KOÇ UNIVERSITY

DOCTORAL THESIS

Essays on Online Product Ratings

Author:
Ali ÇAKAL

Supervisor:
Assoc. Prof. Ali Umut GÜLER

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

Business Administration
Quantitative Marketing

January 3, 2025

Essays on Online Product Ratings

Koç University

Graduate School of Business

This is to certify that I have examined this copy of a doctoral dissertation by

Ali Çakal

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Assoc. Prof. Ali Umut Güler (Advisor)

Prof. Ayşegül Özsomer Tunalı (Second Advisor)

Prof. Seda Ertaç Güler

Prof. Nükhet Harmancıoğlu

Asst. Prof. Zeynep Müge Güzel

Asst. Prof. Mert Gümren

Date: ________________________

Declaration of Authorship

I, Ali Çakal, declare that this thesis titled, "Essays on Online Product Ratings" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"The only true voyage of discovery, the only fountain of Eternal Youth, would be not to visit strange lands but to possess other eyes..."

Narrator in *In Search of Lost Time*

KOÇ UNIVERSITY

Abstract

Graduate School of Business
Quantitative Marketing

Doctor of Philosophy

Essays on Online Product Ratings

by Ali ÇAKAL

This dissertation focuses on the drivers of online product ratings. Existing literature shows that self-selection, social influence, and rating systems influence product ratings and reviews. By documenting the impact of multi-dimensional rating systems on product ratings and exploring strategic rating behavior on online platforms in two separate essays, I aim to implement and extend the literature on the influencers of online ratings and reviews.

In the first study, I explore the effects of multi-dimensional rating systems on product ratings and review content in online platforms. Using empirical data, I demonstrate that the introduction of multi-dimensional rating systems leads to lower overall ratings compared to traditional single-dimensional systems. Additionally, I find that while users write fewer reviews in multi-dimensional systems, those reviews exhibit higher quality, with a greater focus on product-specific attributes rather than seller-specific attributes. These findings suggest that multi-dimensional rating systems not only provide a more granular understanding of product performance but also encourage more detailed and informative feedback. This study has significant implications for the design of rating systems, highlighting how structural changes can influence user behavior and enhance the usefulness of online reviews for both consumers and platform managers. By uncovering these dynamics, the study contributes to the growing literature on online consumer behavior and the optimization of rating systems.

In the second study, I examine whether strategic rating behavior exists when rating products online, focusing on how users' evaluations are influenced by the number and valence of prior ratings. I present evidence that users tend to provide lower ratings to products with a higher number of previous ratings, which suggests strategic efforts to affect overall ratings. This behavior is not explained by either herding or differentiation effects. Additionally, I find that while ratings are subject to strategic behavior, review sentiment and length remain unaffected, offering a more authentic reflection of user experiences. These findings have important implications for the design of rating systems on e-commerce platforms, emphasizing the need to mitigate biases and enhance the reliability of ratings. The study contributes to the literature on the drivers of online ratings by uncovering a novel form of social influence —strategic rating behavior— that adds a new dimension to the understanding of rating dynamics and highlights the value of written reviews as a source of more genuine consumer feedback.

Overall, this dissertation deepens our understanding of the drivers of online product ratings by examining both the structural influences of rating systems and the behavioral dynamics of users. It emphasizes the critical role of system design in shaping ratings and review quality, as well as the strategic behavior that can affect user evaluations. These insights not only contribute to the theoretical literature on online ratings but also provide practical guidance for improving the reliability and authenticity of user feedback on e-commerce platforms, benefiting both researchers and industry professionals.

KOÇ UNIVERSITY

Özetçe

Graduate School of Business
Quantitative Marketing

Doctor of Philosophy

Essays on Online Product Ratings

by Ali ÇAKAL

Bu tez, çevrimiçi ürün puanlamalarını etkileyen faktörleri incelemektedir. Mevcut literatür, öz-seçim, sosyal etki ve ürün puanlama sistemlerinin ürün puanlamaları ve yorumları üzerinde etkili olduğunu göstermektedir. Çok boyutlu ürün puanlama sistemlerinin ürün puanlamaları üzerindeki etkisini belgeleyen ve çevrimiçi platformlarda stratejik puanlama davranışını araştıran iki ayrı makale ile, çevrimiçi ürün puanlamaları ve yorumlarını etkileyen faktörler ile ilgili literatüre katkı yapmayı amaçlıyorum.

İlk çalışmada, çok boyutlu ürün puanlama sistemlerinin e-ticaret platformlarındaki ürün puanlamaları ve yorumları üzerindeki etkileri incelenmektedir. Ampirik veriler kullanılarak, çok boyutlu ürün puanlama sistemlerinin, geleneksel tek boyutlu sistemlere kıyasla genel puanlamalarda azalmaya yol açtığı gösterilmektedir. Ayrıca, kullanıcıların çok boyutlu ürün puanlama sistemlerinde daha az ürün yorumu yazmalarına rağmen, bu yorumların daha yüksek kalitede olduğu ve daha çok ürünle ilgili özelliklere odaklandığı bulunmuştur. Bu bulgular, çok boyutlu derecelendirme sistemlerinin daha ayrıntılı ve bilgilendirici geri bildirimleri teşvik ettiğini göstermektedir. Bu çalışma, ürün derecelendirme sistemlerinin tasarımı için önemli sonuçlar doğurmakta ve yapısal değişikliklerin kullanıcı davranışını nasıl etkileyebileceğini ve çevrimiçi incelemelerin hem tüketiciler hem de platform yöneticileri için nasıl daha faydalı hale getirilebileceğini vurgulamaktadır. Bu dinamikleri ortaya koyarak, çalışma, çevrimiçi tüketici davranışı ve ürün değerlendirme sistemlerinin optimizasyonu üzerine büyüyen literatüre katkıda bulunmaktadır.

İkinci çalışmada, kullanıcıların ürünleri çevrimiçi olarak puanlama davranışlarında stratejik davranıp davranmadıkları incelenmekte, özellikle kullanıcıların puanlarının, ürüne kendilerinden önce verilen puan sayısı ve valansı tarafından nasıl etkilendiği üzerinde durulmaktadır. Araştırma bulguları, kullanıcıların daha fazla puan toplamış ürünlere daha düşük puanlar verdiklerini ve bunun, ürünlerin ortalama puanlarını etkilemeye yönelik stratejik bir çaba olduğunu göstermektedir. Bu davranışın, sürü etkisi veya farklılaşma etkileri ile açıklanmadığı görülmektedir. Ayrıca, puanlamadaki stratejik davranışa rağmen, ürün yorumlarının uzunluğu ve sentimentinin ürüne daha önce verilen puanlardan etkilenmediği görülmektedir ve bu da ürün yorumlarının puanlara göre kullanıcı deneyimlerinin daha otantik bir yansımasını sunduğunu göstermektedir. Bu bulgular, e-ticaret platformlarındaki ürün puanlama sistemlerinin tasarımı açısından önemli çıkarımlar sunmakta, ürün puanlama sistemlerinin güvenilirliğini artırmak gerektiğini vurgulamaktadır. Bu çalışma, çevrimiçi ürün puanlamalarını etkileyen faktörler literatürüne, stratejik puanlama davranışını ekleyerek, ürün puanlama dinamiklerinin anlaşılmasında yeni bir boyut eklemek gibi bir katkı sağlamaktadır ve ürün yorumlarının ürün puanlamalarına göre daha gerçekçi bir tüketici geri bildirimi sağladığını vurgulamaktadır.

Genel olarak, bu tez, hem puanlama sistemlerinin yapısal etkilerini hem de kullanıcıların davranışsal dinamiklerini inceleyerek çevrimiçi ürün puanlamalarının yönlendiricileri hakkındaki anlayışımızı derinleştirmektedir. Puanlama ve yorum kalitesini şekillendirmede sistem tasarımının kritik rolünü vurgulamakta ve aynı zamanda kullanıcı değerlendirmelerini etkileyebilecek stratejik davranışları ele almaktadır. Bu bulgular, çevrimiçi puanlamalar üzerine teorik literatüre katkı sağlamakla kalmayıp, aynı zamanda e-ticaret platformlarında kullanıcı geri bildirimlerinin güvenilirliğini ve özgünlüğünü artırmaya yönelik pratik rehberlik sunarak hem araştırmacılara hem de endüstri profesyonellerine fayda sağlamaktadır.

Acknowledgements

I would like to thank Ali Umut Güler for his excellent supervision. I am indebted to the Güler Family, Ayşegül Özsomer Tunalı, Nükhet Harmancıoğlu and all of my Ph.D. friends for their good hearts and support...

This thesis was supported by the TÜBİTAK-BİDEB 2211 National PhD Scholarship Program, for which I am sincerely grateful.

