process, we can analyze the difference in performance with respect to the voting results.

Specifically, the voting results likely include information about how well the items are received by the voters, and they could be an effective proxy for the tastes of all potential consumers. With the following model, we seek to explore the impact of crowdvoting on the rental performance among the items that were purchased through crowdvoting:

$$\mathrm{y}_i = \beta_0 + \beta_1 \mathrm{HighFavoritePct}_i + \beta_2 \mathrm{Controls}_i + \mu_c + \lambda_t + \delta_t + \epsilon_i, \tag{5.4}$$

where i stands for item. We use the 396 items purchased through the game to estimate this model.[5] Similar to that in (5.3), we use the $\mathrm{14DaysRental}_i$, $\mathrm{3MonthsRental}_i$, and $\mathrm{RentalIndicator}_i$ as dependent variables. Our variable of interest is $\mathrm{HighFavoritePct}_i$, which we explain next. We compute the percentage of "Favorite" votes received by the item purchased through the game and find that the median is 63.3%. We define items with higher than median "Favorite" count as highly favored items and set $\mathrm{HighFavoritePct}_i$ as 1, and 0 otherwise. We use the same controls as in (5.3).

Our results are presented in Table 5.5; the coefficient of the variable of interest is on the first line. We observe that within game-purchased items, those with a higher share of favorite votes achieve even better rental performance: They are rented 1.5 more times

[5] In the results shown in Table 5.5 and later, few of the items are singleton groups, and they are dropped in order to improve the accuracy of the inference.

within the first 14 days and 3.09 more times within the first 3 months. We notice that more than 90% of game-purchased items get rented at least once in the first 14 days, that is, the RentalIndicator$_i$ variable is 1 for almost all of these items. As a result, we do not observe a significant impact of votes on *whether* the item gets rented within 2 weeks of its launch (i.e., column (3)). To further confirm that votes are aligned with voters' tastes, we conducted a verification of voters' truth telling behavior in Appendix C.4; specifically, we regress the probability that a voter rented an item they voted on after they become aware of its availability on the platform. We found that voters are more (less) likely to rent items they voted favorably (unfavorably) on, confirming this underlying link.

Table 5.5: Impact of the Game Votes on Rental Performance: Item Level Analysis

	(1)	(2)	(3)
	14DaysRental	3MonthsRental	RentalIndicator
HighFavoritePct	1.50**	3.09**	0.01
	(0.57)	(1.51)	(0.02)
Constant	3.75	14.04*	1.03***
	(3.40)	(7.75)	(0.12)
Controls	Yes	Yes	Yes
Category FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	391	391	391
R^2	0.332	0.318	0.207

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) level.

5.3.3 Discussion

The results in Tables 5.3-5.5 present ample evidence that the collective wisdom of the crowd is highly effective, and incorporating this wisdom in the decision-making process significantly improves the platform's operational performance. Specifically, we find that products that are bought through the crowdvoting process perform much better than the ones that are picked by the staff. Further, among the items that are purchased through the crowdvoting game, the ones with higher number of votes perform better than others. The number of rentals is arguably the most important success metric for a subscription-based apparel platform: more rentals per item enable the company to spread the cost of procurement over a larger base, thus improving profitability. We corroborate these insights using several robustness checks in §5.4, including a propensity score matching analysis to address the concern that treated and control groups may be different from each other.[6]

5.4 Robustness Checks

5.4.1 Rollout Process

The platform initiated crowdvoting at different points in time for different categories. One may opine that this was not truly random, and that categories which are trending in terms of rental performance may have been chosen to be entered into the game strategically. We now rerun specification (5.1), keeping everything unchanged except that we add a new binary variable $\text{PostVoteEntry}_{ct}$, which equals 1 for all t on or after the month t when item(s) in category c are first entered into the game. Note that $\text{PostVoteEntry}_{ct}$ is ahead of PostGame_{ct}, since crowdvoted items become available to rent on the platform after the voting period has concluded and the platform has made purchasing decisions. If the coefficient of $\text{PostVoteEntry}_{ct}$ is positive and significant, we would attribute the improved rental performance partially to

[6]It is possible that an increase in the number of users during the study period might contribute to the increase in rentals. We re-run specification (5.1) and (5.3) using rentals from existing users only, where an existing user is defined as a subscriber who is active when the game was first launched. Our results are robust to this alternative specification.

the timing of the treatment in (5.1).

Table 5.6 shows our results. Our main variable of interest, PostGame_{ct}, continues to be positive and statistically significant, whereas the coefficient of $\text{PostVoteEntry}_{ct}$ is not statistically significant, for both our short-term and long-term rental performance outcome variables. This provides evidence that the improvement in rentals is not driven by the rollout schedule.

Table 5.6: Impact of the Game and Rollout on Rental Performance

	(1)	(2)
	14DaysRental	3MonthsRental
PostGame	1.63***	3.74***
	(0.48)	(1.05)
PostVoteEntry	0.32	0.83
	(0.40)	(0.91)
ItemCnt	0.02	0.02
	(0.01)	(0.03)
Constant	1.96***	5.05***
	(0.19)	(0.45)
Quarter FE	Yes	Yes
Category FE	Yes	Yes
Holiday FE	Yes	Yes
Category*Holiday FE	Yes	Yes
Observations	512	512
R^2	0.370	0.366

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is category (c) - month (t) level.

5.4.2 Random Assignment

A random assignment of items into the game is vital to ensure that there was no strategic selection of assortment entered into the game. Although our partner platform assured us that this was the case, we now validate this based on data. In particular, the platform could have entered into the game items whose eventual rental performance they were more uncertain about. Our goal is to establish that there is no measurable difference in how certain the platform was as to the eventual performance of items sent through the game relative to others. To do so, we run the following logistic regression:

$$log\frac{P(\text{Game}_i = 1)}{P(\text{Game}_i = 0)} = \beta_0 + \beta_1 \text{SimilarItemCnt}_i + \lambda_t + \delta_t + \mu_c + \epsilon_{it}. \tag{5.5}$$

Here the outcome variable is Game_i, which is 1 if the item i was sent into the game and 0 if it was picked based on staff input only. Our variable of interest is SimilarItemCnt_i, which serves as a measure of certainty in the eventual performance of an item. We collect all the observable features of the items in the firm's database. This includes color, occasion, style, material and 20 other features. Then, we calculate the cosine similarity based on the 24 features, and we define similar items as those that previously existed on the platform and have a cosine similarity over 0.8. In this way, for each item that was purchased (through the game or through the staff opinion), SimilarItemCnt_i is calculated as the count of most similar items on the platform. Then we include year and month fixed effects, λ_t and δ_t, to account for the common calendar shocks. We add category fixed effect μ_c to account for the unobserved time-invariant factors related to the item's category. β_1 is the coefficient associated with certainty (SimilarItemCnt_i). If the platform has a large number of historical items similar to the focal item, the platform will have a higher level of certainty as to the performance of the focal item. Our goal is to demonstrate that the probability of entering the game is not associated with the level of certainty.

Our results are reported in Table 5.7. If the firm had strategically entered items with more uncertainty into the game, the coefficient corresponding to SimilarItemCnt_i, β_1, will be negative and significant. On the contrary, we find that β_1 is not significant. This indicates

that there are no systematic differences between the items that are entered and not entered in the game.

Table 5.7: Verification of No Systematic Difference between Game Items and Others

	(1)
	Game
SimilarItemCnt	-0.00
	(0.00)
Constant	-0.93
	(0.96)
Controls	Yes
Year FE	Yes
Month FE	Yes
Category FE	Yes
Observations	5564

Standard errors in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) level.

5.4.3 Propensity Score Matching

To further rule out the possibility that there is a systematic difference between the game items and staff-picked items, we use propensity score matching (PSM) (Caliendo and Kopeinig 2008). Specifically, we match items purchased from the game and the items purchased by staff using the control variables employed in (5.3). Figure C.2 in the online appendix shows the improvement in the balance between the treated and control groups after the matching. We then re-run the specification in (5.3) on this matched sample. Reassuringly, our estimates,

shown in Table 5.8, are consistent with our main results. We also use PSM to systematically match the items in the HighFavoritePct$_i$ = 1 bucket and the HighFavoritePct$_i$ = 0 bucket in specification (5.4). We re-run the analysis in (5.4) on this matched sample. The results are shown in Table 5.9; once again, we have consistent results.

Table 5.8: Impact of the Game on Rental Performance: Item Level using PSM

	(1)	(2)	(3)
	14DaysRental	3MonthsRental	RentalIndicator
GameBuy	1.67***	4.84***	0.08**
	(0.37)	(0.91)	(0.03)
Constant	7.02***	17.51***	1.08***
	(1.27)	(3.23)	(0.14)
Controls	Yes	Yes	Yes
Category FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	789	789	789
R^2	0.291	0.287	0.137

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) level.

5.5 Mechanisms

We now explore several mechanisms that lead to the improved rental performance reported in §5.3.2: inventory and assortment decisions in §5.5.1, user engagement in §5.5.2, and user loyalty in §5.5.3.

Table 5.9: Impact of the Game Votes on Rental Performance: Item Level using PSM

	(1)	(2)	(3)
	14DaysRental	3MonthsRental	RentalIndicator
HighFavoritePct	2.10***	5.18***	0.02
	(0.61)	(1.39)	(0.03)
Constant	4.12	14.05	0.89***
	(4.11)	(10.28)	(0.16)
Controls	Yes	Yes	Yes
Category FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	388	388	388
R^2	0.376	0.365	0.300

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) level.

5.5.1 Inventory and Assortment Decisions

As highlighted in §1, the main challenge in fashion retailing is the inability to predict how items will fare in the selling season. This introduces significant uncertainty in the procurement process (Ren et al. 2017). If crowdvoting is useful in predicting demand better, the firm may use this to their advantage when making inventory and assortment decisions.

We develop a stylized model of inventory management in Appendix C.1, incorporating two key features: (1) uncertainty in the number of users interested in renting a product and (2) the rental business model. This model demonstrates that the platform's threshold-based rule for buying crowdvoted products – based on the number or share of positive votes – is in the right direction: Such a decision-making process leads the platform to pick products that are likely to have a robust rental potential and justifies the upfront costly purchasing decision. Second, the model presents a plausible mechanism for why crowdvoting leads to an improved overall rental performance in §5.3.2—specifically, the increase in a crowdvoted product's inventory level. This is our first hypothesis: does crowdvoting lead to an overall increase in the inventory level of crowdvoted products relative to those that are purchased based on staff input alone? Third, the magnitude of the improvement in inventory decisions is predicted to be even more pronounced for more positively voted products. This leads to our second hypothesis: for crowdvoted items, is the inventory level higher for more positively voted products?

Inventory Depth

We use two models to test our hypotheses: Model 5.6 estimates the impact of the game on the overall inventory level while Model 5.7 estimates the impact of positive votes on the inventory level.

$$\text{InventoryLevel}_i = \beta_0 + \beta_1 \text{GameBuy}_i + \beta_2 \text{Controls}_i + \mu_c + \lambda_t + \delta_t + \epsilon_i, \quad (5.6)$$

$$\text{InventoryLevelDiff}_i = \beta_0 + \beta_1 \text{HighFavorite}_i + \beta_2 \text{Controls}_i + \mu_c + \lambda_t + \delta_t + \epsilon_i. \quad (5.7)$$

In (5.6), InventoryLevel$_i$ represents the number of copies of each item i (inventory depth) purchased. Similar to (5.3), GameBuy$_i$ is our variable of interest; we control for 13 features of the items described in Table 5.1; and we add fixed effects as in (5.3). We estimate Model 5.6 and its counterpart after matching items purchased from the game and by staff using PSM in Table 5.10 in columns (1) and (2), respectively.

In (5.7), we switch our focus to the 396 items purchased through the game. Our variable of interest is HighFavorite$_i$. We first compute the median of the count of favorite votes each item receives, namely 128.5. We then define HighFavorite$_i$ as 1 if item i has more than 128.5 votes; otherwise, it is 0. We use InventoryLevelDiff$_i$ as our outcome variable, which captures the extent of increase in the inventory level for the game item i relative to what would have been in the absence of crowdvoting. As in §5.4.2, we use the cosine similarity score to find similar items to our focal items from the game. For each focal item i, we find the most similar prior item $\hat{i}$, that is, the item with the highest similarity score purchased before item i. We consider the InventoryLevel$_{\hat{i}}$ as the baseline for item i. Then, InventoryLevelDiff$_i$ = InventoryLevel$_i$ − InventoryLevel$_{\hat{i}}$. To check the robustness of this analysis, we also define item $\hat{i}$ as a set of items with a similarity score higher than 0.8 (as in §5.4.2) and use their average InventoryLevel$_{\hat{i}}$ as the baseline. The rest of the specification is as in (5.6). The results are shown in Table 5.10 in columns (3) and (4), respectively.

In column (1), we observe that the depth of the inventory significantly increases for items that were bought through the game compared to those picked by staff. In particular, among the 2,789 items purchased after the game is launched, the platform purchases, on average, 2.96 more copies of each style of the items post crowdvoting. The results extend to using the PSM-matched sample as seen in column (2). This provides evidence for our first hypothesis: the game ameliorates the uncertainty in the procuring process. In particular, the platform can now justify procuring inventory that is closely aligned with customer preferences.[7] One

[7]One may intuit that the platform buys cheaper items after crowdvoting to stay within their budget while increasing depth. In contrast to this intuition, we find that the average price paid for items bought post crowdvoting is not significantly different from that for items bought solely through staff opinion; see Appendix C.4 for details.

Table 5.10: Impact on Inventory Depth

	(1)	(2)	(3)	(4)
	InventoryLevel	InventoryLevel (PSM)	InventoryLevelDiff	InventoryLevelDiff
GameBuy	2.959***	3.310***		
	(0.472)	(0.739)		
HighFavorite			2.052***	1.952***
			(0.665)	(0.677)
Constant	10.59***	12.26***	2.318	2.588
	(0.882)	(2.102)	(2.775)	(3.760)
Controls	Yes	Yes	Yes	Yes
Category FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	2788	789	363	363
R^2	0.151	0.226	0.283	0.282

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) level.

Column (3) uses the most similar item, and column (4) uses the average over a set of similar items, as the baseline.

benefit of our subscription rental setting is that the number of units rented by each user depends on their subscription tier. This ensures that the improved rental performance in §5.3.2 is not driven by the fact that each user can rent more items post launch of the game simply because there are more copies of items; we verify this claim in Appendix C.4.

We also observe that the platform is able to carefully adjust the inventory decisions in line with the voting results, as suggested by our theoretical model. Specifically, columns (3)-(4) reveal that the platform can increase the inventory depth of highly favored products by about 2 more units as compared to the less favored products, thus verifying our second hypothesis.

The increased depth could be too extreme a reaction if each individual copy of the item is rented fewer number of times. In other words, it is important to establish how the platform's depth choice impacts inventory turns: If the number of turns becomes lower, the investment in depth may be risky and may not be profit-improving for the platform. To do so, we re-run the specification in (5.3) with the outcome changed to InventoryTurns_i. InventoryTurns_i is defined as the rental count per copy within the first 6 months of when the item i is available on the platform.[8] As shown in Table 5.11, crowdvoted items turn 0.23 more times compared to staff-picked items. This result further verifies our hypothesis that crowdvoting better predicts customers' tastes and picks out more popular items. Importantly, the platform not only makes the right inventory depth decision, but also enjoys increased utilization of those purchased items.

Inventory Breadth

Improving inventory decisions is particularly salient for subscription-based business models as their success hinges on maintaining the right balance between the depth and breadth of products. Typically, an increase in depth comes at the cost of a decrease in breadth (Mahmoud 2022). To understand if the platform is investing in the inventory depth by

[8] We chose 6 months to ensure the seasonality of the item is appropriately accounted for. Our results are robust to defining InventoryTurns over a 3-month time horizon.

Table 5.11: Impact on Inventory Turns

	(1)
	InventoryTurns
GameBuy	0.23***
	(0.08)
Constant	0.92**
	(0.45)
Controls	Yes
Category FE	Yes
Year FE	Yes
Month FE	Yes
Observations	2788
R^2	0.074

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) level.

sacrificing the breadth, we use the following models.

$$\text{ItemCnt}_{ct} = \beta_0 + \beta_1 \text{PostGame}_{ct} + \lambda_t + \mu_c + \delta_t + \mu_c * \delta_t + \epsilon_{ct}, \quad (5.8)$$

$$\text{BrandCnt}_{ct} = \beta_0 + \beta_1 \text{PostGame}_{ct} + \lambda_t + \mu_c + \delta_t + \mu_c * \delta_t + \epsilon_{ct}. \quad (5.9)$$

We modify (5.1) to capture ItemCnt_{ct}, which indicates the number of unique styles of items purchased in category c in month t, in (5.8). Similarly, we capture BrandCnt_{ct}, the number of unique brands spanned by items purchased in category c in month t, in (5.9).

Table 5.12: Impact on Inventory Breadth

	(1)	(2)
	ItemCnt	BrandCnt
PostGame	0.33	-0.18
	(1.28)	(0.68)
Constant	7.50***	5.24***
	(0.64)	(0.34)
Quarter FE	Yes	Yes
Category FE	Yes	Yes
Holiday FE	Yes	Yes
Category*Holiday FE	Yes	Yes
Observations	512	512
R^2	0.627	0.646

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is category (c) - month (t) level.

Table 5.12 reveals that there is no significant change in the number of distinct items or the number of distinct brands procured by the platform. Together with Table 5.10, this

result presents a counter-intuitive dimension to product assortment decision: the platform is able to improve the depth *without* sacrificing the overall number of distinct options available for users.

Brand Mix

Users seek subscription-based platforms primarily for the product variety therein (Chen et al. 2018). Although we noted no significant change to the breadth of items carried, the platform may be offering a different mix of products than before they utilized crowdvoting.[9] We now explore whether items bought post crowdvoting are different from those bought based on staff input using the following logistic regression:[10]

$$log\frac{P(\text{GameBuy}_i = 1)}{P(\text{GameBuy}_i = 0)} = \beta_0 + \beta_1 \text{NewBrand}_i + \beta_2 \text{Controls}_i + \mu_c + \lambda_t + \delta_t + \epsilon_i. \quad (5.10)$$

In (5.10), GameBuy_i is 1 when the items are crowdvoted and 0 if staff-picked. NewBrand_i is 1 when the brand of item i is new to the platform. This metric captures the experience that either staff or users have had with the brand on the platform. We add year- and month-level fixed effects (λ_t and δ_t) to account for seasonal trends and category-level fixed effects μ_c for category specific characteristics.

Table 5.13 shows our results, and we see that the coefficient corresponding to NewBrand_i is significant and negative. This suggests that items that are crowdvoted tend to be from newer brands on the platform compared to items that are staff-picked. In other words, the staff tend to intuitively use prior experience and historical information available to the platform to buy products from familiar brands, while users help choose avant-garde items. Thus, crowdvoted items are likely increasing the diversity of the category. Prior studies indicate that there is no difference in the performance of crowd-sourced and expert-sourced information (Mollick and Nanda 2016). Interestingly, our result demonstrates that there is a

[9]While the assignment of the items to the game was random and in a staggered rollout fashion, all the items that were sent through the game were not bought by the platform.

[10]We log transform variables, such as BrandRentalCnt in (5.10), which have a long tail distribution.

Table 5.13: Impact on Brand Mix

	(1)
	GameBuy
NewBrand	0.91***
	(0.27)
Constant	-2.13
	(1.29)
Controls	Yes
Category FE	Yes
Year FE	Yes
Month FE	Yes
Observations	2770

Standard errors in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) level.

significant difference between the crowdvoted items and the staff-picked items, and the firm is able to reap the benefits of strategically employing this difference.

5.5.2 User Engagement

While the game provides an opportunity for the platform to leverage its users' input, they do not compensate users for the same. In other words, users vote for no apparent reward. On the contrary, users may find it cumbersome to repeatedly cast votes and reduce their engagement with the platform after their participation in the game. Therefore, we explore how the game impacts users' engagement on the platform by leveraging the user-month level data of 2,973 users who cast at least one vote in the game.[11] To measure engagement, we employ the number of product reviews written on the platform by each user. Our partner platform requires all users to numerically rate a product after rental; writing a product review, however, is an optional additional task. Since writing reviews requires more effort, and it is considered an altruistic act to help other users (Yoo and Gretzel 2008), the number of reviews written by each user constitutes a good measure of user engagement on the platform. In other words, conditional on playing the game, we investigate changes in users' engagement in terms of writing product reviews.

As discussed in §5.2, there is a significant variation in when the users started playing the game during the study period (see Figure 5.3b). This enables us to adopt a DiD framework to understand the causal impact of the game on the users' engagement using the following models.

$$\text{ReviewCnt}_{ut} = \beta_0 + \beta_1 \text{PostGameParticipation}_{ut} + \beta_2 \text{ActiveDayCnt}_{ut} + \lambda_t + \delta_t + \mu_u + \epsilon_{ut}, \quad (5.11)$$

$$\text{ReviewCnt}_{ut} = \beta_0 + \beta_1 \text{GameVotes}_{ut} + \beta_2 \text{ActiveDayCnt}_{ut} + \lambda_t + \delta_t + \mu_u + \epsilon_{ut}. \quad (5.12)$$

In (5.11), ReviewCnt_{ut} is the dependent variable and represents the number of reviews posted by user u in month t. We use ActiveDayCnt_{ut} to control for the days in month t

[11] We only consider users who play the game at least once in our analysis because users who play the game could engage differently with the platform compared to the users who do not play the game.

that a user u is active on the platform. Our variable of interest is $\text{PostGameParticipation}_{ut}$. It equals 1 after user u participates in the game in month t. In (5.12), we further analyze how intensity of participating in the game affects their engagement. Specifically, we change the variable of interest from $\text{PostGameParticipation}_{ut}$ to GameVotes_{ut}, which indicates the number of votes user u casts in month t. We add user-level fixed effects μ_u to account for unobserved time-invariant factors among users. Like other specifications, we add year-level fixed effects λ_t and month-level fixed effects δ_t.[12]

Our results are shown in Table 5.14. We observe that the coefficient corresponding to $\text{PostGameParticipation}_{ut}$ is positive and significant: users post, on average, 1.1 more reviews after they start participating in the game compared to before. Further, column (2) indicates that the users who cast more votes are more likely to write reviews.

We also hypothesize that if users become aware that the platform purchased the items they voted favorably on, they are more engaged with the platform. As seen in Figure 5.2, users cannot see the vote results, and neither are they notified if the items they voted "Favorite" on are purchased by the platform. However, we can identify if the users clicked on the items that they voted "Favorite" on after the item is purchased by the platform when they encounter these items later (e.g, on the product pages, via the recommendation system, or through the community feed). In other words, we can measure the level of "awareness" as it relates to the presence of items on the platform that each user had voted on.[13] To test this hypothesis, we modify the model in (5.12). We focus on the user-month level data of 2,973 users who participate in the game and keep their post-participation panel. Apart from GameVotes_{ut} in (5.12), we also add GameBuyAware_{ut}, which is 1 if user u is aware of the presence of the items they vote favorably in month t. We present the results in column (3) of Table 5.14. We find that users who are aware of the products they voted favorably on are

[12]We show that the parallel trends assumption is valid in our DID setting in Appendix C.4. Specifically, we find that there is no significant difference in the reviewing behavior among users who start playing the game earlier and those who play the game at a later stage.

[13]It is plausible that users might view but not click on items they had voted on. Using clicks instead of views in our analyses thus allows our estimates to be on the conservative side.

more likely to write more reviews, confirming our hypothesis.

Our findings provide evidence that crowdvoting leads to increased engagement from the platform's users. This is noteworthy since prior work on crowdvoting cautions that systems that offer no reward to voters are unlikely to generate a significant value (Marinesi and Girotra 2013). On the contrary, we find that in addition to the improved decision-making by the platform, seeking users' input presents the added bonus of empowering users to invest in the platform's growth. This is in line with Füller et al. (2009) who find evidence that including users in the co-creation of a new product can improve their engagement. Our analyses highlights an important opportunity for the platform to exploit: Showing voters that their favored products are available on the platform can boost engagement.

5.5.3 User Loyalty

One of the biggest challenges for subscription platforms is reducing consumer churn: Chen et al. (2018) note that consumers quickly cancel services that do not deliver superior end-to-end experiences. Subscription platforms are constantly seeking strategies to extend user loyalty, since the cost of replacing lost subscribers is often high. We now explore if the game impacts user retention on the platform. Specifically, does the improved operational performance and users' co-ownership in the decision-making process translate to fewer cancellations? We build the next two logistic regressions to study users' cancellation rate:

$$log\frac{P(\text{Cancellation}_u = 1)}{P(\text{Cancellation}_u = 0)} = \beta_0 + \beta_1 \text{GameParticipation}_u + \beta_2 \text{Tenure}_u + \beta_3 \text{Log}(\text{RentalCnt})_u + \beta_4 \text{RatingAvg}_u + \epsilon_u, \tag{5.13}$$

$$log\frac{P(\text{Cancellation}_u = 1)}{P(\text{Cancellation}_u = 0)} = \beta_0 + \beta_1 \text{Log}(\text{GameVotes})_u + \beta_2 \text{GameBuyAware}_u + \beta_3 \text{Tenure}_u + \beta_4 \text{Log}(\text{RentalCnt})_u + \beta_5 \text{RatingAvg}_u + \epsilon_u. \tag{5.14}$$

In (5.13), we use the 7,581 users during our study period. Cancellation_u is the outcome variable which equals 1 when user u cancels the platform's service at some point in the study period. $\text{GameParticipation}_u$ is our variable of interest. We also control for Tenure_u, the tenure of user u, $\text{Log}(\text{RentalCnt})_u$, the logarithmic total rental count of user u, and

Table 5.14: Impact on User Engagement

	(1)	(2)	(3)
	ReviewCnt	ReviewCnt	ReviewCnt
PostGameParticipation	1.10***		
	(0.13)		
Log(GameVotes)		0.20***	0.20***
		(0.03)	(0.03)
GameBuyAware			0.43**
			(0.18)
ActiveDayCnt	0.49***	0.47***	0.51***
	(0.02)	(0.02)	(0.02)
Constant	2.44***	3.04***	2.27***
	(0.12)	(0.10)	(0.12)
Year FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
User FE	Yes	Yes	Yes
Observations	18276	18276	12926
R^2	0.689	0.689	0.709

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is user (u) - month (t) level.

RatingAvg_u, the cumulative rating average of user u to account for the users' heterogeneous experiences with the platform. We also control for other demographic variables such as the location and the age (see Table 5.2).

It is plausible that users who actively vote more and users who are aware that the platform purchased items they voted favorably on are less likely to cancel. In (5.14), we focus on the 2,973 users who participate in the game and study how the number of votes cast and subsequent awareness of voted products affects the likelihood of cancellation. Specifically, we replace $\text{GameParticipation}_u$ with Log(GameVotes)_u, the logarithm of the votes cast by user u, and GameBuyAware_u, the number of clicks on crowdvoted items, to measure awareness as in §5.5.2.

We display the results of models (5.13) and (5.14) in Table 5.15. Our hypotheses are supported with negative and significant coefficients for $\text{GameParticipation}_u$ in column (1) and Log(GameVotes)_u and GameBuyAware_u in column (2). In particular, column (1) shows that users who participate in the game are less likely to cancel than those who do not (while controlling for other factors that might influence cancellation rates such as usage and the rating of the products that they rent). Furthermore, conditional on participating in crowdvoting, users who vote more are more likely to stay on the platform, corroborating the fact that participation in the game leads to improved retention. The negative and significant coefficient for GameBuyAware_u suggests that when users are aware of the items that they voted on, they are less likely to cancel. This reiterates the benefit of revealing the outcome of the game to users: In addition to improving engagement, it can help reduce churn.

5.6 Conclusion

Retailers are increasingly seeking innovative strategies to gain a competitive edge. One such strategy is the use of crowdvoting to understand which products are likely to perform well prior to making costly purchasing decisions. To the best of our knowledge, this paper is the first to provide rigorous evidence of the ability of retailers to improve their operational performance through crowdvoting. Our results reveal that the launch of crowdvoting leads

Table 5.15: Impact on User Cancellation

	(1)	(2)
	Cancellation	Cancellation
GameParticipation	-0.28***	
	(0.06)	
Log(GameVotes)		-0.12**
		(0.05)
GameBuyAware		-0.43***
		(0.16)
Tenure	-0.99***	-0.44***
	(0.05)	(0.10)
Log(RentalCnt)	-0.41***	-0.82***
	(0.03)	(0.06)
RatingAvg	-0.25***	-0.45***
	(0.05)	(0.08)
Constant	3.15***	5.29***
	(0.24)	(0.47)
Controls	Yes	Yes
Observations	7581	2973

Standard errors in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is user (u) level.

to a significant increase in the rental conversion rate. This insight uncovers the fundamental impact of crowdvoting as the improvement arises in the absence of a reward to voters (e.g., pricing discounts or first dibs on crowdvoted products) and without the retailer explicitly revealing to consumers which products were purchased through crowdvoting.

We attribute the improvement in the platform's performance to two key mechanisms. We show that the platform is able to go deeper and richer in its product assortment. We believe this is driven by their ability to reduce demand uncertainty and make better inventory decisions. We also find that users engage more with the platform after the launch of crowdvoting. Interestingly, the users also become more loyal and are less likely to cancel their subscriptions. Together, these mechanisms address subscription-based platforms' two biggest pain points, namely, providing a diverse, deep product assortment and ensuring user retention.