Contents

List of Figures

List of Tables

To my patient and loving parents, Hülya and Yakup Çakal

Chapter 1

Introduction

As online sales accelerate, online product reviews have become more important than ever. Online product reviews provide a succinct and convenient way to understand customers' experiences with products. Furthermore, these reviews drive sales, as documented in the literature on online reviews (Chevalier and Mayzlin, 2006, Clemons, Gao, and Hitt, 2006, Chintagunta, Gopinath, and Venkataraman, 2010). Therefore, eliciting truthful and high-quality reviews is crucial for better-informed purchase decisions.

Given the significance of online product reviews, it is important to understand the factors influencing them. The literature on the drivers of online product reviews has identified three main factors: self-selection, social influence, and the rating system under which product ratings are generated.

Under *self-selection*, existing research examines the evolution of book ratings over time and shows that early raters tend to give higher ratings than later raters. This phenomenon is attributed to self-selection: early raters are often more enthusiastic readers of the rated books and thus provide higher ratings. Alternatively, lower ratings from later raters may reflect unmet expectations resulting from the higher ratings given by early raters (Li and Hitt, 2008, Wu and Huberman, 2008, Hu and Li, 2011).

Under *social influence*, prior studies identify herding (Lee, Hosanagar, and Tan, 2015, Wang, Zhang, and Hann, 2018) and differentiating (Moe and Schweidel, 2012) behaviors in response to others' ratings. Herding refers to aligning one's ratings with those of others, while differentiating involves deliberately diverging from others' ratings.

Regarding the impact of *rating systems* on ratings, previous studies present mixed findings. For example, Chen, Hong, and Liu, 2018 found that the overall ratings increased when Tripadvisor transitioned from a single-dimensional rating system to a multi-dimensional one. In contrast, Schneider et al., 2021, through a series of lab experiments, found mixed results concerning overall ratings under multi-dimensional versus single-dimensional systems.

In Chapter 2, I provide evidence from an e-commerce website's transition from a single-dimensional to a multi-dimensional rating system, showing that overall ratings decreased following the transition. Furthermore, users provided less written reviews under multi-dimensional rating system, suggesting a substitutability between attribute ratings and written reviews. Additionally, reviewers produced higher-quality reviews under the multi-dimensional system, mentioning more product-related attributes and fewer seller-related aspects in their reviews. These findings extend existing research on the impact of rating systems on ratings and reviews and offer insights into designing rating systems that elicit higher-quality reviews.

In Chapter 3, I examine a new dimension of social influence in online ratings: strategic rating. Using user and product rating data from a beer-rating website, I show that users tend to give lower ratings to products with more and higher ratings. However, the length and sentiment of their reviews appear unaffected. These findings also provide guidance for designing rating systems to elicit more truthful ratings.

Finally, in Chapter 4, I offer concluding remarks and highlight how understanding the drivers of online product ratings can enhance rating system design, improve review quality, and promote more informed consumer decisions.

References

Chen, Pei-Yu, Yili Hong, and Ying Liu (2018). “The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments”. In: *Management Science* 64.10, pp. 4629–4647.

Chevalier, Judith A and Dina Mayzlin (2006). “The effect of word of mouth on sales: Online book reviews”. In: *Journal of marketing research* 43.3, pp. 345–354.

Chintagunta, Pradeep K, Shyam Gopinath, and Sriram Venkataraman (2010). “The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets”. In: *Marketing science* 29.5, pp. 944–957.

Clemons, Eric K, Guodong Gordon Gao, and Lorin M Hitt (2006). “When online reviews meet hyperdifferentiation: A study of the craft beer industry”. In: *Journal of management information systems* 23.2, pp. 149–171.

Hu, Ye and Xinxin Li (2011). “Context-dependent product evaluations: an empirical analysis of internet book reviews”. In: *Journal of Interactive Marketing* 25.3, pp. 123–133.

Lee, Young-Jin, Kartik Hosanagar, and Yong Tan (2015). “Do I follow my friends or the crowd? Information cascades in online movie ratings”. In: *Management Science* 61.9, pp. 2241–2258.

Li, Xinxin and Lorin M Hitt (2008). “Self-selection and information role of online product reviews”. In: *Information systems research* 19.4, pp. 456–474.

Moe, Wendy W and David A Schweidel (2012). “Online product opinions: Incidence, evaluation, and evolution”. In: *Marketing Science* 31.3, pp. 372–386.

Schneider, Christoph et al. (2021). “When the stars shine too bright: The influence of multidimensional ratings on online consumer ratings”. In: *Management Science* 67.6, pp. 3871–3898.

Wang, Chong, Xiaoquan Zhang, and Il-Horn Hann (2018). “Socially nudged: A quasi-experimental study of friends’ social influence in online product ratings”. In: *Information Systems Research* 29.3, pp. 641–655.

Wu, Fang and Bernardo A Huberman (2008). “How public opinion forms”. In: *International Workshop on Internet and Network Economics*. Springer, pp. 334–341.

Chapter 2

How Do Multi-Dimensional Rating Systems Affect Consumer Feedback? An Empirical Investigation

2.1 Introduction

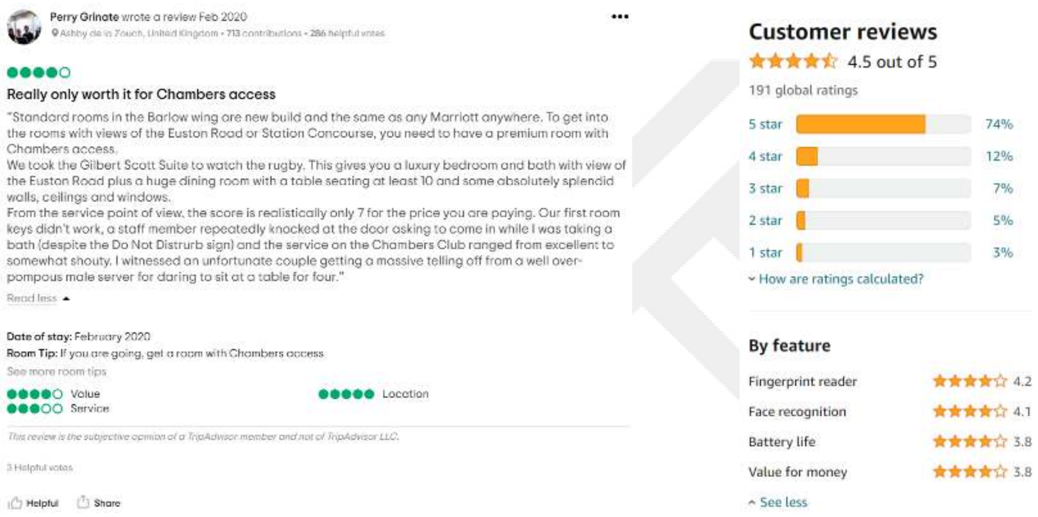
Product ratings and reviews are pivotal in shaping customers' purchase decisions on e-commerce platforms. With the share of internet sales in total retail sales doubling from 17% in 2018 to 34% in 2021 (Office for National Statistics, 2021), the importance of online word-of-mouth (WOM) continues to grow. Notably, 93% of customers report reading online reviews before making a purchase (Qualtrics, 2021). Reflecting this trend, extensive research has demonstrated strong links between online WOM—both ratings and reviews—and sales (Chevalier and Mayzlin, 2006, Yong Liu, 2006, Dellarocas, X. Zhang, and Awad, 2007, Duan, Gu, and Whinston, 2008, Chintagunta, Gopinath, and Venkataraman, 2010, Clemons, Gao, and Hitt, 2006, Y. Chen, Q. Wang, and Xie, 2011, Cho, Sosa, and Hasija, 2022, L. Li, Gopinath, and Carson, 2022, N. Lee, Bollinger, and Staelin, 2023) as well as business success (Naumzik, Feuerriegel, and Weinmann, 2022).

In addition to understanding the impact of WOM on sales, researchers have explored factors shaping ratings and reviews. These factors include social influence (Moe and Trusov, 2011, Muchnik, Aral, and Taylor, 2013, Y.-J. Lee, Hosanagar, and Tan, 2015, C. Wang, X. Zhang, and Hann, 2018, Bhukya and Paul, 2023), visibility of reviews (Schlosser, 2005, Goes, M. Lin, and Au Yeung, 2014, Huang, Hong, and Burtch, 2017), peer review volume (Burtch et al., 2018), user characteristics (e.g., early vs. late adopters) (F. Wu and Huberman, 2008, X. Li and Hitt, 2008), product popularity (Hu and X. Li, 2011), review environments (Moe and Schweidel, 2012), the disconfirmation effect (Ho, J. Wu, and Tan, 2017), and the design of review systems (P.-Y. Chen, Hong, and Ying Liu, 2018, Schneider et al., 2021, Lembregts, Schepers, and Keyser, 2024). This study contributes to the last area by examining the effect of transitioning from a single-dimensional to a multi-dimensional rating system on an e-commerce platform.