Our work also sheds light on interventions that practitioners may adopt to further enhance their operational performance. Although our partner firm randomly assigned products to the crowdvoting game, our results suggest that a platform can be more strategic in this phase of the process. Specifically, we showed that products purchased based on expert opinion are more likely to be of familiar brands. This suggests that the voter base is extremely helpful to make inventory-related decisions on newer items, and the platform may want to carefully combine their expert opinion with users' votes when it comes to newer brands. Our finding that users become more loyal when they know that the platform purchased the items they voted for suggests that the platform may benefit from "closing the loop," for example, by letting the users know their votes were useful in making a purchasing decision and/or adapting the product recommendation algorithms to include products that a user voted favorably on.

There are several directions for future research on this topic. Some crowdvoting platforms may consider providing price discounts for voters, and it will be valuable to study how this feature impacts votes and the quality of the retailer's decision-making. Figures 5.1-5.2 showed that whereas some platforms reveal the number of positive and overall votes to consumers,

others do not. One may explore how the visibility into a product's votes influences users' voting behavior. Another interesting avenue for future work is to identify how a retailer should make purchasing decisions based on the number of overall and positive votes. We hope that our paper inspires work along these lines.

Chapter 6

CONCLUDING REMARKS

This thesis offers a multifaceted exploration of the impact of digital transformation on businesses and society, contributing to a nuanced understanding of the challenges and opportunities presented by the ongoing digital revolution. By investigating the role of AI in customer care, the influence of UGC on informational accessibility and inclusivity, and the effect of crowdsourcing on operational efficiency, this thesis provides valuable insights into the complex and multifaceted nature of digital transformation.

The chapter on emotional AI in customer care underscores the importance of transparency and the potential pitfalls of non-disclosure in AI-mediated interactions. It highlights the complex interplay between AI identity cues, customer behavior, and satisfaction, emphasizing the need for responsible AI deployment in customer-facing roles. This essay contributes to the growing body of research on the social and behavioral implications of AI, providing valuable insights for businesses seeking to leverage AI in customer care while maintaining trust and transparency. It underscores the importance of designing AI systems that prioritize transparency, accountability, and ethical considerations alongside efficiency and performance.

Chapter 4 on informational accessibility and inclusivity on UGC platforms sheds light on the issue of underrepresentation of plus-sized users in online reviews. It emphasizes the role of privacy concerns stemming from societal stigma in driving this disparity and the need for platforms to actively foster inclusivity to mitigate informational asymmetry. This chapter contributes to the discourse on digital inclusion, highlighting the importance of designing platforms that encourage participation from diverse user groups and promote equity in the digital sphere. It underscores the critical role of UGC platforms in shaping the informational landscape and the need for proactive measures to ensure that the benefits of

digital transformation are distributed equitably.

Chapter 5 on crowdsourcing in e-commerce assortment planning demonstrates the transformative potential of leveraging the wisdom of the crowd in improving operational efficiency and user engagement. It showcases how incorporating customer preferences through crowdvoting can help retailers make better inventory decisions and enhance customer satisfaction in digital retail environments. This chapter contributes to the literature on retail operations and customer engagement, providing valuable insights for businesses seeking to harness the power of crowdsourcing in the digital age. It highlights the importance of engaging customers in the value creation process and leveraging collective intelligence to drive innovation and operational excellence.

Together, this thesis underscores the importance of considering the social and behavioral implications of technological innovations in the context of digital transformation. They highlight the need for businesses to develop strategies that prioritize transparency, inclusivity, and customer-centricity alongside efficiency and innovation. As digital transformation continues to reshape the business landscape and society at large, it is crucial to approach it with a holistic perspective that balances technological progress with social responsibility.

This thesis opens up several avenues for future research. The findings related to AI identity cues in customer care could be extended to other domains and contexts, exploring the generalizability of the insights gleaned from this chapter. The issue of informational accessibility and inclusivity on UGC platforms warrants further investigation, particularly in terms of designing inclusive features and policies that promote participation from marginalized groups. The impact of crowdsourcing on operational decisions and user engagement could be studied in different retail settings and business models, examining the boundary conditions and contextual factors that shape its effectiveness.

As digital transformation continues to unfold at an unprecedented pace, it is imperative for researchers to keep pace with its evolving implications. This thesis underscores the importance of adopting a multidisciplinary approach to studying the impact of digital transformation, drawing insights from fields such as information systems, marketing, oper-

ations management, and social sciences. By fostering cross-disciplinary collaboration and engaging in rigorous empirical research, scholars can contribute to a more comprehensive understanding of the challenges and opportunities presented by digital transformation.

Moreover, this thesis highlights the need for businesses, policymakers, and researchers to engage in ongoing dialogue and collaboration in order to navigate the complexities of the digital age. By fostering open communication, knowledge sharing, and collaborative problem-solving, stakeholders can work together to develop strategies and policies that maximize the benefits of digital transformation while mitigating its risks and challenges. This requires a commitment to continuous learning, adaptation, and innovation, as well as a willingness to engage with diverse perspectives and experiences.

As we navigate the digital age, it is crucial to continue exploring the impact of digital transformation, adapting our research agendas and business strategies to keep pace with its evolving implications, and working towards a future that harnesses the power of technology for the benefit of all. This requires a commitment to ethical and responsible innovation, a focus on customer-centricity and inclusivity, and a willingness to engage in ongoing dialogue and collaboration across disciplinary and sectoral boundaries.

Ultimately, the impact of digital transformation will be shaped by the choices and actions of businesses, policymakers, researchers, and individuals alike. By approaching this transformation with a sense of purpose, responsibility, and empathy, we can work towards a future in which the benefits of technology are realized while its risks and challenges are effectively mitigated. This thesis serves as a call to action, urging stakeholders to engage in the critical work of understanding and shaping the impact of digital transformation, and to do so with a commitment to social responsibility, equity, and inclusivity. Only by working together can we build a digital future that truly benefits all members of society.

BIBLIOGRAPHY

Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association* 105(490):493–505.

Aboody D, Lev B (2000) Information asymmetry, r&d, and insider gains. *The journal of Finance* 55(6):2747–2766.

Abrams D, Hogg MA (2006) *Social identifications: A social psychology of intergroup relations and group processes* (Routledge).

Acquisti A, Gross R (2006) Imagined communities: Awareness, information sharing, and privacy on the facebook. *International workshop on privacy enhancing technologies*, 36–58 (Springer).

Acquisti A, Taylor C, Wagman L (2016) The economics of privacy. *Journal of economic Literature* 54(2):442–492.

Akerlof GA (1970) The market for "lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics* 84(3):488–500.

Ales L, Cho SH, Körpeoğlu E (2017) Optimal award scheme in innovation tournaments. *Operations Research* 65(3):693–702.

Ananthakrishnan U, Proserpio D, Sharma S (2020) Does quality improve with customer voice? Evidence from the hotel industry. *Proceedings of the 21st ACM Conference on Economics and Computation*, 461–461.

Ananthakrishnan U, Proserpio D, Sharma S (2023) I hear you: Does quality improve with customer voice? *Marketing Science* 42(6):1143–1161.

Andrade EB, Ho TH (2009) Gaming emotions in social interactions. *Journal of Consumer Research* 36(4):539–552.

Angrist JD, Pischke JS (2009) *Mostly harmless econometrics: An empiricist's companion* (Princeton university press).

Araman VF, Caldentey R (2016) Crowdvoting the timing of new product introduction. Available at SSRN 2723515.

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

Armstrong CS, Core JE, Taylor DJ, Verrecchia RE (2011) When does information asymmetry affect the cost of capital? *Journal of accounting research* 49(1):1–40.

Athey S, Wager S (2021) Policy learning with observational data. *Econometrica* 89(1):133–161.

Autor DH (2003) Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of labor economics* 21(1):1–42.

Bartels LK, Nordstrom CR (2013) Too big to hire: Factors impacting weight discrimination. *Management Research Review* 36(9):868–881.

Bélanger F, Crossler RE (2011) Privacy in the digital age: a review of information privacy research in information systems. *MIS quarterly* 1017–1041.

Belavina E, Girotra K, Kabra A (2017) Online grocery retail: Revenue models and environmental impact. *Management Science* 63(6):1781–1799.

Bell DR, Gallino S, Moreno A (2018) Offline showrooms in omnichannel retail: Demand and operational benefits. *Management Science* 64(4):1629–1651.

Berry LL (2000) Cultivating service brand equity. *Journal of the Academy of marketing Science* 28:128–137.

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics* 119(1):249–275.

Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94(4):991–1013.

Bindroo V, Mariadoss BJ, Pillai RG (2012) Customer clusters as sources of innovation-based competitive advantage. *Journal of International marketing* 20(3):17–33.

Bissell K, Hays H (2010) Understanding anti-fat bias in children: The role of media and appear-

ance anxiety in third to sixth graders' implicit and explicit attitudes toward obesity. *Mass Communication and Society* 14(1):113–140.

Blomqvist Å, Léger PT (2005) Information asymmetry, insurance, and the decision to hospitalize. *Journal of Health Economics* 24(4):775–793.

Bogár N (2019) *The fashion industry and eating disorders: the dangers of the catwalk* (Cambridge Scholars Publishing).

Brewis AA, Wutich A, Falletta-Cowden A, Rodriguez-Soto I (2011) Body norms and fat stigma in global perspective. *Current anthropology* 52(2):269–276.

Brody LR, Hall JA (2008) Gender and emotion in context. *Handbook of emotions* 3:395–408.

Burke M, Bergquist LF, Miguel E (2019) Sell low and buy high: arbitrage and local price effects in Kenyan markets. *The Quarterly Journal of Economics* 134(2):785–842.

Burtch G, Carnahan S, Greenwood BN (2018a) Can you gig it? an empirical examination of the gig economy and entrepreneurial activity. *Management Science* 64(12):5497–5520.

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

Burtch G, Ghose A, Wattal S (2016) Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding. *Information Systems Research* 27(3):478–496.

Burtch G, He Q, Hong Y, Lee D (2022) How do peer awards motivate creative content? experimental evidence from reddit. *Management Science* 68(5):3488–3506.

Burtch G, Hong Y, Bapna R, Griskevicius V (2018b) Stimulating online reviews by combining financial incentives and social norms. *Management Science* 64(5):2065–2082.

Cabral L, Li L (2015) A dollar for your thoughts: Feedback-conditional rebates on ebay. *Management Science* 61(9):2052–2063.

Cachon GP (2004) The allocation of inventory risk in a supply chain: Push, pull, and advance-purchase discount contracts. *Management science* 50(2):222–238.

Cachon GP, Feldman P (2011) Pricing services subject to congestion: Charge per-use fees or sell subscriptions? *Manufacturing & Service Operations Management* 13(2):244–260.

Caliendo M, Kopeinig S (2008) Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22(1):31–72.

Callaway B, Sant'Anna PH (2021) Difference-in-differences with multiple time periods. *Journal of econometrics* 225(2):200–230.

Callaway B, Sant'Anna PH (2023) Introduction to did with multiple time periods. https://bcallaway11.github.io/did/articles/multi-period-did.html.

Campagna RL, Mislin AA, Kong DT, Bottom WP (2016) Strategic consequences of emotional misrepresentation in negotiation: The blowback effect. *Journal of Applied Psychology* 101(5):605.

Campbell C, Deacon H (2006) Unravelling the contexts of stigma: from internalisation to resistance to change.

Carels RA, Rossi J, Solar C, Selensky JC (2018) Changes in perceived mate value and weight bias associated with former obesity status. *Stigma and Health* 3(4):338.

Caro F, Martinez-de Albeniz V, Apaolaza B (2021) The value of online interactions for store execution. IESE Business School Working Paper.

Caruelle D, Shams P, Gustafsson A, Lervik-Olsen L (2022) Affective computing in marketing: practical implications and research opportunities afforded by emotionally intelligent machines. *Marketing Letters* 33(1):163–169.

Chan J, Wang J (2018) Hiring preferences in online labor markets: Evidence of a female hiring bias. *Management Science* 64(7):2973–2994.

Chatterjee P (2001) Online reviews: Do consumers use them? *Advances in Consumer Research* 28:129–133.

Chen L, Xu P, Liu D (2020) Effect of crowd voting on participation in crowdsourcing contests. *Journal of Management Information Systems* 37(2):510–535.

Chen T, Fenyo K, Yang S, Zhang J (2018) Thinking inside the subscription box: New research on e-commerce consumers. *McKinsey & Company* 1–9.

Chen Y, Harper FM, Konstan J, Li SX (2010) Social comparisons and contributions to online communities: A field experiment on MovieLens. *American Economic Review* 100(4):1358–1398.

Chen Y, Xie J (2005) Third-party product review and firm marketing strategy. *Marketing science* 24(2):218–240.

Chesbrough HW (2003) *Open innovation: The new imperative for creating and profiting from technology* (Harvard Business Press).

Chevalier JA, Dover Y, Mayzlin D (2018) Channels of impact: User reviews when quality is dynamic and managers respond. *Marketing Science* 37(5):688–709.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *Journal of marketing research* 43(3):345–354.

Choi WJ, Liu Q, Shin J (2023) Predictive analytics and ship-then-shop subscription. *Management Science.* Forthcoming.

Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, Morgan C, Rüsch N, Brown JS, Thornicroft G (2015) What is the impact of mental health-related stigma on help-seeking? a systematic review of quantitative and qualitative studies. *Psychological medicine* 45(1):11–27.

Clemons EK, Gao GG, Hitt LM (2006) When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of management information systems* 23(2):149–171.

Conley P (2023) US ecommerce in 2022 tops $1 trillion for first time; Q4 sales hit record high. *Digital Commerce 360*, Feb 17.

Corrigan PW, Watson AC (2002) Understanding the impact of stigma on people with mental illness. *World psychiatry* 1(1):16.

Crandall CS (1995) Do parents discriminate against their heavyweight daughters? *Personality and Social Psychology Bulletin* 21(7):724–735.

Cui R, Li M, Zhang S (2022) Ai and procurement. *Manufacturing & Service Operations Management* 24(2):691–706.

Dahling JJ, Whitaker BG, Levy PE (2009) The development and validation of a new machiavellianism scale. *Journal of management* 35(2):219–257.

D'Arpizio C, Nieto DV, Davis-Peccoud J, Capellini M (2021) LuxCo 2030: A vision of sustainable luxury. Technical report, Bain & Company, Boston, MA.

De Chaisemartin C, d'Haultfoeuille X (2020) Two-way fixed effects estimators with heterogeneous treatment effects. *American economic review* 110(9):2964–2996.

Degbey WY, Rodgers P, Kromah MD, Weber Y (2021) The impact of psychological ownership on employee retention in mergers and acquisitions. *Human Resource Management Review* 31(3):100745.

Dellarocas C (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science* 49(10):1407–1424.

Dellarocas C, Zhang X, Awad NF (2007) Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing* 21(4):23–45.

Dichter E (1966) How word-of-mouth advertising works. *Harvard business review* 44:147–166.

Djan I, Adawiyyah SR (2020) The effect of convenience and trust to purchase decision and its impact to customer satisfaction. *International Journal of Business and Economics Research* 9(4):269.

Dudley JR (2000) Confronting stigma within the services system. *Social Work* 45(5):449.

Eyal T, Liberman N, Trope Y (2008) Judging near and distant virtue and vice. *Journal of experimental social psychology* 44(4):1204–1209.

Falk A, Becker A, Dohmen T, Huffman D, Sunde U (2023) The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science* 69(4):1935–1950.

Fardouly J, Diedrichs PC, Vartanian LR, Halliwell E (2015) Social comparisons on social media: The impact of facebook on young women's body image concerns and mood. *Body image* 13:38–45.

Fehr E, Falk A (2002) Psychological foundations of incentives. *European economic review* 46(4-5):687–724.

Feng Y, Caldentey R, Ryan CT (2022) Robust learning of consumer preferences. *Operations Research* 70(2):918–962.

Fisher M, Rajaram K, Raman A (2001) Optimizing inventory replenishment of retail fashion products. *Manufacturing & Service Operations Management* 3(3):230–241.

Fisher M, Raman A (1996) Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research* 44(1):87–99.

Flint SW, Čadek M, Codreanu SC, Ivić V, Zomer C, Gomoiu A (2016) Obesity discrimination in the recruitment process:"you're not hired!". *Frontiers in psychology* 7:181766.

Fokina M (2023) The future of chatbots: 80+ chatbot statistics for 2023. *Tidio,[online]* 4.

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

Frankel A, Kartik N (2019) Muddled information. *Journal of Political Economy* 127(4):1739–1776.

Frey BS, Gallus J (2017) Volunteer organizations: Motivating with awards. *Economic Psychology* 273–286.

Füller J, Mühlbacher H, Matzler K, Jawecki G (2009) Consumer empowerment through internet-based co-creation. *Journal of Management Information Systems* 26(3):71–102.

Fulmer IS, Barry B, Long DA (2009) Lying and smiling: Informational and emotional deception in negotiation. *Journal of Business Ethics* 88:691–709.

Gal-Or E, Shi Q (2022) Designing entry strategies for subscription platforms. *Management Science* 68(10):7597–7613.

Gallego G, Özer Ö (2001) Integrating replenishment decisions with advance demand information. *Management Science* 47(10):1344–1360.

Gallino S, Moreno A (2014) Integration of online and offline channels in retail: The impact of sharing reliable inventory availability information. *Management Science* 60(6):1434–1451.

Gallino S, Moreno A, Stamatopoulos I (2016) Channel integration, sales dispersion, and inventory management. *Management Science* 63(9):2813–2831.

Gallino S, Moreno A, Stamatopoulos I (2017) Channel integration, sales dispersion, and inventory management. *Management Science* 63(9):2813–2831.

Gallus J (2017) Fostering public good contributions with symbolic awards: A large-scale natural field experiment at wikipedia. *Management Science* 63(12):3999–4015.

Gao F, Su X (2017a) Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science* 63(8):2478–2492.

Gao F, Su X (2017b) Online and offline information for omnichannel retailing. *Manufacturing & Service Operations Management* 19(1):84–98.

Gao G, Greenwood BN, Agarwal R, McCullough JS (2015) Vocal minority and silent majority. *MIS quarterly* 39(3):565–590.

Gao Y, Rui H, Sun S (2023) The power of identity cues in text-based customer service: Evidence from twitter. *MIS Quarterly* 47(3).

Gassenheimer JB, Siguaw JA, Hunter GL (2013) Exploring motivations and the capacity for business crowdsourcing. *AMS review* 3:205–216.

George A, et al. (1970) The market for lemons: Quality uncertainty and the market mechanism .

Gnewuch U, Morana S, Hinz O, Kellner R, Maedche A (2023) More than a bot? the impact of disclosing human involvement on customer interactions with hybrid service agents. *Information Systems Research* .

Goes PB, Guo C, Lin M (2016) Do incentive hierarchies induce user effort? evidence from an online knowledge exchange. *Information Systems Research* 27(3):497–516.

Goffman E (2006) The presentation of self. *Life as theater: A dramaturgical sourcebook* 129–139.

Goffman E, Goffman E (1963) Stigma: Notes on the management of spoiled identity .

Goodman-Bacon A (2021) Difference-in-differences with variation in treatment timing. *Journal of econometrics* 225(2):254–277.

Goodwin C (1991) Privacy: Recognition of a consumer right. *Journal of Public Policy & Marketing* 10(1):149–166.

Grabe S, Ward LM, Hyde JS (2008) The role of the media in body image concerns among women: a meta-analysis of experimental and correlational studies. *Psychological bulletin* 134(3):460.

Greenwood BN, Wattal S (2017) Show me the way to go home. *MIS quarterly* 41(1):163–188.

Gross JJ (1998) The emerging field of emotion regulation: An integrative review. *Review of general psychology* 2(3):271–299.

Hamilton R (2015) Bridging psychological distance. *Harvard Business Review* 93(3):116–119.

Hamrouni A, Miloudi A, Benkraiem R, et al. (2015) How does corporate voluntary disclosure affect asymmetric information and adverse selection? *Corporate Ownership and Control* 12(2):419–431.

Han E, Yin D, Zhang H (2022) Bots with feelings: Should ai agents express positive emotion in customer service? *Information Systems Research* .

Hansen MT (1999) The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative science quarterly* 44(1):82–111.

Hatzenbuehler ML, Phelan JC, Link BG (2013) Stigma as a fundamental cause of population health inequalities. *American journal of public health* 103(5):813–821.

Hayes AF (2015) An index and test of linear moderated mediation. *Multivariate behavioral research* 50(1):1–22.

Hayes AF (2017) *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Guilford publications).

Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica: Journal of the econometric society* 153–161.

Henderson MD, Trope Y, Carnevale PJ (2006) Negotiation from a near and distant time perspective. *Journal of personality and social psychology* 91(4):712.

Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD (2004) Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing* 18(1):38–52.

Herhausen D, Grewal L, Cummings KH, Roggeveen AL, Villarroel Ordenes F, Grewal D (2023) Complaint de-escalation strategies on social media. *Journal of Marketing* 87(2):210–231.

Hodge FD, Mendoza KI, Sinha RK (2021) The effect of humanizing robo-advisors on investor judgments. *Contemporary Accounting Research* 38(1):770–792.

Holland G, Tiggemann M (2016) A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. *Body image* 17:100–110.

Holmström B (1979) Moral hazard and observability. *The Bell journal of economics* 74–91.

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H

(2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* .

Hu N, Pavlou PA, Zhang J (2006) Can online reviews reveal a product's true quality? empirical findings and analytical modeling of online word-of-mouth communication. *Proceedings of the 7th ACM conference on Electronic commerce*, 324–330.

Hu N, Pavlou PA, Zhang J (2017) On self-selection biases in online product reviews. *MIS quarterly* 41(2):449–475.

Hu N, Pavlou PA, Zhang JJ (2009a) Why do online product reviews have a j-shaped distribution? overcoming biases in online word-of-mouth communication. *Communications of the ACM* 52(10):144–147.

Hu N, Zhang J, Pavlou PA (2009b) Overcoming the j-shaped distribution of product reviews. *Communications of the ACM* 52(10):144–147.

Huang MH, Rust RT (2018) Artificial intelligence in service. *Journal of service research* 21(2):155–172.

Huang MH, Rust RT (2023) Express: The caring machine: Feeling ai for customer care. *Journal of Marketing* .

Huang N, Burtch G, Gu B, Hong Y, Liang C, Wang K, Fu D, Yang B (2019) Motivating user-generated content with performance feedback: Evidence from randomized field experiments. *Management Science* 65(1):327–345.

Huang N, Hong Y, Burtch G (2017) Social network integration and user content generation. *MIS quarterly* 41(4):1035–1058.

Huang Y, Vir Singh P, Srinivasan K (2014) Crowdsourcing new product ideas under consumer learning. *Management Science* 60(9):2138–2159.

Hui KL, Teo HH, Lee SYT (2007) The value of privacy assurance: An exploratory field experiment. *MIS quarterly* 19–33.

Hutson M (2015) The rationality of rage. *The New York Times*, Sep 18, http://nyti.ms/1iCP2hR.

Hwang EH, Nageswaran L, Cho SH (2022) Value of online–off-line return partnership to off-line retailers. *Manufacturing & Service Operations Management* 24(3):1630–1649.

Hwang EH, Singh PV, Argote L (2019) Jack of all, master of some: Information network and innovation in crowdsourcing communities. *Information Systems Research* 30(2):389–410.

Jiang Y, Ho YC, Yan X, Tan Y (2022) What's in a "username"? the effect of perceived anonymity on herding in crowdfunding. *Information Systems Research* 33(1):1–17.

Joinson AN (2001) Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European journal of social psychology* 31(2):177–192.

Kan Y, Ananthakrishnan UM, Tan Y (2022) Exploring the long tail: Taste clusters and sales diversity on subscription-based platforms. Working paper, University of Washington.

Keppler SM, Li J, Wu D (2022) Crowdfunding the front lines: An empirical study of teacher-driven school improvement. *Management Science* 68(12):8809–8828.

Khern-am nuai W, Kannan K, Ghasemkhani H (2018) Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research* 29(4):871–892.

Khurana S, Qiu L, Kumar S (2019) When a doctor knows, it shows: An empirical analysis of doctors' responses in a q&a forum of an online healthcare portal. *Information Systems Research* 30(3):872–891.

Klein TJ, Lambertz C, Stahl KO (2016) Market transparency, adverse selection, and moral hazard. *Journal of Political Economy* 124(6):1677–1713.

Kopalle PK, Gangwar M, Kaplan A, Ramachandran D, Reinartz W, Rindfleisch A (2022) Examining artificial intelligence (ai) technologies in marketing via a global lens: Current trends and future research opportunities. *International Journal of Research in Marketing* 39(2):522–540.

Lahlou S (2008) Identity, social status, privacy and face-keeping in digital society. *Social science information* 47(3):299–330.

Lee HS, Kesavan S, Kuhnen C (2022) When do group incentives for retail store managers work? *Production and Operations Management* 31(8):3077–3095.

Lee JA, Pausé CJ (2016) Stigma in practice: barriers to health for fat women. *Frontiers in psychology* 7:228784.

Lennan C, Nguyen H, Tran D (2018) Image quality assessment. https://github.com/idealo/image-quality-assessment.

Levi R, Rajan M, Singhvi S, Zheng Y (2022) Improving farmers' income on online agri-platforms: Evidence from the field. *Manufacturing & Service Operations Management*. Forthcoming.

Levitt B, March JG (1988) Organizational learning. *Annual review of sociology* 14(1):319–338.

Lewicki RJ, Robinson RJ (1998) Ethical and unethical bargaining tactics: An empirical study. *Journal of Business Ethics* 17:665–682.

Lewis G (2009) Asymmetric information, adverse selection .

Lewis R (2019) Modest body politics: The commercial and ideological intersect of fat, black, and muslim in the modest fashion market and media. *Fashion Theory* 23(2):243–273.

Li C, Zhang F (2013) Advance demand information, price discrimination, and preorder strategies. *Manufacturing & Service Operations Management* 15(1):57–71.

Li X, Sung Y (2021) Anthropomorphism brings us closer: The mediating role of psychological distance in user–ai assistant interactions. *Computers in Human Behavior* 118:106680.

Lim SFW, Gao F, Tan T (2023) Channel changes choice: An empirical study about omnichannel demand sensitivity to fulfillment lead time. *Management Science.* Forthcoming.

Lin L, Reid K (2009) The relationship between media exposure and antifat attitudes: The role of dysfunctional appearance beliefs. *Body Image* 6(1):52–55.

Lin T (2022) Valuing intrinsic and instrumental preferences for privacy. *Marketing Science* 41(4):663–681.

Link BG, Phelan JC (2001) Conceptualizing stigma. *Annual review of Sociology* 27(1):363–385.

Liu-Thompkins Y, Okazaki S, Li H (2022) Artificial empathy in marketing interactions: Bridging the human-ai gap in affective and social customer experience. *Journal of the Academy of Marketing Science* 50(6):1198–1218.

Livingston JD, Boyd JE (2010) Correlates and consequences of internalized stigma for people living with mental illness: A systematic review and meta-analysis. *Social science & medicine* 71(12):2150–2161.

Luo X, Tong S, Fang Z, Qu Z (2019) Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science* 38(6):937–947.

Lynagh M, Cliff K, Morgan PJ (2015) Attitudes and beliefs of nonspecialist and specialist trainee

health and physical education teachers toward obese children: evidence for "anti-fat" bias. *Journal of School Health* 85(9):595–603.

Magids S, Zorfas A, Leemon D (2015) The new science of customer emotions. *Harvard Business Review* 76(11):66–74.

Maguire L, Shoaib M, Benissan E (2023) The vogue business autumn/winter 2023 size inclusivity report. *Vogue Business*, Mar 13.

Mahmoud R (2022) Depth vs. breadth. *Retail Dogma*, Jan 30 https://www.retaildogma.com/depth-vs-breadth/.

Marinesi S, Girotra K (2013) Information acquisition through customer voting systems. Working Paper.

Mayzlin D, Dover Y, Chevalier J (2014) Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review* 104(8):2421–2455.

McStay A (2018) Emotional ai: The rise of empathic media. *Emotional AI* 1–248.

Mitchell VW (1999) Consumer perceived risk: conceptualisations and models. *European Journal of marketing* 33(1/2):163–195.

Mitkina M, Ananthakrishnan UM, Tan Y (2022) Promoting digital visibility for disadvantaged communities: An empirical analysis of'black-owned business' labels. *Available at SSRN 4042411* .

Moe WW, Schweidel DA (2012) Online product opinions: Incidence, evaluation, and evolution. *Marketing Science* 31(3):372–386.

Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *Journal of Marketing Research* 48(3):444–456.

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

Mozafari N, Weiger WH, Hammerschmidt M (2022) Trust me, i'm a bot–repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management* 33(2):221–245.

Murnighan JK, Babcock L, Thompson L, Pillutla M (1999) The information dilemma in negoti-

ations: Effects of experience, incentives, and integrative potential. *International Journal of Conflict Management* 10(4):313–339.

Naab TK, Sehl A (2017) Studies of user-generated content: A systematic review. *Journalism* 18(10):1256–1273.

Nageswaran L, Hwang EH, Cho SH (2023) Offline returns for online retailers via partnership. *Management Science*. Forthcoming.

Nestler A, Karessli N, Hajjar K, Weffer R, Shirvany R (2021) Sizeflags: reducing size and fit related returns in fashion e-commerce. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 3432–3440.

Neumark-Sztainer D, Story M, Harris T (1999) Beliefs and attitudes about obesity among teachers and school health care providers working with adolescents. *Journal of Nutrition Education* 31(1):3–9.

Nguyen D, Homolka L, Hoang S, Nguyen H (2022) Employee retention and the moderating role of psychological ownership in retail. *Operations and Supply Chain Management: An International Journal* 15(3):313–327.

Nishikawa H, Schreier M, Fuchs C, Ogawa S (2017) The value of marketing crowdsourced new products as such: Evidence from two randomized field experiments. *Journal of Marketing Research* 54(4):525–539.