2.1.1 Multi-Dimensional Rating Systems in Practice

Traditional single-dimensional rating systems solicit an overall product rating. Recently, platforms such as TripAdvisor, Amazon, and Capterra have adopted multi-dimensional rating systems that elicit ratings for specific product attributes. For instance, TripAdvisor allows users to rate attributes such as *value*, *location*, *service*, and *cleanliness* (Figure 2.1.A) (P.-Y. Chen, Hong, and Ying Liu, 2018), while Amazon includes varied attribute ratings depending on the product category (Figure 2.1.B). Capterra, a software rating platform, asks users to evaluate *ease of use*, *customer service*, *features*, and *value for money*. Similarly, many local e-commerce platforms now incorporate multi-dimensional rating systems for specific product categories.

Multi-dimensional rating systems vary in implementation. Some platforms make attribute ratings optional (e.g., RateBeer), while others require users to submit overall ratings and written reviews but leave attribute ratings optional (e.g., TripAdvisor, Amazon). Others mandate ratings for specific attributes while keeping the rest optional (e.g., Capterra). Differences in system design across platforms and product categories suggest that rating system structures likely influence both ratings and review content. Understanding these causal effects is essential for creating systems that elicit truthful and unbiased reviews.



(A) TripAdvisor (B) Amazon

FIGURE 2.1: Multi-Dimensional Rating Systems on TripAdvisor and Amazon

2.1.2 Research Questions and Key Findings

This study explores how introducing a multi-dimensional rating system affects overall ratings and reviews. Specifically, we analyze the transition of a major Turkish e-commerce platform from a single-dimensional to a multi-dimensional rating system in its mobile phone category. The new system, implemented on December 24, 2020, includes four attributes: battery, camera, screen, and processor. Users must first provide an overall rating, after which attribute rating options and a review text box appear. Rating attributes and writing reviews are optional, allowing for a quasi-natural experiment with the pre-change data serving as the control group and post-change data as the treatment group.

The transition to a multi-dimensional rating system led to a decline in overall ratings, primarily driven by users who opted to rate specific attributes. Interestingly, all attribute ratings were consistently lower than the overall ratings in both single- and multi-dimensional systems. This indicates that the inclusion of lower-evaluated attributes in the rating process pulls down overall ratings, even when users provide the overall rating first.

The adoption of the multi-dimensional system reduced the likelihood of users writing reviews, suggesting a substitutive relationship between attribute ratings and written reviews. However, the content of reviews shifted significantly: mentions of the rated attributes increased, reflecting a heightened focus on product-specific attributes. Furthermore, the likelihood of discussing product attributes that are not included in the multi-dimensional rating system slightly increased. Conversely, mentions of seller-related attributes declined significantly. Despite the decline in review frequency, the average length of reviews, measured by word count, remained stable. This indicates that while fewer reviews were written, their relevance and quality improved.

2.1.3 Contributions

This study demonstrates that multi-dimensional rating systems can reduce overall ratings, especially when they include attributes that receive lower evaluations. Notably, this effect occurs even when overall ratings are collected prior to attribute ratings, challenging assumptions about the sequence of rating inputs.

By analyzing the influence of multi-dimensional rating systems on review content, this research underscores how rating system design shapes the aspects that reviewers emphasize in their reviews, thereby enriching our understanding of online word-of-mouth dynamics.

The findings contribute to the literature on review system design (Gutt et al., 2019, Jiang and Guo, 2015, Zheng et al., 2018, Tunc, Cavusoglu, and Raghunathan, 2017) by providing empirical evidence of how changes to rating systems affect reviewer behavior. These insights are valuable for managers seeking to optimize their platforms for more truthful and unbiased feedback.

Unlike most previous studies focusing on experience goods like books, movies, and restaurants, this research examines a search good: mobile phones. The tangible quality cues associated with search goods create distinct rating dynamics, offering unique insights into how multi-dimensional systems influence reviewer behavior in this context.

The rest of the paper is organized as follows. In Section 2.2, we review the literature relevant to our paper. In Section 2.3, we present the data, descriptive statistics, and the empirical setting. We put forth our empirical strategy in Section 2.4, followed by the results and robustness checks in Section 2.5 and 2.6, respectively. In Section 2.7, we summarize our findings and discuss its scholarly and managerial implications. Finally, in Section 2.8, we conclude.

2.2 Relevant Literature

2.2.1 Online Ratings

Previous research identifies three key factors influencing online ratings: social influence, self-selection, and review system design. Studies on social influence consistently demonstrate that observing others' ratings impacts the valence of a reviewer's own rating. For example, Muchnik, Aral, and Taylor, 2013, in a large-scale field experiment on a news aggregation website, find significant positive herding effects in some categories (e.g., politics, business, culture, and society) but not in others (e.g., economics, IT, fun). They also attribute this effect to "friends" (users who previously liked the poster's comments), rather than "enemies" (those who disliked them).

Similarly, Y.-J. Lee, Hosanagar, and Tan, 2015 examine the effects of prior ratings by strangers versus friends on a social movie website and find that users align their ratings more closely with their friends. Moreover, they observe that herding behavior after seeing a crowd rating decreases as the number of friends rating a movie increases. Expanding on this, C. Wang, X. Zhang, and Hann, 2018 analyze data from an online rating platform for books, movies, and songs, revealing that friendships formed on the platform lead to more similar ratings. They also show that social influence is stronger among users with smaller friend networks.

Another stream of research explores temporal trends in ratings, often attributing them to self-selection effects. For instance, F. Wu and Huberman, 2008 examine the evolution of book ratings on Amazon, noting that later raters deviate from earlier ones. X. Li and Hitt, 2008 also find that early raters tend to provide higher ratings than later raters, attributing this to differences in preference structures between early and late consumers. Similarly, Hu and X. Li, 2011 study Amazon book reviews, showing that later reviews often contradict earlier ones, attributing this divergence to self-selection bias or expectations shaped by previous reviews.

Research on multi-dimensional rating systems provides insights directly relevant to our work. P.-Y. Chen, Hong, and Ying Liu, 2018 argue, based on the expectancy-disconfirmation theory, that multi-dimensional systems lead to higher overall ratings due to their informativeness. Examining Tripadvisor's transition to a multi-dimensional system, they find an increase in overall ratings. Conversely, Schneider et al., 2021 conduct lab experiments and argue that multi-dimensional systems increase the accessibility of certain attributes, which can either elevate or lower overall ratings, depending on the attributes introduced. Their findings suggest that overall ratings need not necessarily increase in a multi-dimensional system, as shown by P.-Y. Chen, Hong, and Ying Liu, 2018, but may decrease under specific conditions.

Our research contributes to the existing body of work by demonstrating that overall ratings differ significantly between multi- and single-dimensional rating systems. In contrast to the findings of P.-Y. Chen, Hong, and Ying Liu, 2018, our results reveal that overall ratings do not necessarily increase in multi-dimensional systems. Instead, they may decrease when lower-evaluated attributes are incorporated into the rating framework.

Furthermore, our study highlights the interaction between rating systems and written reviews. We provide evidence that attribute ratings in multi-dimensional systems can act as substitutes for written reviews. Despite this substitution effect, the informativeness and quality of the remaining reviews appear to improve, as reviewers focus more on product-specific attributes rather than seller-related attributes.

By extending the literature on rating systems, our findings underscore the critical role of attribute selection in designing multi-dimensional systems. Additionally, we complement prior research by elucidating how rating system design influences not only overall ratings but also the quality and focus of written reviews.

2.2.2 Review Content

Previous research has demonstrated that social observability significantly influences what reviewers include in their reviews. For instance, leveraging the integration of TripAdvisor and Yelp with Facebook, Huang, Hong, and Burtch, 2017 show that when reviews become visible to a reviewer's Facebook connections, reviewers tend to express fewer negative and more positive emotions in their reviews. Similarly, Goes, M. Lin, and Au Yeung, 2014 use networking data from a product review platform to reveal that as the number of a reviewer's "subscribers" increases, they tend to use fewer emotional words, thereby increasing the objectivity of their reviews. Additionally, C. Wang, X. Zhang, and Hann, 2018 find that an exogenous increase in the user population of an online review site led to higher-quality reviews, as measured by the number of helpfulness votes they received.

Self-related factors also shape review content. For example, Pu et al., 2020 demonstrate that identity disclosure reduces the frequency of reviews but increases the effort invested in them. Meanwhile, Proserpio and Zervas, 2017 show that when hotels respond to customer reviews, reviewers produce fewer but longer reviews, suggesting they feel compelled to justify negative feedback. Berger, 2014 offers a comprehensive review, discussing how impression management and emotion regulation shape the content of reviews.

Departing from prior research that primarily examines self- or social-related factors, we investigate how transitioning to a multi-dimensional rating system influences review content. Specifically, we analyze changes in the likelihood of reviewers mentioning attributes included in the multi-dimensional system, as well as other product- and seller-related attributes, following the switch from a single-dimensional system. On the one hand, the increased salience of attributes in the multi-dimensional system might lead reviewers to mention these attributes more often. On the other hand, rating these attributes could potentially substitute for mentioning them in reviews. Our findings indicate that reviewers are more likely to mention these attributes in their reviews, even after explicitly rating them.

Additionally, we observe an increase in mentions of other product-related attributes, while the likelihood of mentioning seller-related attributes declines sharply. These results suggest that the introduction of a multi-dimensional rating system shifts the focus of reviews toward the product itself, rather than the seller or platform. When prompted to review a recent purchase, reviewers might face ambiguity, unsure whether to critique the product or the e-commerce platform. By introducing attribute-specific ratings, the multi-dimensional system appears to resolve this confusion, encouraging reviewers to focus on the product.

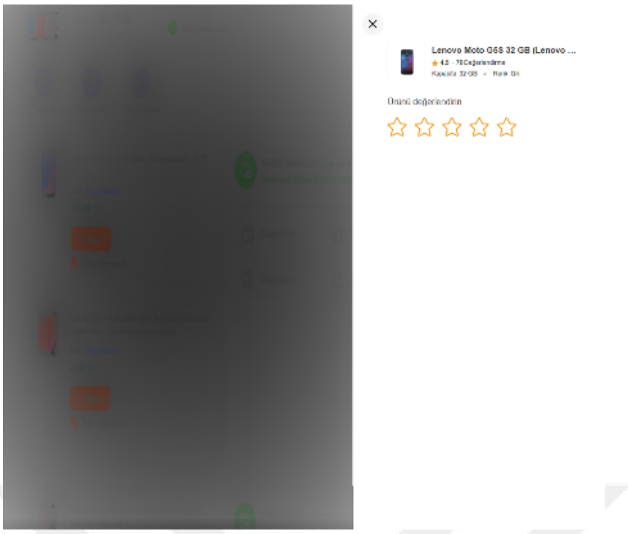
2.3 Empirical Setting

2.3.1 Rating System Change in the E-Commerce Website

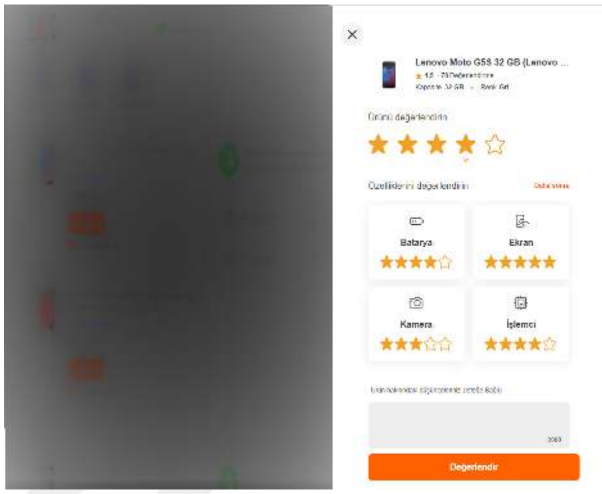
We collected data from a major Turkish e-commerce website, which sells all kinds of goods. This website switched to a multi-dimensional rating system for its mobile phones on December 24, 2021 for which it had a single-dimensional rating system before. In the single-dimensional rating system, customers who wanted to review a mobile phone had to first provide an overall rating (1-5 stars), and then it was optional to write a review.

In the multi-dimensional rating system, reviewers still have to provide an overall rating (1-5 stars) while rating attributes (1-5 stars) and writing a review are optional. The attributes included in the multi-dimensional rating system are *battery*, *camera*, *screen*, and *processor*.

As is shown in Figure 2.2, the design of the multi-dimensional rating system is such that those customers who want to review a mobile phone first see the screen in Figure 2.2.A where they can only see the overall ratings section. After they provide an overall rating, the attribute ratings and review writing sections pop up on the same screen (Figure 2.2.B) and rating these attributes as well as writing a review are optional. This optionality is indicated by the small texts on the upper-right corners of the attribute rating and review parts, which say ‘Later’ and ‘Optional’, respectively.



(A) First-Stage: Only the Overall Rating Section is on the Screen



(B) Second-Stage: After Providing Overall Rating, Attribute Ratings and Review Writing Sections Pop Up on the Same Screen

FIGURE 2.2: Multi-Dimensional Rating System Implemented on Mobile Phone Category in the E-Commerce Website

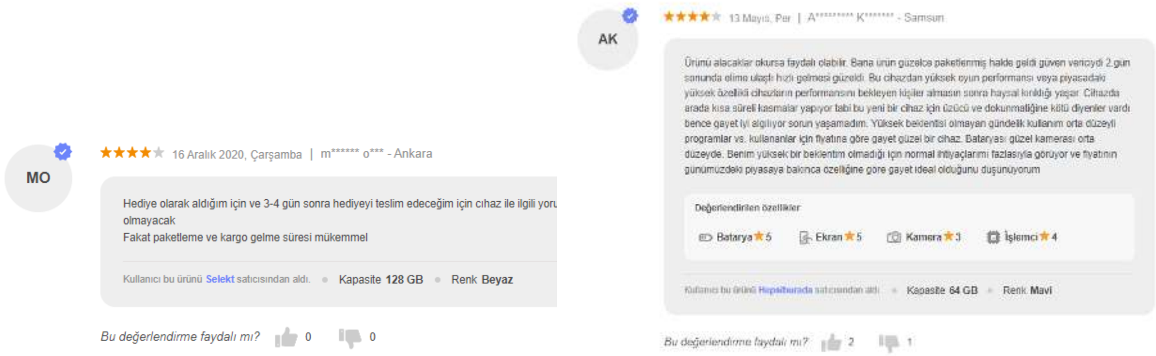
We collected all customer reviews from 2010 to 2021 for 295 mobile phone models on the e-commerce website for a total of 130,206 reviews. Before the rating system change, each review contained information on the initials of the reviewer, the city she lived, her age, whether she purchased the mobile phone, the date of the review, the overall rating she provided, and a written review if the reviewer wrote one (Figure 2.3.A).

After the rating system change, reviews consist of the same information and in addition, they contain the attribute ratings if the reviewer opted to rate them (Figure 2.3.B). Rating attributes is flexible on the platform such that the reviewers may opt to rate either one, two, three or all four attributes.

Another feature of the multi-dimensional rating system is that once a reviewer clicked on a rating for an attribute, she cannot take it back (unrate) unless she closes the window and logins again to review the mobile phone. However, she can change her rating if she clicks on another rating (e.g. she can switch from a 3-star rating to a 4-star rating).

To examine the changes in the content of the reviews upon the implementation of the multi-dimensional rating system, we created various dummy variables which indicate whether a certain attribute is mentioned in the review. These dummy variables equal 1 if the reviewer mentions the specific attribute in her review and 0 otherwise. Table 2.1 shows the variables created and which words were used to determine the value of the dummy variable.

As an example, we created *ReviewBattery* variable which equals 1 if the review includes the words battery or charge (and some misspellings of these). These variables include the attributes included in the multi-dimensional rating system (*ReviewBattery*, *ReviewCamera*, *ReviewScreen* and *ReviewProcessor*), other product-related



(A) A Review from *Before* the Rating System Change (Single-Dimensional Rating System)

(B) A Review from *After* the Rating System Change (Multi-Dimensional Rating System)

FIGURE 2.3: Exemplar Reviews from Single- and Multi-Dimensional Rating Systems

attributes (*ReviewPrice* and *ReviewOtherTechnical*), and a platform/seller-related attribute (*ReviewShipment*).

	Variable	Words
Attributes included in the multi-dimensional rating system	*ReviewBattery*	battery, charge
	ReviewCamera	camera, photo, picture
	ReviewScreen	screen
	ReviewProcessor	processor
Other product-related attributes that are not included in the multi-dimensional rating system	*ReviewPrice*	price
	ReviewOtherTechnical	ram, gps, touch, face, memory, bluetooth, application, game, voice, cable
Seller-related attribute that is not included in the multi-dimensional rating system	*ReviewShipment*	shipment

TABLE 2.1: Variables Indicating whether a Certain Attribute is Mentioned in a Review

2.3.2 Descriptive Statistics

The key statistics of the data are displayed in Table 2.2. Before the rating system change (in the single-dimensional rating system), the mean overall rating equals 4.62 and 81% of reviewers had provided a written review. However, when we closely look at the data, we see that before September 4, 2020, all reviewers had written a review, hence most likely, it was not optional before that date to write a review but mandatory. Before it became optional to write a review, the mean overall rating equals 4.57 while between September 4, 2020 and the date of the rating system change, December 24, 2020, the reviewers provided an average of 4.69 points for the mobile phones. Between it became optional to write a review and the implementation of the multi-dimensional rating system, 53% of reviewers opted to write a review. After the rating system change (in the multi-dimensional rating system), the mean overall rating is 4.64 and 35% of reviewers opted to write a review.