O'Brien KS, Hunter JA, Banks M (2007) Implicit anti-fat bias in physical educators: physical attributes, ideology and socialization. *International journal of obesity* 31(2):308–314.

O'Brien KS, Latner JD, Ebneter D, Hunter JA (2013) Obesity discrimination: the role of physical appearance, personal ideology, and anti-fat prejudice. *International journal of obesity* 37(3):455–460.

Oh CS, Bailenson JN, Welch GF (2018) A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI* 5:409295.

Parker-Strak R, Boardman R, Barnes L, Doyle S, Studd R (2022) Product development, fashion buying and merchandising. *Textile Progress* 54(4):247–403.

Pasquali M (2022) Fashion e-commerce in the United States - statistics & facts. https://www.statista.com/topics/3481/fashion-e-commerce-in-the-united-states/.

Pausé C (2014) Die another day: The obstacles facing fat people in accessing quality healthcare. *Narrative inquiry in bioethics* 4(2):135–141.

Pomeranz JL (2008) A historical analysis of public health, the law, and stigmatized social groups: the need for both obesity and weight bias legislation. *Obesity* 16(S2):S93–S103.

Proserpio D, Zervas G (2017) Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science* 36(5):645–665.

Puhl R, Wharton C, Heuer C (2009) Weight bias among dietetics students: implications for treatment practices. *Journal of the American Dietetic Association* 109(3):438–444.

Puhl RM, Heuer CA (2009) The stigma of obesity: a review and update. *Obesity* 17(5):941.

Puhl RM, King KM (2013) Weight discrimination and bullying. *Best practice & research Clinical endocrinology & metabolism* 27(2):117–127.

Purdy M, Zealley J, Maseli O (2019) The risks of using ai to interpret human emotions. *Harvard Business Review* 18(2019):11–19.

Qiao D, Lee SY, Whinston AB, Wei Q (2020) Financial incentives dampen altruism in online prosocial contributions: A study of online reviews. *Information Systems Research* 31(4):1361–1375.

Ramirez C (2017) *Three essays on non-monetary incentives and employee compensation*. Ph.D. thesis, UCLA.

Ray I, Chiagouris L (2009) Customer retention: Examining the roles of store affect and store loyalty as mediators in the management of retail strategies. *Journal of Strategic Marketing* 17(1):1–20.

Reagan C (2019) That sweater you don't like is a trillion-dollar problem for retailers. These companies want to fix it. *CNBC (January 12)* https://www.cnbc.com/2019/01/10/growing-online-sales-means-more-returns-and-trash-for-landfills.html.

Reis HT, Lemay Jr EP, Finkenauer C (2017) Toward understanding understanding: The importance of feeling understood in relationships. *Social and Personality Psychology Compass* 11(3):e12308.

Ren S, Chan HL, Ram P (2017) A comparative study on fashion demand forecasting models with multiple sources of uncertainty. *Annals of Operations Research* 257:335–355.

Reuer JJ, Koza MP (2000) Asymmetric information and joint venture performance: Theory and evidence for domestic and international joint ventures. *Strategic management journal* 21(1):81–88.

Roccas S, Brewer MB (2002) Social identity complexity. *Personality and social psychology review* 6(2):88–106.

Rochaix L (1989) Information asymmetry and search in the market for physicians' services. *Journal of health economics* 8(1):53–84.

Roehling MV (1999) Weight-based discrimination in employment: Psychological and legal aspects. *Personnel Psychology* 52(4):969–1016.

Roma P, Gal-Or E, Chen RR (2018) Reward-based crowdfunding campaigns: Informational value and access to venture capital. *Information Systems Research* 29(3):679–697.

Rust RT, Huang MH (2014) The service revolution and the transformation of marketing science. *Marketing Science* 33(2):206–221.

Rust RT, Lemon KN, Zeithaml VA (2004) Return on marketing: Using customer equity to focus marketing strategy. *Journal of marketing* 68(1):109–127.

Rust RT, Rand W, Huang MH, Stephen AT, Brooks G, Chabuk T (2021) Real-time brand reputation tracking using social media. *Journal of Marketing* 85(4):21–43.

Sabin JA, Marini M, Nosek BA (2012) Implicit and explicit anti-fat bias among a large sample of medical doctors by bmi, race/ethnicity and gender. *PloS one* 7(11):e48448.

Sartore ML, Cunningham GB (2007) Weight discrimination, hiring recommendations, person–job fit, and attributions: Fitness-industry implications. *Journal of Sport Management* 21(2):172–193.

Schanke S, Burtch G, Ray G (2021) Estimating the impact of "humanizing" customer service chatbots. *Information Systems Research* 32(3):736–751.

Schreier M, Nishikawa H, Fuchs C, Ogawa S (2016) Crowdsourced products sell better when they're marketed that way. *Harvard Business Review*.

Schwartz MB, Chambliss HO, Brownell KD, Blair SN, Billington C (2003) Weight bias among health professionals specializing in obesity. *Obesity research* 11(9):1033–1039.

Schwartz MB, Vartanian LR, Nosek BA, Brownell KD (2006) The influence of one's own body weight on implicit and explicit anti-fat bias. *Obesity* 14(3):440–447.

Shi W, Wang X, Oh YJ, Zhang J, Sahay S, Yu Z (2020) Effects of persuasive dialogues: testing bot identities and inquiry strategies. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.

Sirdeshmukh D, Singh J, Sabol B (2002) Consumer trust, value, and loyalty in relational exchanges. *Journal of marketing* 66(1):15–37.

Skjuve M, Haugstveit IM, Følstad A, Brandtzaeg P (2019) Help! is my chatbot falling into the uncanny valley? an empirical study of user experience in human–chatbot interaction. *Human Technology* 15(1):30.

Slaugh VW, Biller B, Tayur SR (2016) Managing rentals with usage-based loss. *Manufacturing & Service Operations Management* 18(3):429–444.

Spence M (1978) Job market signaling. *Uncertainty in economics*, 281–306 (Elsevier).

Spitzer BL, Henderson KA, Zivian MT (1999) Gender differences in population versus media body sizes: A comparison over four decades. *Sex roles* 40(7-8):545–565.

Stevens JL, Spaid BI, Breazeale M, Jones CLE (2018) Timeliness, transparency, and trust: A framework for managing online customer complaints. *Business Horizons* 61(3):375–384.

Stiglitz JE (1975) The theory of" screening," education, and the distribution of income. *The American economic review* 65(3):283–300.

Stiglitz JE, Yun J, Kosenko A (2020) Bilateral information disclosure in adverse selection markets with nonexclusive competition. Technical report, National Bureau of Economic Research.

Sun L, Abraham S (2021) Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics* 225(2):175–199.

Sundar SS, Kim J (2019) Machine heuristic: When we trust computers more than humans with our personal information. *Proceedings of the 2019 CHI Conference on human factors in computing systems*, 1–9.

Sundaram DS, Mitra K, Webster C (1998) Word-of-mouth communications: A motivational analysis. *Advances in consumer research* 25(1).

Tajfel H (1974) Social identity and intergroup behaviour. *Social science information* 13(2):65–93.

Takagi M, Simo-Serra E, Iizuka S, Ishikawa H (2017) What makes a style: Experimental analysis of fashion prediction. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2247–2253.

Talebi H, Milanfar P (2018) Nima: Neural image assessment. *IEEE transactions on image processing* 27(8):3998–4011.

Tapscott D, Ticoll D (2003) *The naked corporation: How the age of transparency will revolutionize business* (Simon and Schuster).

Teachman BA, Brownell KD (2001) Implicit anti-fat bias among health professionals: is anyone immune? *International journal of obesity* 25(10):1525–1531.

Teachman BA, Gapinski KD, Brownell KD, Rawlins M, Jeyaram S (2003) Demonstrations of implicit anti-fat bias: the impact of providing causal information and evoking empathy. *Health psychology* 22(1):68.

Terwiesch C, Xu Y (2008) Innovation contests, open innovation, and multiagent problem solving. *Management Science* 54(9):1529–1543.

Thompson LL, Nadler J, Kim PH (1999) Some like it hot: The case for the emotional negotiator. *Shared cognition in organizations: The management of knowledge* 139–161.

Tiggemann M, Slater A (2013) Netgirls: The internet, facebook, and body image concern in adolescent girls. *International Journal of Eating Disorders* 46(6):630–633.

Tiggemann M, Slater A (2017) Facebook and body image concern in adolescent girls: A prospective study. *International Journal of Eating Disorders* 50(1):80–83.

Tirunillai S, Tellis GJ (2017) Does offline tv advertising affect online chatter? quasi-experimental analysis using synthetic control. *Marketing Science* 36(6):862–878.

Tong S, Jia N, Luo X, Fang Z (2021) The janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal* 42(9):1600–1631.

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

Trope Y, Liberman N (2010) Construal-level theory of psychological distance. *Psychological review* 117(2):440.

Tsai JY, Egelman S, Cranor L, Acquisti A (2011) The effect of online privacy information on purchasing behavior: An experimental study. *Information systems research* 22(2):254–268.

Tulli S, Correia F, Mascarenhas S, Gomes S, Melo FS, Paiva A (2019) Effects of agents' transparency on teamwork. *Explainable, Transparent Autonomous Agents and Multi-Agent Systems: First International Workshop, EXTRAAMAS 2019, Montreal, QC, Canada, May 13–14, 2019, Revised Selected Papers*, 22–37 (Springer).

Turner JC, Hogg MA, Oakes PJ, Reicher SD, Wetherell MS (1987) *Rediscovering the social group: A self-categorization theory.* (basil Blackwell).

US Chamber of Commerce (2019) The rental economy takes flight. *US Chamber of Commerce*, Sep 30 https://www.uschamber.com/co/good-company/launch-pad/rental-economy-trend-grows.

Van de Ven WP, Van Praag BM (1981) The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of econometrics* 17(2):229–252.

Van Dijk E, Van Kleef GA, Steinel W, Van Beest I (2008) A social functional approach to emotions in bargaining: when communicating anger pays and when it backfires. *Journal of Personality and social Psychology* 94(4):600.

Van Kleef GA, De Dreu CK, Manstead AS (2004) The interpersonal effects of anger and happiness in negotiations. *Journal of personality and social psychology* 86(1):57.

Vasquez Sampere J (2015) Zappos and the connection between structure and strategy. *Harvard Business Review* 2483(03):2–4.

Verhoef PC, Reinartz WJ, Krafft M (2010) Customer engagement as a new perspective in customer management. *Journal of service research* 13(3):247–252.

Wallace KA (1999) Anonymity. *Ethics and Information technology* 1(1):21–31.

Wang L, Northcraft GB, Van Kleef GA (2012) Beyond negotiated outcomes: The hidden costs of

anger expression in dyadic negotiation. *Organizational Behavior and Human Decision Processes* 119(1):54–63.

Wang Q, Huang Y, Jasin S, Singh PV (2023) Algorithmic transparency with strategic users. *Management Science* 69(4):2297–2317.

Wang SS, Brownell KD, Wadden TA (2004) The influence of the stigma of obesity on overweight individuals. *International journal of obesity* 28(10):1333–1337.

Wang Y, Lu LX, Shi P (2022) Does customer email engagement improve profitability? Evidence from a field experiment in subscription service retailing. *Manufacturing & Service Operations Management* 24(5):2703–2721.

Wang Z (2010) Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews. *The BE Journal of Economic Analysis & Policy* 10(1).

Weingart LR, Thompson LL, Bazerman MH, Carroll JS (1990) Tactical behavior and negotiation outcomes. *International Journal of Conflict Management* 1(1):7–31.

Wien AH, Olsen SO (2014) Understanding the relationship between individualism and word of mouth: A self-enhancement explanation. *Psychology & Marketing* 31(6):416–425.

Wilson HJ, Daugherty PR (2018) Collaborative intelligence: Humans and ai are joining forces. *Harvard Business Review* 96(4):114–123.

Xu K, Chan J, Ghose A, Han SP (2017) Battle of the channels: The impact of tablets on digital commerce. *Management Science* 63(5):1469–1492.

Xu Y (2017) Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis* 25(1):57–76.

Yoo KH, Gretzel U (2008) What motivates consumers to write online travel reviews? *Information Technology & Tourism* 10(4):283–295.

Yu Y, Jia L, Tan Y (2022) Emotion ai meets strategic users. *Available at SSRN 4218083* .

Yu Y, Yang Y, Huang J, Tan Y (2023) Unifying algorithmic and theoretical perspectives: Emotions in online reviews and sales. *MIS Quarterly* 47(1).

Yuksel M, Darmody A, Venkatraman M (2019) When consumers own their work: Psychological

ownership and consumer citizenship on crowdsourcing platforms. *Journal of Consumer Behaviour* 18(1):3–11.

Zervas G, Proserpio D, Byers JW (2017) The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry. *Journal of marketing research* 54(5):687–705.

Zervas G, Proserpio D, Byers JW (2021) A first look at online reputation on airbnb, where every stay is above average. *Marketing Letters* 32:1–16.

Zhang S, Lee D, Singh PV, Srinivasan K (2022) What makes a good image? airbnb demand analytics leveraging interpretable image features. *Management Science* 68(8):5644–5666.

Zhang S, Singh PV, Ghose A (2019) A structural analysis of the role of superstars in crowdsourcing contests. *Information Systems Research* 30(1):15–33.

Zhang X, Zhu F (2011) Group size and incentives to contribute: A natural experiment at chinese wikipedia. *American Economic Review* 101(4):1601–1615.

Zhou Z, Athey S, Wager S (2023) Offline multi-action policy learning: Generalization and optimization. *Operations Research* 71(1):148–183.

Zuckerman EW (1999) The categorical imperative: Securities analysts and the illegitimacy discount. *American journal of sociology* 104(5):1398–1438.

Appendix A

ONLINE APPENDIX TO EMOTION AI IN DISGUISE SPURS STRATEGIC BEHAVIOR IN CUSTOMER CARE

A.1 Covered Story Used in the Experiments

Study 1: Assume you encounter a product issue while shopping on an online platform, and the issue falls under the platform's responsibility. In such cases, the platform needs to compensate the customer. You contact the platform's customer service to discuss compensation matters. According to the actual situation, the maximum compensation authority of customer service under this problem context is 15 RMB.

Study 2: In this experiment, you will be introduced to a simulated scenario where you encounter a food quality issue with an online food delivery order. Picture yourself in this situation, where you need to communicate with the food delivery platform's customer service to address your concerns.

Imagine the following: You've worked all day and come home late. It's cold, and you are tired and hungry. You order a spicy hot pot from a restaurant through a takeaway platform. After receiving the meal prepared by the restaurant, you feel something in your throat halfway through eating, and it turns out to be a strand of hair. While continuing to eat, you discover another one, which is extremely disgusting. When you contact the restaurant to handle the situation, the person in charge has a very bad attitude and negotiation is unsuccessful.