	OverallRating	*YesReview*	*BatteryRa*	*CameraRa*	*ScreenRa*	*ProcessorRa*
SD Rating Sys(Mand)	4.57	100%	-	-	-	-
SD Rating Sys(Opt)	4.69	53%	-	-	-	-
MD Rating Sys	4.64	35%	4.59	4.46	4.61	4.52

TABLE 2.2: Descriptive Statistics: Overall and Attribute Ratings

After the rating system change, 35% of reviewers opted to rate at least one attribute. Among these, 9.3% rated only 1 attribute, 2.4% rated either 2 or 3 attributes and the remaining 88.3% rated all 4 attributes. On average, reviewers rated battery 4.59, camera 4.46, screen 4.61 and processor 4.52.

Seemingly, attribute ratings are generally below overall ratings provided both in the single-dimensional and multi-dimensional rating systems. Corroborating these statistics, we show in later analyses that when we confine our analyses to the time before the rating system change, mentioning either of these four attributes in reviews is negatively associated with overall ratings (see Section 2.6.4).

Figure 2.4 displays the change in the overall ratings over time within mobile phone model. As a model-free evidence, we see that overall ratings exhibit an upward-sloping trend before the rating system change, but they decrease after the multi-dimensional rating system is implemented. The jump at the time of the rating system change suggests a sudden decrease in overall ratings when the e-commerce website started using a multi-dimensional rating system.

Furthermore, unlike the increasing trend of overall ratings in the single-dimensional rating system, overall ratings seem to slightly decrease or not change in the multi-dimensional rating system. Assuming that factors other than the implementation of the multi-dimensional rating system have not changed at the time of the implementation of the multi-dimensional rating system, this figure supplies us with a model-free evidence of the impact of multi-dimensional rating system on overall ratings.

Figure 2.5 shows the evolution of attribute ratings over time. Generally, attribute ratings exhibit a constant trend over time, possibly because they reflect the general consensus of customers on them, and they are less likely to be subject to the biases the overall ratings are. Such stability of attribute ratings assure that the analysis we conduct on the impact of introducing these attributes in a rating system on the overall ratings are not due to the changes in attribute ratings, but rather because of the implementation of them into the rating system.

For the statistics regarding the content of reviews, we focus on the time between it became optional to write a review and the time the multi-dimensional rating system implemented for the single-dimensional case, and the date after the implementation of the multi-dimensional rating system for the multi-dimensional case. We do not consider the content of reviews during the time it was mandatory to write a review because we base our analyses of the changes in the content of reviews on the dates when writing a review is optional.

According to Table 2.3, on average, in the single-dimensional case, 16%, 14%, 9% and 1.8% of those who wrote a review mentioned battery, camera, screen and processor in their reviews, respectively. In the case of multi-dimensional rating system,

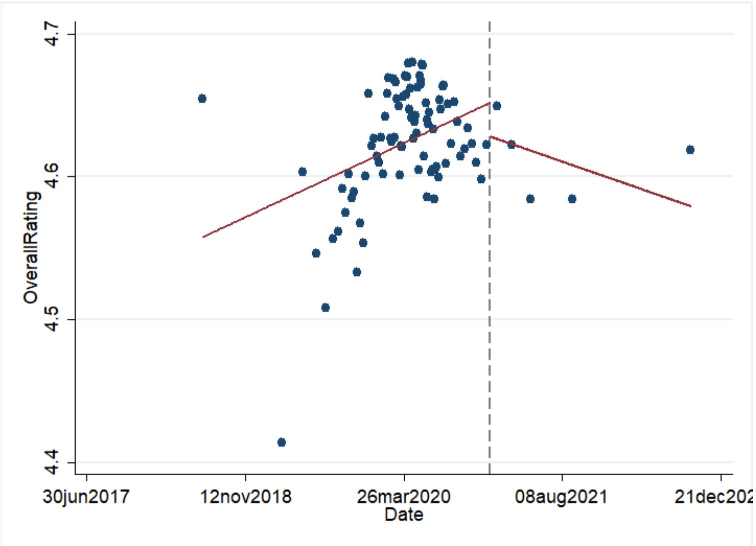
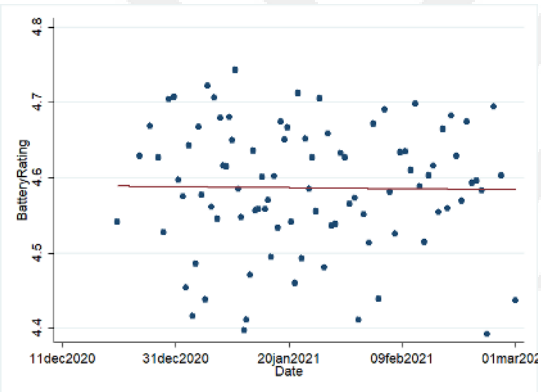
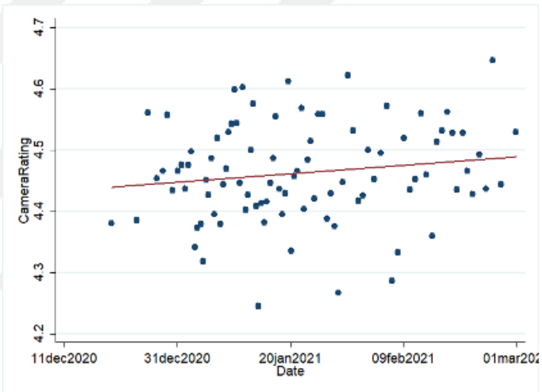


FIGURE 2.4: Change in the Overall Ratings After the Implementation of the Multi-Dimensional Rating System



(A) Battery Ratings



(B) Camera Ratings

(C) Screen Ratings

(D) Processor Ratings

FIGURE 2.5: Attribute Ratings Over Time

the corresponding numbers are 17.9%, 15%, 10.2% and 3.3% (see also Figure 2.6 for changes in the likelihood of mentioning the attributes examined after the introduction of the multi-dimensional rating system). We see a slight increase in the tendency to mention battery, camera and screen after the multi-dimensional rating system is introduced while on average, reviewers mentioned processor almost twice as much as they did in the single-dimensional regime.

	ReviBatt	*ReviCame*	*ReviScre*	*ReviProc*	*ReviPric*	*ReviTech*	*ReviShip*
SD Rat Sys(Opt)	16%	14%	9%	1.8%	7.5%	16%	35.6%
MD Rat Sys	17.9%	15%	10.2%	3.3%	8.2%	17.5%	31.6%

TABLE 2.3: Descriptive Statistics: Likelihood of Mentioning Attributes in Written Reviews

Likelihood of mentioning other-product related attributes that are not part of the multi-dimensional rating system appears to slightly increase after the rating system change. While 7.5% and 16% of reviewers mention price/performance and other technical attributes, respectively in the single-dimensional case, the corresponding numbers are 8.2% and 17.5% in the multi-dimensional case. Although raw in nature, the introduction of a multi-dimensional rating system with the battery, camera, screen and processor attributes does not seem to lead to a decrease in the likelihood of mentioning other product-related attributes.

Another relevant attribute is 'shipment'. On average, 35.6% of reviewers mention shipment in their reviews in the single-dimensional case, and this number falls to 31.6% in the multi-dimensional case. In e-commerce platforms, reviewers might be confused in that they may not be sure of what to review, the seller or the product. Although not formal, observing a sharper decrease in the likelihood of mentioning shipment suggests that the introduction of a multi-dimensional rating system may lead the reviewers to review the product, not the platform/seller.

2.4 Empirical Analysis

We investigate whether and how overall ratings change after the implementation of the multi-dimensional rating system. Our base model is:

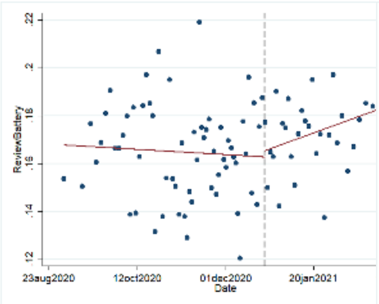
$$OverallRating_{ijt} = \beta_0 + \beta_1 BeforeAfter + \beta_2 t + \beta_3 t^2 + \alpha_j + \epsilon_{ijt} \tag{2.1}$$

where i and j index the reviewer (or his/her review) and the mobile phone model, respectively and t indexes time in days. $OverallRating_{ijt}$ is the dependent variable denoting the overall rating provided by reviewer i for mobile phone model j in day t.