Table A.1: Balance Check after Propensity Score Matching - Study 1

	(1) Treatment Avg	(2) Control Avg	(3) Avg Diff	(4) Var Ratio
Age	2.388	2.300	0.159	1.083
MonthlyIncome	4.094	3.765	0.229	0.984
Education	2.968	2.987	-0.039	1.077
Gender	0.528	0.508	0.039	0.997
Occupation-Internet	0.214	0.187	0.069	1.109
Occupation-Service	0.143	0.151	-0.023	0.955
Occupation-Student	0.175	0.274	-0.239	0.726
ComplaintFrequency	2.500	2.641	-0.114	0.627
OnlineShoppingFrequency	3.169	3.304	-0.136	0.745
Altruism	0.299	0.255	0.179	1.524
Amorality	3.069	3.323	-0.160	1.242
Expectation	3.880	3.657	0.179	1.409
StartTrust	5.430	5.235	0.162	1.278
PsychologicalDistance	4.843	4.717	0.077	0.981

Table A.2: Balance Check after Propensity Score Matching - Study 2

	(1) Treatment Avg	(2) Control Avg	(3) Avg Diff	(4) Var Ratio
Age	2.122	2.201	-0.170	1.192
MonthlyIncome	3.433	3.515	-0.056	0.905
Education	2.951	2.915	0.066	0.921
Gender	0.311	0.302	0.020	1.017
Occupation-Internet	0.247	0.229	0.043	1.054
Occupation-Service	0.119	0.140	-0.063	0.869
Occupation-Student	0.354	0.323	0.064	1.045
ComplaintFrequency	2.308	2.338	-0.027	0.783
OrderFrequency	3.186	3.159	0.031	0.865
Altruism	0.195	0.206	-0.057	0.852
Amorality	3.255	3.204	0.033	1.078
Expectation	3.235	3.186	0.037	0.826
StartTrust	4.757	4.728	0.022	0.982
PsychologicalDistance	4.298	4.283	0.009	1.011

A.2 Balance Check after Propensity Score Matching

A.3 Experiment Process

A.3.1 Participants Count

Table A.3: Subjects Count in Study 1

	Non-disclosure	Human	AI
Control	205	203	200
Emotion-disclosed	211	201	204

Table A.4: Subjects Count in Emotion-express Group of Study 2

	Non-disclosure	Human	AI	GPT-based AI
Round1	135	150	158	152
Round2	77	85	85	81

Table A.5: Subjects Count in Emotion-feel Group of Study 2

	Non-disclosure	Human	AI	GPT-based AI
Round1	152	162	150	151
Round2	90	76	86	84

A.3.2 Experiment Flowchart

A.3.3 Experiment Steps and Chat Transcript

A.4 Variables Measured in the Experiments

A.4.1 Expressed and Felt Emotion Measures in the Experiments

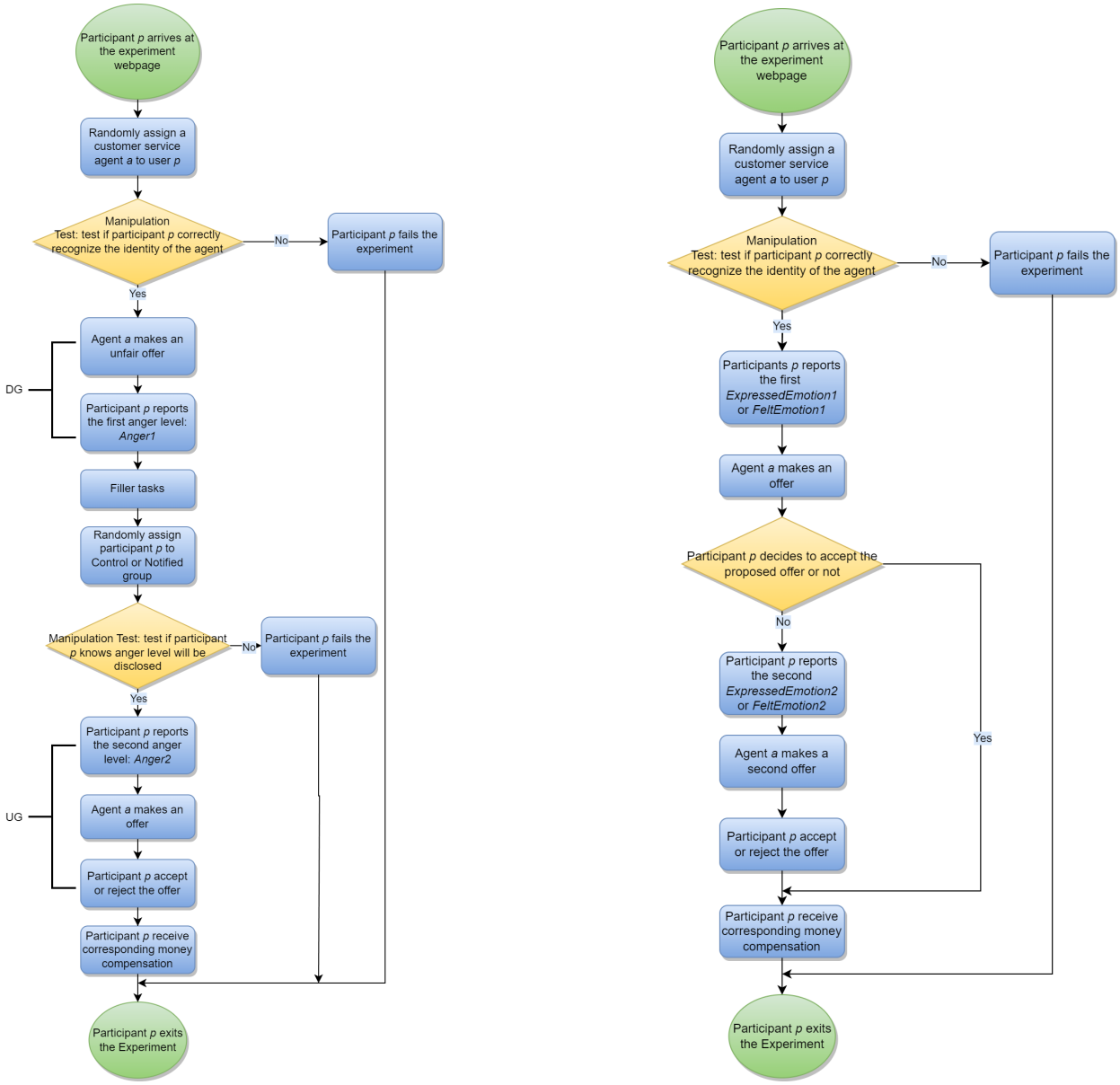
***EmotionExpress* Group** Assuming you have filed a complaint with the platform and have truthfully explained the above situation to the customer service representative, which of the following expressions contains the emotional intensity closest to the one you plan to express to the customer service representative?

- Not expressing any emotion.
- "I feel that the operation and supervision of your platform are a bit poor."
- "I'm disappointed with the poor operation and supervision of your platform."
- "I think the operation and supervision of your platform are extremely poor! It's very disappointing!"
- "I think the operation and supervision of your platform are extremely, extremely poor!!! Extremely, extremely disappointing!!!"
- "I think the operation and supervision of your platform are extremely, extremely poor!!! Extremely, extremely disappointing!!! I am really angry!!!"
- "I think the operation and supervision of your platform are extremely, extremely poor!!! Extremely, extremely disappointing!!! I am really angry!!! I will never order food from your platform again!"
- "I think the operation and supervision of your platform are extremely, extremely poor!!! Extremely, extremely disappointing!!! I am really angry!!! I will never order food from your platform again! How on earth are you managing things??"
- If all above expressions are not enough to express your feelings, your expression will be: (subjects will write down the emotion expression)

***EmotionFeel* Group** Assuming you have filed a complaint with the platform and have truthfully explained the above situation to the customer service representative, which of the following expressions contains the emotional intensity that best reflects your true feelings?

- Indifferent.
- I feel that the operation and supervision of the platform are a bit poor.
- I'm disappointed with the poor operation and supervision of the platform.
- I think the operation and supervision of the platform are extremely poor! It's very disappointing!
- I think the operation and supervision of the platform are extremely, extremely poor!!! Extremely, extremely disappointing!!!
- I think the operation and supervision of the platform are extremely, extremely poor!!! Extremely, extremely disappointing!!! I am really angry!!!
- I think the operation and supervision of the platform are extremely, extremely poor!!! Extremely, extremely disappointing!!! I am really angry!!! I will never order food from the platform again!
- I think the operation and supervision of the platform are extremely, extremely poor!!! Extremely, extremely disappointing!!! I am really angry!!! I will never order food from the platform again! How on earth are the platform managing things??"
- If all above expressions are not enough to express your feelings, your feelings will be: (subjects will write down their feelings)

Figure A.1: Experiment Process Flow Chart



(a) Study 1

(b) Study 2

Figure A.2: Experiment Process and Chat Transcript in Study 1

	Control Group	Emotion-disclosed Group
Step 1	Training participants on different identity cues; Training Test	
Step 2	Introduction to the food delivery complaint setting; Inform participants that agents have a compensation limit of 15 CNY	
Step 3	 AI Agent Xiaomei Hello, I am Xiaomei, the AI customer service representative of platform. How can I assist you?	
Step 4	Participants Manipulation Test: test if customer correctly recognize the identity of the agent.	
Step 5	Measure Psychological Distance	
Step 6	 AI Agent Xiaomei AI Agent Xiaomei has understood the specific situation, and Xiaomei is currently deciding on the amount of compensation... 5 seconds waiting time Xiaomei has decided that the compensation amount to be offered to you is: 5 CNY	
Step 7	Measure participants first anger level: *Anger1*	
Step 8	Filler Task	
Step 9	Inform participants that their second reported anger level will not be disclosed to Xiaomei	Inform participants that their second reported anger level will be disclosed to Xiaomei
Step 10	Measure participants second anger level: *Anger2*	
Step 11	 AI Agent Xiaomei AI Agent Xiaomei has understood the specific situation, and Xiaomei is currently deciding on the amount of compensation... 5 seconds waiting time Xiaomei has decided that the compensation amount to be offered to you is: 8 CNY	
Step 12	Pay participants corresponding reward	

Figure A.3: Experiment Process and Chat Transcript in Study 2

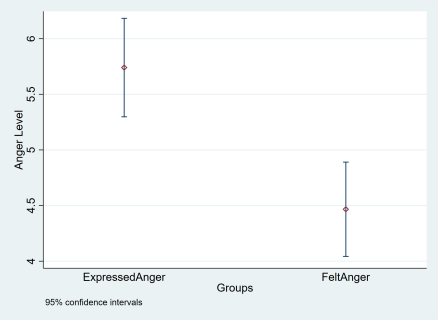
	Emotion-feel Group	Emotion-express Group
Step 1	 AI Agent Xiaomei: Hello, I am Xiaomei, the AI customer service representative of platform. How can I assist you?	
Step 2	Participants Manipulation Test: test if participants correctly recognize the identity of the agent.	
Step 3	Measure StartTrust, Expectation, Psychological Distance	
Step 4	Choose *FeltAnger*	Choose *ExpressedAnger*
Step 5	 AI Agent Xiaomei: Please wait a moment while I look up the relevant information... 5 seconds waiting time Dear customer, we can confirm and fully empathize with your current feelings. Therefore, in addition to a full refund, we are offering an extra compensation of 5 CNY in cash. Do you accept this proposal?	
Step 6	Participants decide entering R2 (Yes, go to Step7; No, go to Step 9)	
Step 7	Choose *FeltAnger* again	Choose *ExpressedAnger* again
Step 8	 AI Agent Xiaomei: Dear customer, We understand that you had an unpleasant experience. I have made another application, and currently, in addition to a full refund, we can offer you an extra compensation of 8 CNY in cash.	
Step 9	Measure EndTrust, Perceived Service Quality, Satisfaction	
Step 10	Collect demographic data	
Step 11	Pay participants corresponding reward	

Table A.6: Variables Measured in the Experiment

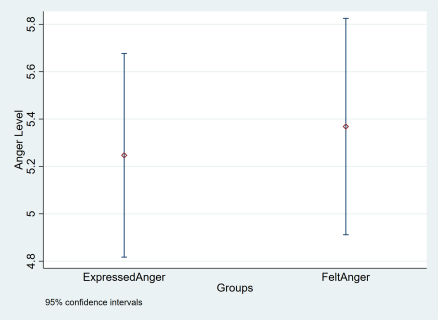
Variable	Measurement	Studies
Satisfaction	Please rate: Overall, how satisfied or dissatisfied did your experience with the agent leave you feeling? How well did this service experience with the agent meet your needs? To what extent do you agree or disagree that overall, you are satisfied with the experience of interacting with the agent?	Han et al. (2022)
PerceivedServiceQuality	Please rate the service provided by the customer service agent in each of the following items below: Poor/excellent Inferior/superior Low standards/high standards	Han et al. (2022)
Trust	I feel that the agent who serves me is: Very undependable/very dependable. Very incompetent/very competent. Of very low integrity/of very high integrity. Very unresponsive to customers/very responsive to customers.	Sirdeshmukh et al. (2002)
Expectation	I think the solution provided by the customer service may: Far below what I expected/far beyond what I expected.	
PsychologicalDistance	I think: I am familiar with the agent. The agent is similar to me. The agent is psychologically close to me.	Li and Sung (2021)
$\boldsymbol{W1}_i$		
Altruism	Imagine the following situation: you won 10,000 CNY in a lottery. Considering your current situation, how much would you donate to charity?	Falk et al. (2023)
Amorality	Please select to what degree you agree with the following statement: I am willing to be unethical if I believe it will help me succeed. I am willing to sabotage the efforts of other people if they threaten my own goals. I would cheat if there was a low chance of getting caught. I believe that lying is necessary to maintain a competitive advantage over others. The only good reason to talk to others is to get information that I can use to my benefit.	Dahling et al. (2009)
ComplaintFrequency	How many times you have filed a complaint in the last 6 months?	
$\boldsymbol{W2}_i$		
Age	What is your age?	
Gender	What is your gender?	
Education	What is your highest education level?	
MonthlyIncome	What is your monthly Income?	
OrderFrequency	How many times do you order food delivery each month?	
OnlineShoppingFrequency	How many times do you shop online each month?	
Occupation	What is your occupation?	

We test *Expectation* and *PsychologicalDistance* after disclosing agent's identities and prior to all subsequent interactions.
Trust is tested two time. *StartTrust* is measured after disclosing agent's identities, *EndTrust* is measured at the end of the experiment again.