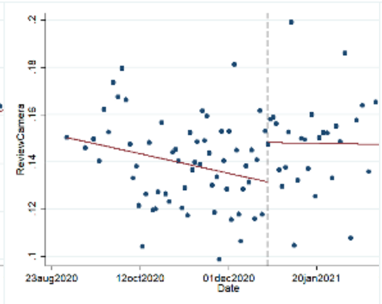
BeforeAfter is the main variable of interest and it is a dummy variable which equals 1 if the date is after the rating system change (i.e. multi-dimensional rating system), and 0 if it is before the rating system change (i.e. single-dimensional rating system). To control for mobile phone model characteristics, we include mobile phone model-fixed effects, α_j, in our model.

Mobile phone model-fixed effects allow us to inspect the change in overall ratings after the multi-dimensional rating system is introduced within mobile phone model. Furthermore, we include a time trend (t) to be able to capture the change in overall ratings around the time trend. Previous studies have shown that ratings exhibit either a downward or an upward trend over time mostly due to selection issues (see the review in Section 2.2.1). Hence, by including a time trend, we make sure that the estimated change in the overall ratings is not due to the trend in overall ratings but rather due to the implementation of the multi-dimensional rating system.

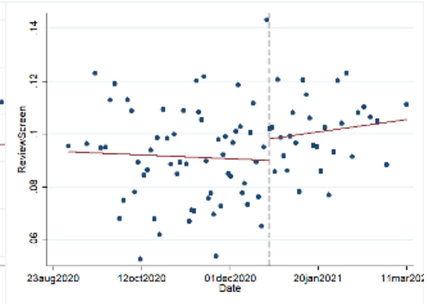
We also include t^2 to allow nonlinearity of the overall ratings over time (Figure 2.4). Viewing the overall ratings in the single-dimensional rating system as



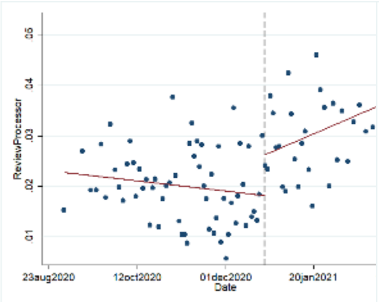
(A) Battery



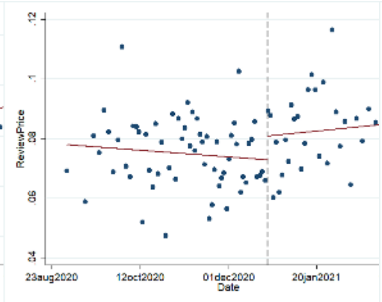
(B) Camera



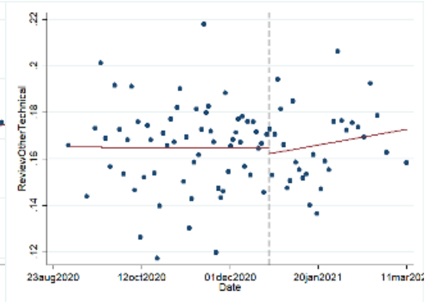
(C) Screen



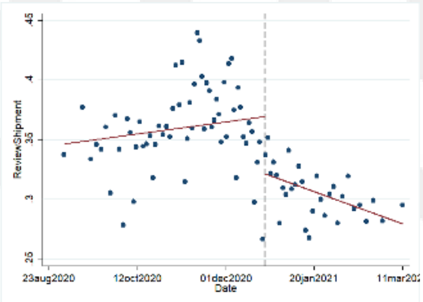
(D) Processor



(E) Price



(F) Other Technical Attributes



(G) Shipment

FIGURE 2.6: Likelihood of Mentioning Attributes Included in the Multi-Dimensional Rating System as well as Other Product- and Seller-Related Attributes Over Time

the control group and the overall ratings in the multi-dimensional rating system as the treatment group, the estimate of β_1 yields the impact of introducing the multi-dimensional rating system on overall ratings.

Our identifying assumption is that the implementation of the multi-dimensional rating system is exogenous. Specifically, to be able to interpret our results as causal, we assume that the e-commerce website did not incentivize or disincentivize the customers to review the mobile phones. In support of this assumption, we find no divergence in the number of reviews posted before and after the implementation of the multi-dimensional rating system, except for a few days (the results without including these days in Section 2.6.1).

In addition, we assume that the rating system change did not cause much difference in customers' purchasing behavior. This assumption is supported by the results that for those who did not opt to rate attributes, the decrease in overall ratings were small while overall ratings of the reviewers opting to rate attributes decreased rather sharply (Table 2.5). Hence, reviewers not rating the attributes acted more similarly as those reviewers before the rating system change.

Also, when we confine our analysis to a small time interval around the rating system change, the effect is still there. This result reassures the reliability of our main findings given that those reviewers who rated the product shortly after the implementation of the multi-dimensional rating system purchased their product before the introduction of the multi-dimensional rating system, hence they were exposed to the single-dimensional rating system during their search for the product.

We estimate the following base model to see whether and how the likelihood of mentioning certain attributes changed upon the implementation of the multi-dimensional rating system:

$$Y_{ijt|YesReview=1} = \beta_0 + \beta_1 BeforeAfter + \beta_2 t + \alpha_j + \epsilon_{ijt} \tag{2.2}$$

where Y is the attribute-specific dependent variable which we include in the analysis if the reviewer has provided a written review ($YesReview = 1$) and the rest is same as in Equation (2.1).

To investigate the changes in the likelihood of mentioning various attributes, Y becomes either *ReviewBattery*, *ReviewCamera*, *ReviewScreen*, *ReviewProcessor*, *ReviewPrice*, *ReviewOtherTechnical* or *ReviewShipment* (Table 2.1 displays which words in the reviews are used to determine the value of these dummy variables).

We make a similar identifying assumption on the exogeneity of the introduction of the multi-dimensional rating system on the likelihood of mentioning these attributes as we do for overall ratings. In support of this assumption, as can be seen in Figure 2.6, the likelihood of mentioning all the attributes we are interested in change right after the implementation of the multi-dimensional rating system.

Reviewers who review the product right after the multi-dimensional rating system is implemented have purchased the product before its implementation. Hence, they were exposed to the reviews in the single-dimensional rating system before purchase, and they review the product in a multi-dimensional rating system. Given the discrete jumps in the likelihood of mentioning each attribute around the implementation of the multi-dimensional rating system, our assumption finds at least some support.

2.5 Results

2.5.1 Effect of Multi-Dimensional Rating System on Overall Ratings

The results for the impact of the multi-dimensional rating system on overall ratings are presented in Table 2.4. The first column presents our estimate of the base model (Equation 2.1). On average, overall ratings decrease by 0.06 points after the implementation of the multi-dimensional rating systems. The negative sign of the change in overall ratings is in accordance with the mean overall ratings reported in Table 2.2.

Apparently, introducing a multi-dimensional rating system including attributes that are lower evaluated than the overall product caused a decrease in overall ratings. In Columns 2-4, we include additional control variables that may influence rating behavior. In Column 2, we include *City* dummies, in column 3 the dummy variable indicating whether the reviewer provided a written review (*YesReview*), and in column 4 both *City* and *YesReview* variables. Across all specifications, the estimate of the focal variable, *BeforeAfter*, is both statistically and quantitatively similar as in our base model.

The coefficient of time trend, t, is negative and statistically significant at 1%, in line with the findings of extant research that document a decreasing trend of ratings

VARIABLES	(1) OverallRating	(2) OverallRating	(3) OverallRating	(4) OverallRating
BeforeAfter	-0.0595***	-0.0402***	-0.0577***	-0.0391***
	(0.0108)	(0.00975)	(0.0103)	(0.00924)
t	-0.00786***	-0.00333***	-0.00817***	-0.00366***
	(0.00154)	(0.00122)	(0.00163)	(0.00140)
t^2	1.83e-07***	7.55e-08***	1.91e-07***	8.33e-08**
	(3.57e-08)	(2.83e-08)	(3.77e-08)	(3.25e-08)
UserCity		-0.000230***		-0.000228***
		(7.60e-05)		(7.67e-05)
YesReview			0.0102	0.00809
			(0.0138)	(0.0141)
Constant	88.84***	41.34***	92.16***	44.88***
	(16.59)	(13.26)	(17.51)	(15.13)
Observations	130,206	119,620	130,206	119,620
R-squared	0.001	0.001	0.001	0.001
Mobile Phone FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 2.4: Impact of Implementation of the Multi-Dimensional Rating System on Overall Ratings

over time (X. Li and Hitt, 2008, F. Wu and Huberman, 2008, Hu and X. Li, 2011). The coefficient of the square of time trend, t^2, is also statistically significant and positive across all specifications, suggesting a curvilinear evolution of ratings over time. This is in accordance with the evolution of ratings over time seen in Figure 2.4.

UserCity also appears statistically significant, possibly due to socio-economic differences across consumers living in different cities. Finally, *YesReview* is not statistically significant, suggesting that there is not significant differences across the ratings given by users who wrote a review versus not.

Table 2.5 reports the estimates of the heterogeneity across reviewers in terms of their attribute rating behavior. Specifically, in this regression, we investigate whether and how the overall ratings of the reviewers not opting to rate any attributes, opting to rate only one attribute, *Battery*, and those opting to rate all attributes in the multi-dimensional rating system differ from the overall ratings of the reviewers in the single-dimensional rating system.

The coefficients of *After * Not Rate Attributes*, *After * Rate Only Battery*, *After * Rate All Attributes* are all negative and significant at $p<0.1$. On average, reviewers opting to rate only battery or all attributes decreased their overall ratings by 0.16 and 0.11 points, respectively while the decrease is smaller for the reviewers not opting to rate attributes, 0.03. Seemingly, the impact of the multi-dimensional rating system on overall ratings is more pronounced for the reviewers opting to rate attributes.

Although smaller, it is interesting to see that those reviewers who do not opt to rate attributes also decrease their overall ratings. Possibly, just viewing the attributes caused the reviewers reevaluate the mobile phone they are rating. Given the lower ratings associated with the attributes (see Table 2.2), some of the reviewers not opting to rate attributes appear to change their overall ratings, too.

These results show the effect of treatment on those who self-select into the treatment. In other words, we estimate the impact of implementing a multi-dimensional rating system on two groups of reviewers who can either self-select into rating attributes or not. Ideally, running a randomized experiment in which reviewers are randomly assigned to treatment groups such that one group of reviewers are asked

VARIABLES	(1) OverallRating
Before	-
After * Not Rate Attributes	-0.0331***
	(0.0120)
After * Rate Only Battery	-0.166***
	(0.0304)
After * Rate All Attributes	-0.101***
	(0.0126)
t	-0.00783***
	(0.00153)
t^2	1.82e-07***
	(3.55e-08)
Constant	88.52***
	(16.51)
Observations	129,936
R-squared	0.002
Mobile Phone FE	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 2.5: Impact of Implementation of the Multi-Dimensional Rating System on the Overall Ratings of Various Types of Reviewers

to review the product in a single-dimensional rating system and another group of reviewers are asked to review the product in a multi-dimensional rating system would yield the average treatment effect of the multi-dimensional rating system on overall ratings.

In this study, we believe that those reviewers who opt to rate only one attribute provide us an opportunity to study the impact of multi-dimensional rating system on the reviewers who do not opt to rate attributes. When we look at which attributes have been rated by the reviewers opting to rate only one attribute, we see that all of these reviewers rated battery. Because battery is the first attribute displayed on the review page (Figure 2.2), we suppose that these reviewers first thought of rating attributes as mandatory, and when they realized that it is in fact optional by recognizing the small text saying ‘Later’ on the upper-right corner (Figure 2.2), they did not continue to rate other attributes.

In support of this conjecture, as we mention in Section 2.3.2, among those who rated at least one attribute, around 90% of them rated all four attributes, 9% rated only battery, and just 1% rated two or three attributes. Hence, we also analyze the impact of the multi-dimensional rating system on overall ratings for the reviewers who only rated one attribute (i.e. battery).

Interestingly, reviewers who only rate one attribute, battery, decreased their overall ratings by 0.16 points after the rating system change. Viewing these reviewers as reviewers who were not to rate any attributes but somehow ‘treated’, we see that the multi-dimensional rating system also affects such reviewers’ overall ratings.

One may argue that these reviewers became upset upon the mandatory (although not) nature of rating attributes, hence they provided a rating for battery (before realizing that rating attributes is in fact optional) without elaborating much on it. However, if they were upset, we believe that they would not turn back and change their overall ratings. In fact, they seem to have changed their overall ratings and provided lower overall ratings than the reviewers who opted to rate all attributes.

Overall, these findings provide evidence that the implementation of the multi-dimensional rating system caused a decrease in overall ratings. Moreover, not only the reviewers who opted to rate attributes lowered their overall ratings, but also the reviewers who did not opt to rate any attribute provided lower overall ratings, albeit not as low as the former group of reviewers.

Possibly, mere viewing the attributes led at least some of the reviewers who did not rate any attributes to reconsider their overall ratings and change it. Interpreting the reviewers who rated only one attribute as the reviewers who were unmotivated (as the reviewers who did not rate any attribute) but somehow 'forced' to rate one attribute (battery), we find that they also decrease their overall ratings. Furthermore, the decrease is even more pronounced than the overall ratings provided by the reviewers who rated all four attributes. This discrepancy might be attributed to the difference between the number of attributes rated (Schneider et al., 2021).

2.5.2 Effect of Multi-Dimensional Rating System on Written Reviews

Table 2.6 presents the results pertaining to the impact of multi-dimensional rating system on the likelihood of writing a review and number of words used in reviews. According to the results in Column 1, on average, reviewers provided 5.1% less written reviews in the multi-dimensional rating system. Column 2 adds t^2 to the regression to allow for nonlinearity, and the results are both qualitatively and quantitatively similar. Hence, it appears that attribute ratings substitute for written reviews in multi-dimensional rating systems.

Across Column 3-4, we investigate the change in the number of words the reviewers used in their reviews. Results from both specifications show that there is not a significant change in the number of words used in the reviews, on average.

VARIABLES	(1) YesReview	(2) YesReview	(3) NoWords	(4) NoWords
BeforeAfter	-0.0513***	-0.0674***	-0.0385	0.697
	(0.00968)	(0.00969)	(0.653)	(0.814)
t	-0.00190***	-0.203***	-0.00937	7.504*
	(0.000122)	(0.0616)	(0.00711)	(4.448)
t^2		4.52e-06***		-0.000169*
		(1.39e-06)		(0.000100)
Constant	42.86***	2,279***	228.5	-83,292*
	(2.719)	(685.4)	(158.0)	(49,426)
Observations	71,410	71,410	32,044	32,044
R-squared	0.051	0.051	0.000	0.000
Mobile Phone FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 2.6: Changes in the Likelihood of Writing a Review and the Number of Words in Reviews upon the Implementation of the Multi-Dimensional Rating System

Regarding the change in the content of reviews, Tables 2.7 and 2.8 show that the likelihood of mentioning attributes that are included in the multi-dimensional rating system increases. Reviewers mention about battery, camera, screen and processor from 1% to 2% more in the multi-dimensional rating system compared to the reviews in the single-dimensional rating system. Furthermore, reviewers also mention more about price by 1% and there is not a significant change in the likelihood of mentioning other technical attributes. Regarding the seller-related attribute, shipment, there is a substantial decrease by around 8% in the mentioning likelihood of reviewers, on average.

VARIABLES	(1) ReviewBattery	(2) ReviewCamera	(3) ReviewScreen	(4) ReviewProcessor
BeforeAfter	0.0177***	0.0208***	0.0115**	0.0182***
	(0.00664)	(0.00787)	(0.00550)	(0.00315)
t	-5.85e-05	-0.000148**	-1.17e-05	-5.82e-05
	(9.03e-05)	(6.85e-05)	(5.50e-05)	(3.61e-05)
Constant	1.463	3.436**	0.350	1.313
	(2.009)	(1.522)	(1.223)	(0.802)
Observations	32,044	32,044	32,044	32,044
R-squared	0.000	0.000	0.000	0.002
Mobile Phone FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 2.7: Impact of Implementation of the Multi-Dimensional Rating System on the Likelihood of Mentioning Attributes Included in the Multi-Dimensional Rating System

VARIABLES	(1) ReviewPrice	(2) ReviewOtherTechnical	(3) ReviewShipment
BeforeAfter	0.0103**	0.00413	-0.0871***
	(0.00464)	(0.00720)	(0.0123)
t	-3.39e-05	-1.80e-05	0.000253*
	(5.04e-05)	(7.67e-05)	(0.000137)
Constant	0.829	0.564	-5.252*
	(1.121)	(1.705)	(3.034)
Observations	32,044	32,044	32,044
R-squared	0.000	0.000	0.004
Mobile Phone FE	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 2.8: Impact of Implementation of the Multi-Dimensional Rating System on the Likelihood of Mentioning Other Product- and Seller-Related Attributes that are not part of the Multi-Dimensional Rating System

Likely, the attributes included in the multi-dimensional rating system made product-related attributes (both those included in the multi-dimensional rating system and others) more accessible to the reviewers, and platform/seller-related attributes less accessible. In other words, by making product-related attributes the main theme of a review, the multi-dimensional rating system appears to orient reviewers' attention to these attributes, and away from platform/seller-related attributes.

2.6 Robustness Checks

In this section, we present several robustness checks. First, we replicate the main regression on the impact of the multi-dimensional rating system on overall ratings without including earlier ratings of the multi-dimensional rating system. Second, we examine whether a negative trend in overall ratings already exists before the rating system change. Third, we present results from a differences-in-differences model where data for the same set of mobile phones' ratings from another major e-commerce website is used. Finally, we inspect the association between the overall ratings and whether a review mentions any one of the attributes included in the multi-dimensional rating system in the single-dimensional rating system.