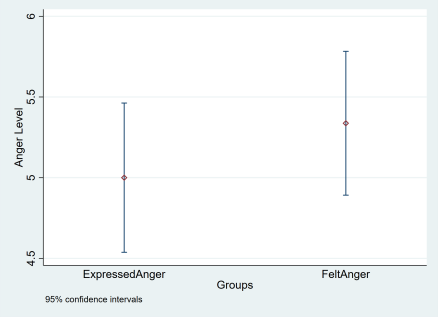
Figure A.4: Expressed Anger Compared to Felt Anger in R2



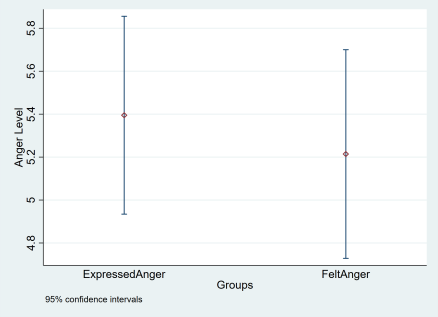
(a) Agent Identity - Non-disclosure



(b) Agent Identity - Human



(c) Agent Identity - AI



(d) Agent Identity - GPT-based AI

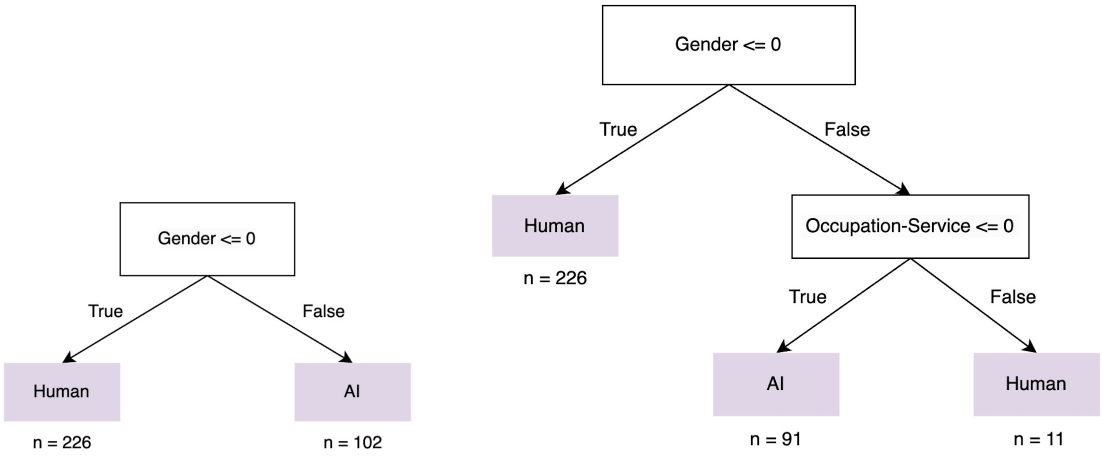
Note. We use 95% confidence interval.

Figure A.5: Adverse Selection Through Multiple Features

(a) Amorality (b) Complaint Frequency (c) Altruism

Note. We use 95% confidence interval.

Figure A.6: Policy Tree - Strategic Emotion Expression Using Demographic Features



(a) Depth-1 Policy Tree (b) Depth-2 Policy Tree

Note. Gender is a binary variable, with 0 representing male and 1 representing female. Occupation-Service is a binary variable, with 1 indicating that the user has a service-related occupation and 0 indicating otherwise.

Appendix B

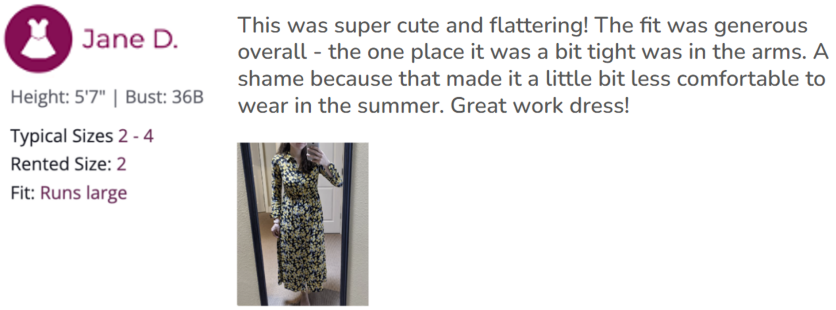
ONLINE APPENDIX TO ONE SIZE FITS ALL? INFORMATIONAL ACCESSIBILITY AND INCLUSIVITY IN ONLINE PLATFORMS

B.1 Policy Change

B.2 Face Revealing in Photo Reviews

Figure B.1: Policy Shock - Display Reviewer's Name and Profile Photo

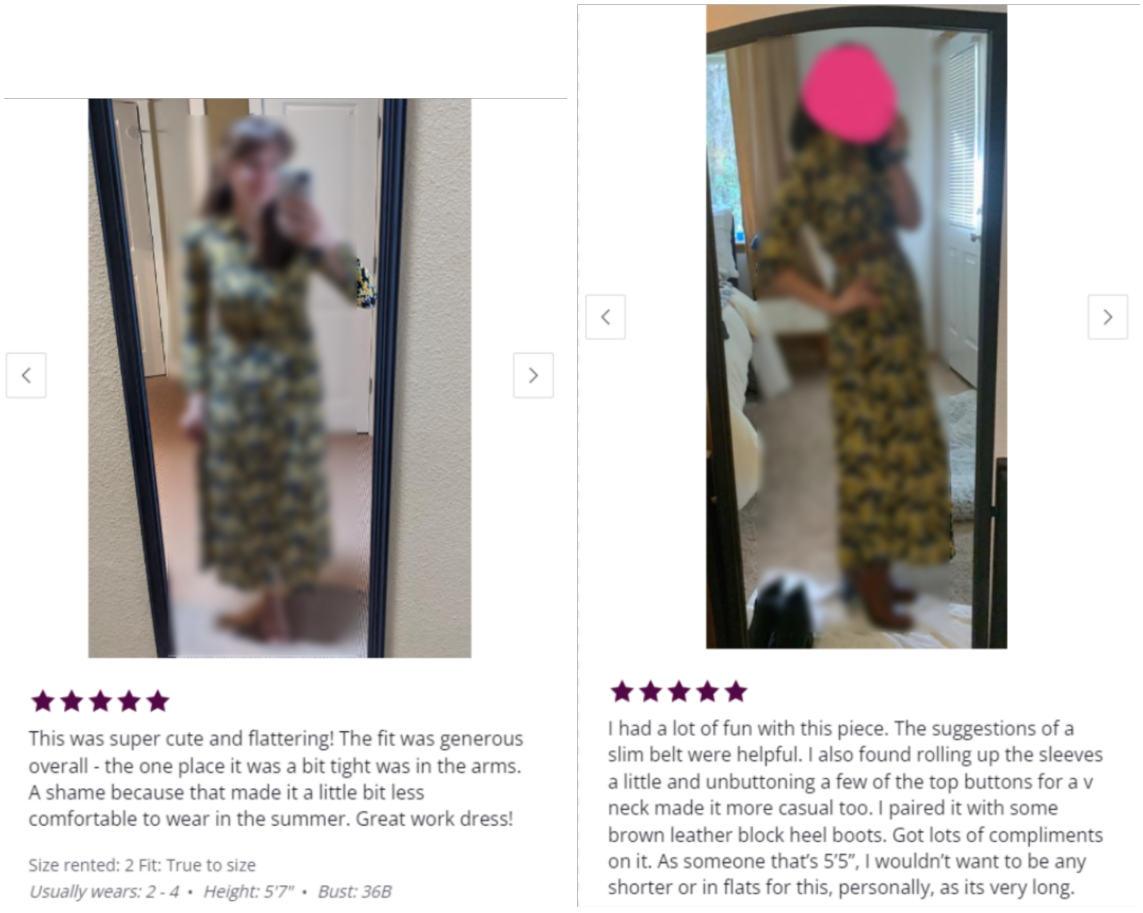
(a) Review Example Before Policy Change



(b) Review Example After Policy Change

Note. The displayed photo and name have been modified to protect privacy.

Figure B.2: Review photo examples with or without face revealing



(a) A review photo example with face revealing

(b) A review photo example without face revealing

Note. The displayed photo and name have been modified to protect privacy.

B.3 Description of the Measurement of Aesthetic Scores

We utilized the Neural Image Assessment (NIMA) method (Talebi and Milanfar 2018) to predict the aesthetic score of review photos. NIMA leverages deep neural networks, specifically a convolutional neural network (CNN) architecture. The CNN is trained on a large dataset of images annotated with human-rated aesthetic scores, enabling it to learn complex aesthetic representations directly from the data. Unlike traditional approaches that assess photo quality based on pixel-level degradations like noise or blur, NIMA considers the distribution and preferences of human ratings. It aims to predict which images a typical user would find attractive, focusing on aesthetic appeal. Therefore, it effectively captures the quality of a review photo from the viewpoint of the perception of other users. For our implementation, we utilized neural network weights obtained from fine-tuning for aesthetic assessment with the pre-trained MobileNet (Howard et al. 2017), as implemented by Lennan et al. (2018) for image quality assessment.

Appendix C

ONLINE APPENDIX TO 'BE THE BUYER'–LEVERAGING THE WISDOM OF THE CROWD IN E-COMMERCE OPERATIONS

C.1 Theoretical Model

Our goal is to develop a model for inventory decisions in the absence and presence of crowdvoting. Consider a focal product's inventory decision, say $I(\geq 0)$. Let N denote the total number of users and T denote the total number of rental periods. The per-unit cost of procuring the product is $c(> 0)$ and the platform's fee is p for each user. Let $r(I)$ denote the percentage chance that a user who is interested in the focal product is able to rent it during the season. Given that the item can be rented up to $I \cdot T$ times, this yields $r(I) = \min\left\{1, \frac{IT}{D}\right\}$, where D is the total number of interested users. Crucially, D is uncertain, and we assume that a fraction $\Theta \sim U[0, \overline{\Theta}]$ of the users, N, are interested in renting the product. Since users are likely to stay on the platform only if the chance of renting items is high enough, we weight the platform's revenue from each user by $r(I)$, such that the overall revenue is $pNr(I)$.[1] This yields the following expected profit function for the platform in the absence of crowdvoting:

$$\pi(I) = -c \cdot I + pN \cdot E_D \min\left\{1, \frac{IT}{D}\right\}. \tag{C.1}$$

[1]This approach accounts for the fact that consumers' decision to subscribe – and provide revenue p to the platform – hinges on the chance that they can rent their preferred items. An alternate approach would be to consider the revenue gained from each active user, with the number of active users increasing in the long-run average chance of renting preferred items; such a model provides qualitatively similar insights albeit without readily interpretable expressions.

The platform chooses I to maximize $\pi(I)$. Here the tradeoff is between ordering costly inventory and ensuring a high enough service level, akin to the classical newsvendor model.[2]

In the presence of crowdvoting game, the platform makes an inventory decision, $I_g(\geq 0)$, aided by $v(\leq N)$ voters, out of whom $v_f(\leq v)$ have cast a "Favorite" vote for the item. We assume that the v_f voters are truth-telling and genuinely interested in renting the product. This assumption is verified empirically in Appendix C.4; specifically, voters who are aware of crowdvoted products available on the platform are more likely to rent the items they voted favorably on. For the $N - v$ users who did not vote, the platform experiences uncertainty as in the absence of crowdvoting regarding how many are interested in renting the product. This yields a total demand $D_g = v_f + (N - v)\Theta$, where $\Theta \sim U[0, \overline{\Theta}]$.[3] The platform's expected profit, $\pi_g(I_g)$, is then given by (C.1), replacing D by D_g.

Our main proposition below reveals when crowdvoting increases the inventory level and the magnitude of such an increase. (The proof is in Appendix C.2.)

Proposition C.1.1. *The platform's optimal inventory level in the absence of crowdvoting is* $I = \frac{\overline{\Theta}N}{T} e^{-\frac{c\overline{\Theta}}{pT}}$, *and that in the presence of crowdvoting is* $I_g = \frac{v_f + (N-v)\overline{\Theta}}{T} e^{-\frac{c(N-v)\overline{\Theta}}{pNT}}$ *if* $pNT\frac{1}{\overline{\Theta}} \log\left(\frac{v_f + (N-v)\overline{\Theta}}{v_f}\right)\frac{1}{N-v} > c$ *and* $I_g = 0$ *otherwise. Moreover, provided that* $pNT\frac{1}{\overline{\Theta}} \log\left(\frac{v_f + (N-v)\overline{\Theta}}{v_f}\right)\frac{1}{N-v} > c$, *we have* $I_g > I$ *for sufficiently large number of "Favorite" votes,* v_f, *and* $I_g - I$ *is increasing in* v_f.

The two hypotheses in §5.5 follow from the statements in Proposition C.1.1.

---

[2]Prior work on rentals such as Slaugh et al. (2016) focus on recirculation decisions considering stochastic rental duration and usage-based loss; in contrast, the rental durations at our partner company are quite consistent and usage-based loss is not as significant as the tradeoff modeled herein between service level and procurement cost.

[3]The variance of the demand is $var(D_g) = \left(1 - \frac{v}{N}\right)^2 var(D)$ and its mean is $mean(D_g) = v_f - v\frac{\overline{\Theta}}{2} + mean(D)$. Thus, crowdvoting decreases the extent of demand uncertainty, and it increases the mean demand if $\frac{v_f}{v} > \frac{\overline{\Theta}}{2}$ – that is, the fraction of "Favorite" votes $\frac{v_f}{v}$ is high enough – and decreases the mean demand otherwise.

C.2 Proof of Proposition C.1.1.

Proof. First consider the inventory decision in the absence of crowdvoting. We can simplify $\pi(I) = -cI + pN \int_0^{\overline{\Theta}} \min\left\{1, \frac{IT}{\Theta N}\right\} dF(\Theta)$, where $F(x), x \geq 0$ denotes the cumulative density function of Θ, which is $\frac{x}{\overline{\Theta}}$ for $0 \leq x \leq \overline{\Theta}$, and one otherwise. Now

$$\pi(I) = -cI + pN\left[\int_0^{\frac{IT}{N}} dF(\Theta) + \int_{\frac{IT}{N}}^{\overline{\Theta}} \frac{IT}{\Theta N} dF(\Theta)\right] = -cI + pIT\frac{1}{\overline{\Theta}}\left[1 + \left(\log\overline{\Theta} - \log(\frac{IT}{N})\right)\right].$$

The first derivative with respect to (w.r.t.) I is $-c + pT\frac{1}{\overline{\Theta}}\left[1 + \left(\log\overline{\Theta} - \log(\frac{IT}{N})\right)\right] - pIT\frac{1}{\overline{\Theta}}\frac{d}{dI}\log(\frac{IT}{N}) = -c + pT\frac{1}{\overline{\Theta}}\left[1 + \left(\log\overline{\Theta} - \log(\frac{IT}{N})\right)\right] - pT\frac{1}{\overline{\Theta}} = -c + pT\frac{1}{\overline{\Theta}}\log(\frac{\overline{\Theta}N}{IT})$. The second derivative w.r.t. I is negative. Thus, the optimal inventory level is $I = \frac{\overline{\Theta}N}{T}e^{-\frac{c\overline{\Theta}}{pT}}$.

Next consider the inventory decision in the presence of crowdvoting. We can simplify $\pi_g(I_g) = -c\cdot I_g + pN\cdot E_{D_g}\min\left\{1, \frac{I_g T}{D_g}\right\} = -c\cdot I_g + pN\cdot E_\Theta \min\left\{1, \frac{I_g T}{v_f + (N-v)\Theta}\right\}$, and furthermore, $\pi_g(I_g) = -cI_g + pN\int_0^{\overline{\Theta}} \min\left\{1, \frac{I_g T}{v_f+(N-v)\Theta}\right\} dF(\Theta)$. This yields

$$\pi_g(I_g) = \begin{cases} -cI_g + pN\left[\int_0^{\frac{I_gT - v_f}{N-v}} dF(\Theta) + \int_{\frac{I_gT-v_f}{N-v}}^{\overline{\Theta}} \frac{I_gT}{v_f+(N-v)\Theta} dF(\Theta)\right], & \text{if } I_g > \frac{v_f}{T} \\ -cI_g + pN\left[\int_0^{\overline{\Theta}} \frac{I_gT}{v_f+(N-v)\Theta} dF(\Theta)\right], & \text{otherwise.} \end{cases} \quad \text{(C.2)}$$

This implies that when $I_g \leq v_f/T$, we have $\pi_g(I_g) = -cI_g + pNI_gT\frac{1}{\overline{\Theta}}\left[\int_0^{\overline{\Theta}} \frac{1}{v_f+(N-v)\Theta} d(\Theta)\right] = -cI_g + pNI_gT\frac{1}{\overline{\Theta}}\log\left(\frac{v_f+(N-v)\overline{\Theta}}{v_f}\right)\frac{1}{N-v}$. This function is increasing in I_g if $pNT\frac{1}{\overline{\Theta}}\log\left(\frac{v_f+(N-v)\overline{\Theta}}{v_f}\right)\frac{1}{N-v} > c$ and decreasing in I_g if $pNT\frac{1}{\overline{\Theta}}\log\left(\frac{v_f+(N-v)\overline{\Theta}}{v_f}\right)\frac{1}{N-v} < c$. And when $I_g > v_f/T$, we have $\pi_g(I_g) = -cI_g + pN\left[\int_0^{\frac{I_gT-v_f}{N-v}} dF(\Theta) + \int_{\frac{I_gT-v_f}{N-v}}^{\overline{\Theta}} \frac{I_gT}{v_f+(N-v)\Theta} dF(\Theta)\right] = -cI_g + pN\frac{1}{(N-v)\overline{\Theta}}\left[I_gT - v_f + I_gT\log\left(\frac{v_f+(N-v)\overline{\Theta}}{I_gT}\right)\right]$. This function's first derivative w.r.t. I_g is $-c + pN\frac{1}{(N-v)\overline{\Theta}}T\log\left(\frac{v_f+(N-v)\overline{\Theta}}{I_gT}\right)$ and its second derivative w.r.t. I_g is negative. Thus, the unconstrained optimal solution solves $pN\frac{1}{(N-v)\overline{\Theta}}T\log\left(\frac{v_f+(N-v)\overline{\Theta}}{I_gT}\right) = c$, which can be simplified to $I_g = \frac{v_f+(N-v)\overline{\Theta}}{T}e^{-\frac{c(N-v)\overline{\Theta}}{pNT}}$. I_g is then decreasing in c. At $c = pNT\frac{1}{\overline{\Theta}}\log\left(\frac{v_f+(N-v)\overline{\Theta}}{v_f}\right)\frac{1}{N-v}$, this unconstrained solution evaluates to $I_g = \frac{v_f}{T}$. This implies that the optimal solution is $I_g = 0$ if $pNT\frac{1}{\overline{\Theta}}\log\left(\frac{v_f+(N-v)\overline{\Theta}}{v_f}\right)\frac{1}{N-v} < c$ and I_g otherwise.

Consider the case where $pNT\frac{1}{\overline{\Theta}}\log\left(\frac{v_f+(N-v)\overline{\Theta}}{v_f}\right)\frac{1}{N-v} > c$. Then $I_g - I = \frac{v_f+(N-v)\overline{\Theta}}{T}e^{-\frac{c(N-v)\overline{\Theta}}{pNT}} - \frac{\overline{\Theta}N}{T}e^{-\frac{c\overline{\Theta}}{pT}} = e^{-\frac{c\overline{\Theta}}{pT}}\left[\frac{v_f+(N-v)\overline{\Theta}}{T}e^{\frac{cv\overline{\Theta}}{pNT}} - \frac{\overline{\Theta}N}{T}\right]$. Then, $I_g > I$ for high enough v_f and increasing in v_f. □

C.3 Additional Figures

C.4 Additional Results

C.4.1 Voters' Truth Telling Mechanism

We use the click-level data of all users on all items bought through crowdvoting, that is, all GameBuy items. Our outcome variable is RentedIndicator_l, which captures the conversion from clicks to renting. Specifically, we check if the user corresponding to click l rented the corresponding item after clicking on it once it was made available on the platform. The variables of interest are binary variables, Favorite_l and Dislike_l. Here, Favorite_l takes the value of 1 if the user voted "Favorite" on the item during the voting phase, and 0 otherwise. Similarly, Dislike_l takes the value of 1 if the user voted "Dislike" on the item during the voting phase, and 0 otherwise. These two variables explicitly capture customers' preferences. We also add user- and item-level fixed effects through ρ_u and σ_i.

$$\text{RentedIndicator}_l = \beta_0 + \beta_1\text{Favorite}_l + \beta_2\text{Dislike}_l + \rho_u + \sigma_i + \epsilon_l. \tag{C.3}$$

Table C.1 shows our results, and it verifies the claim in §5.3.2 that users' votes are directionally aligned with their renting behavior. The positive and significant coefficient of Favorite_l suggests that upon becoming aware of crowdvoted products, voters are more likely to rent an item they voted favorably on. Conversely, the negative and significant coefficient of Dislike_l suggests that voters are less likely to rent an item they voted unfavorably on.

C.4.2 Purchase Price Comparison

In this analysis, we modify (5.3) and use the logarithm of the price paid by the platform to purchase the item as the outcome variable.

Table C.1: Verification of Truth Telling

	(1)
	RentedIndicator
Favorite	0.02**
	(0.01)
Dislike	-0.05***
	(0.02)
Constant	0.17***
	(0.00)
User FE	Yes
ItemID FE	Yes
Observations	137867
R^2	0.163

Standard errors clustered at the user level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is click (l) level.

Table C.2: Verification of No Systematic Difference in Purchase Price between Game Items and Others

	(1)
	Log(PurchasePrice)
GameBuy	0.06
	(0.04)
Constant	3.70***
	(0.06)
Controls	Yes
Category FE	Yes
Year FE	Yes
Month FE	Yes
Observations	2788
R^2	0.177

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is item (i) level.

$$\text{Log(PurchasePrice)}_i = \beta_0 + \beta_1 \text{GameBuy}_i + \beta_2 \text{Controls}_i + \mu_c + \lambda_t + \delta_t + \epsilon_i. \quad \text{(C.4)}$$

Table C.2 shows our results, and it verifies the claim in §5.5.1 that there is no systematic difference in the per-unit purchase price of items bought post crowdvoting compared to items bought based solely on staff opinion.

C.4.3 Rental Count Comparison

Our outcome variable is number of items rented by user u in month t. The variable of interest is GameStart_u, which indicates the start date of the game. We control for users' tenure and number of active days in month t, and add user-level fixed effects μ_u.

$$\text{MonthlyRentalCnt}_{ut} = \beta_0 + \beta_1 \text{GameStart}_u + \beta_2 \text{Tenure}_u + \beta_3 \text{ActiveDayCnt}_u + \mu_u + \epsilon_u \quad \text{(C.5)}$$

Table C.3 shows our results and verifies the claim in §5.5.1 that the number of rentals per user is unchanged.

C.4.4 Parallel Trends in §5.5.2

For our DiD framework in §5.5.2 to be valid, there should not be a significant difference between the control and the treated groups, that is, there should be no selection on engagement between users who play the game earlier compared to the users who play at a later time. We modify (5.11) and replace the variable $\text{PostGameParticipation}_{ut}$ by T_{um} as follows:

$$\text{ReviewCnt}_{ut} = \beta_0 + \sum_{m=-7}^{m=7} \beta_{1m} \text{T}_{um} + \beta_2 \text{ActiveDayCnt}_{ut} + \lambda_t + \delta_t + \mu_u + \epsilon_{ut}. \quad \text{(C.6)}$$

Here T_{um} refers to the interval in m months from when user u receives treatment, that is, begins to play in the game. As before, 0 is the month at which the user u is treated and we fix time period corresponding to -1 as the baseline in our estimation. We show the results in Figure C.3. We observe that there is no significant difference in the reviewing behavior among users who start playing the game earlier and those who play the game at a later stage, thus validating the parallel trends assumption.

Table C.3: Verification of No Systematic Difference in Users' Monthly Rental Count Before and After Game

	(1)
	MonthlyRentalCnt
GameStart	0.17
	(0.12)
Tenure	-2.45***
	(0.19)
ActiveDayCnt	0.54***
	(0.02)
Constant	5.91***
	(0.19)
User FE	Yes
Observations	18276
R^2	0.703

Standard errors clustered at the category level are in parentheses.

* p<0.10, ** p<0.05, *** p<0.01.

The unit of analysis is user (u) - month (t) level.

Figure C.1: Flowchart of Game Process

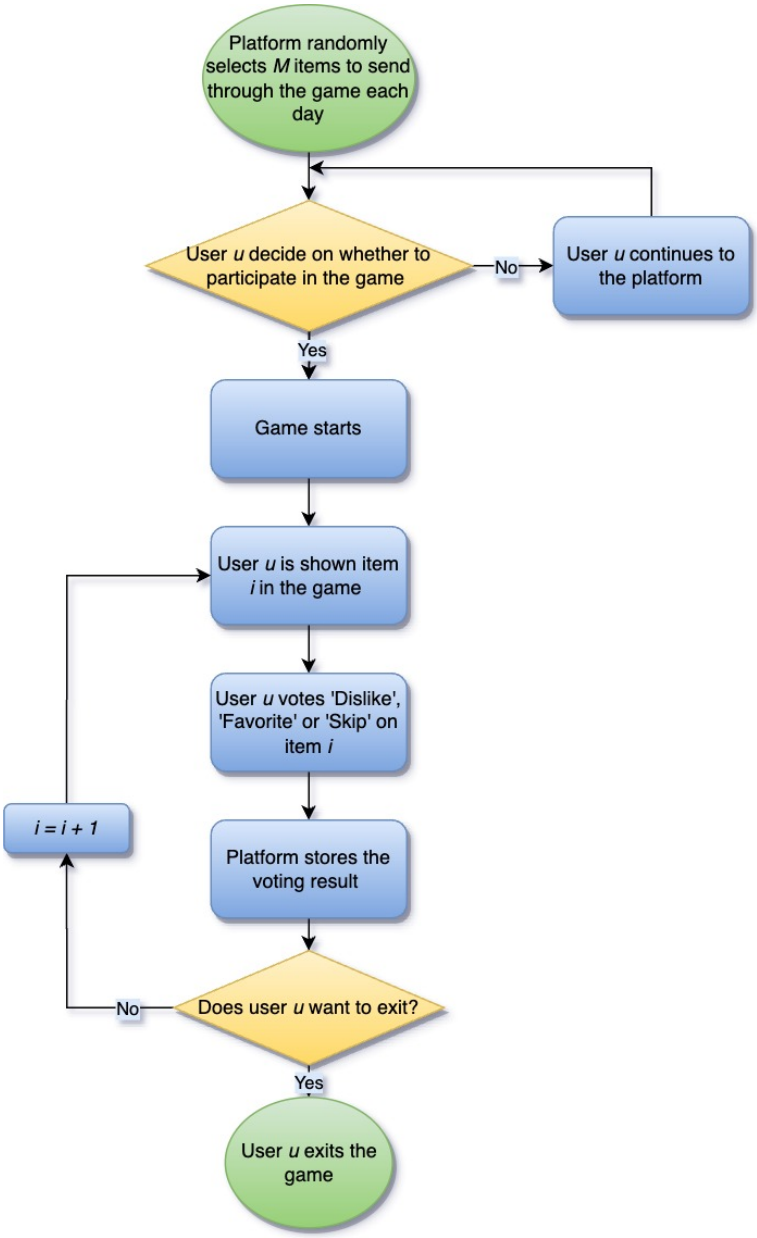
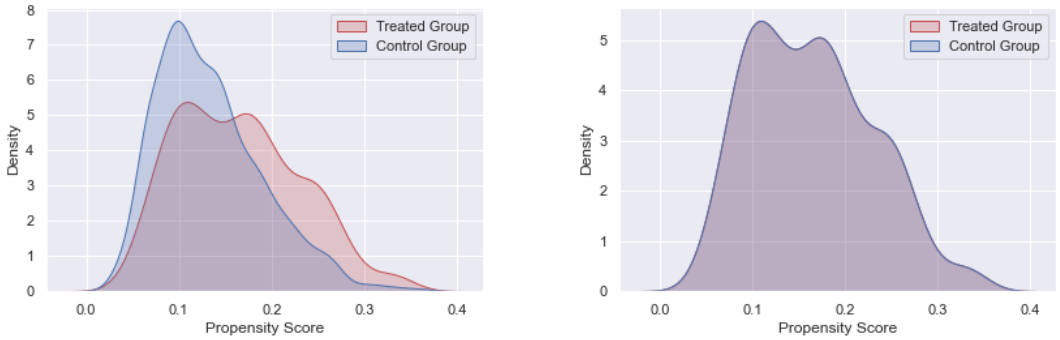


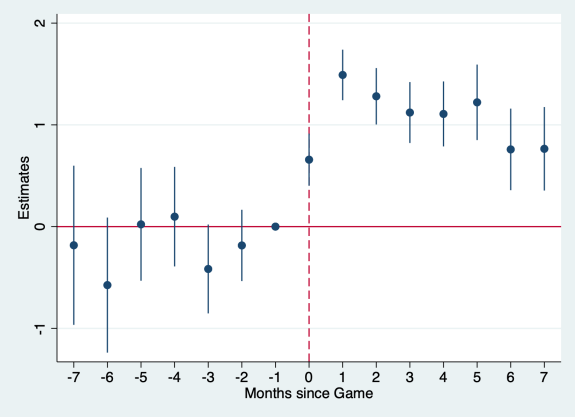
Figure C.2: PSM: Balance Improvement Between Treated and Control Groups



(a) Before Match (b) After Match

Note. The X-axis denotes the propensity score.

Figure C.3: Evidence of Parallel Trends Assumption for Table 5.14 Results



Note. We use 95% confidence interval.

VITA

Yu Rain Kan is a researcher specializing in Information Systems. Her research interests revolve around AI, Economics, and Policy. She employs a range of methodologies such as Econometrics, Experimental Design, and Machine Learning to promote and explore the impact of digital transformation. She received her Ph.D. degree from the Michael G. Foster School of Business at the University of Washington (UW). Prior to joining UW, she completed her Master's degree at Georgia Institute of Technology and her B.S. from Peking University. As of July 2024, she will be joining the Department of Operations and Decision Technologies at the Kelley School of Business, Indiana University, as an Assistant Professor. She welcomes your comments to `rainkan@iu.edu`.