2.6.1 Removing the Early Ratings in the Multi-Dimensional Rating System

It can be argued that when the multi-dimensional rating system is implemented, the e-commerce website has an incentive to take action so that mobile phone buyers would provide more reviews. These actions may include offering coupons upon submitting a review, reminding buyers more to submit a review via email campaigns, etc. In fact, such solicitation of reviews is shown to elicit different ratings than the usual flow (Karaman, 2021).

VARIABLES	(1) OverallRating	(2) OverallRating
BeforeAfter	-0.0617***	-0.0559***
	(0.0113)	(0.0114)
t	-0.00790***	-0.00803***
	(0.00155)	(0.00164)
t^2	1.84e-07***	1.87e-07***
	(3.59e-08)	(3.81e-08)
Constant	89.25***	90.77***
	(16.71)	(17.73)
Observations	127,315	123,599
R-squared	0.001	0.001
Mobile Phone FE	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 2.9: Replicating Main Analysis by Excluding First One and First Two Weeks After the Implementation of the Multi-Dimensional Rating System

In order to test whether the results change when we do not include these earlier ratings into the analysis, we reestimate Equation (2.1) without including earlier days after the multi-dimensional rating system is implemented. Table 2.9 presents results. Column 1 shows the results when we exclude first week of the multi-dimensional rating system, and Column 2 presents the results when we exclude first two weeks. The results are both qualitatively and quantitatively similar to those in Table 2.4, increasing the reliability of the main results.

2.6.2 Pre-Trend Analysis

We present results from estimating Equation (2.1) which limit the date to before the rating system change and does not include *BeforeAfter* and t^2. Table 2.10 shows that there is in fact a positive trend before the rating system change in overall ratings. In contrast with other studies which show that ratings display a negative trend over time, observing a positive trend here may be because of the nature of the good we are studying. Specifically, the increase in overall ratings over time might be the result of higher-technology phones introduced in the market over time.

VARIABLES	(1) OverallRating
t	7.58e-05**
	(3.72e-05)
Constant	2.953***
	(0.816)
Observations	97,914
R-squared	0.001
Mobile Phone FE	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 2.10: Trend in Overall Ratings Before Rating System Change

Most of the existing studies investigating online ratings rely on data from experience goods (see Section 2.2.1 for a brief review). In contrast, mobile phones are mainly search goods which are also high involvement. Given that a prospective buyer can learn about such goods easily just by searching for them, over time, the increased number of reviews and ratings might have helped customers to find better matches in terms of mobile phones. For our interest, observing a positive trend before the implementation of the multi-dimensional rating system further strengthens the reliability of the finding that multi-dimensional rating system caused decrease in overall ratings.

2.6.3 Data from Another E-Commerce Website

To validate the findings from our primary analysis, we replicated the study using rating data from another Turkish e-commerce website that features the same set of mobile phone models. Unlike the focal website, this platform employs a single-dimensional rating system where providing an overall rating is mandatory while writing a review is optional.

We estimated a differences-in-differences (DiD) model using the focal website's switch to a multi-dimensional rating system as the basis for the *BeforeAfter* dummy variable. As can be seen in Table 2.11, the key interaction term (*DiD*) yielded a coefficient of -0.0387 which is statistically significant at 10%, indicating a marginally significant decrease of approximately 0.0387 units in overall ratings for the treated group compared to the control group post-treatment, further validating our main finding.

Additionally, the pre-treatment effect (*Before*) was negligible and not statistically significant, suggesting no notable difference between groups prior to the treatment. However, the effect of being part of the treated group (*Hepsiburada*) was significant and positive, with a coefficient of 0.0450 that is statistically significant at 1%, indicating higher ratings in the focal website by approximately 0.045 units regardless of time.

VARIABLES	(1) OverallRating
DiD	-0.0387*
	(0.0212)
Before	0.00611
	(0.0191)
Hepsiburada	0.0450***
	(0.0164)
Constant	4.639***
	(0.0126)
Observations	81,093
R-squared	0.065
Mobile Phone FE	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 2.11: A Differences-in-Differences Model

2.6.4 Comparison of Ratings given by the Users that Mentioned Attributes in their Reviews versus not in the Single-Dimensional Rating System

Our main result regarding the impact of multi-dimensional rating systems on overall ratings show that overall ratings decrease in multi-dimensional rating system compared to single-dimensional rating system. In Table 2.12, we compare overall ratings given by users who mention either battery, camera, screen, or processor versus users who do not mention these attributes in their reviews before the rating system change, i.e. under single-dimensional rating system.

VARIABLES	(1) OverallRating	(2) OverallRating	(3) OverallRating	(4) OverallRating
ReviewBattery	-0.127***			
	(0.0174)			
ReviewCamera		-0.146***		
		(0.0167)		
ReviewScreen			-0.233***	
			(0.0162)	
ReviewProcessor				-0.108***
				(0.0310)
Constant	4.621***	4.624***	4.624***	4.601***
	(0.00310)	(0.00294)	(0.00175)	(0.000825)
Observations	79,578	79,578	79,578	79,578
R-squared	0.003	0.004	0.007	0.000
Mobile Phone FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 2.12: Comparison of Ratings given by the Users that Mentioned Attributes in their Reviews versus not

In line with the decline in overall ratings in multi-dimensional rating system, we find that mentioning the attributes included in the multi-dimensional rating system in reviews is associated with lower overall ratings under single-dimensional rating system. Hence, it is not surprising to see overall ratings decline after these attributes are included in the rating system.

2.7 Discussion

2.7.1 Theoretical Contributions

Biases in online ratings have been subject to inquiry by researchers. These include extremity bias (Brandes, Godes, and Mayzlin, 2019, Karaman, 2021), social influence bias (Muchnik, Aral, and Taylor, 2013, Y.-J. Lee, Hosanagar, and Tan, 2015, C. Wang, X. Zhang, and Hann, 2018), disconfirmation bias (Ho, J. Wu, and Tan, 2017) and free sampling bias (Z. Lin, Y. Zhang, and Tan, 2019). If we define bias as the deviation from the rating which the reviewer supplies without being affected by external factors that may cause the reviewer take nonproduct-related factors into consideration when evaluating a product, we cannot label the change in overall ratings caused by the multi-dimensional rating system a bias.

However, if we expand the definition of bias to include changes in the weighting of attributes that go into the overall evaluation of a product, what we present in this study constitutes a bias. Specifically, the multi-dimensional rating system causes reviewers to adjust their overall ratings so that it is close to the mean of the attribute ratings, hence not all attributes that may be included in the overall evaluation of the product are represented in the overall ratings, but only those attributes that are part of the multi-dimensional rating system. Thus, this study shows that multi-dimensional rating systems can lead to biases in overall ratings.

In addition to the impact of multi-dimensional rating system on ratings, we also examine how it affects what reviewers write in their reviews. We show that including product-related attributes in a multi-dimensional rating system leads reviewers to mention more about these attributes as well as other product-related attributes. However, we find that reviewers mention significantly less about seller-related attributes in the multi-dimensional rating system compared to the single-dimensional rating system. Therefore, this study introduces multi-dimensional rating systems as another driver of what reviewers talk about in their reviews.

Our findings challenge the findings of previous work on the impact of multi-dimensional rating systems on overall ratings. Drawing upon the higher informativeness of the multi-dimensional rating systems as opposed to the single-dimensional rating system and the expectancy-disconfirmation theory, P.-Y. Chen, Hong, and Ying Liu, 2018 propose that multi-dimensional rating systems should lead to higher overall ratings compared to the single-dimensional rating systems. They argue that multi-dimensional rating systems lead to better match between customers and restaurants, hence more satisfied customers would provide higher overall ratings compared to the single-dimensional rating systems. We show that the overall ratings can be impacted in the negative direction as well upon adopting a multi-dimensional rating system.

There may be several reasons why our results diverge. First, the review website P.-Y. Chen, Hong, and Ying Liu, 2018 study might have implemented a multi-dimensional rating system which includes attributes that are higher rated than the overall experience in restaurants. Second, their assumption that most of the prospective customers read online reviews of restaurants and make their choices upon that may not hold in real life. Finally, their data come from restaurant reviews. Viewing restaurants as experience goods, it may be possible that multi-dimensional rating systems lead to better customer-restaurant matches. However, we study reviews on mobile phones which is a high-involvement search good. Given that prospective customers can easily learn about various attributes of mobile phones, the 'better match' explanation may not be applicable here.

Schneider et al., 2021 also investigate the impact of multi-dimensional rating systems on overall ratings in a series of lab experiments. They propose accessibility-diagnosticity as the mechanism behind the change in overall ratings. However, in some of their experiments, they find contradictory findings with accessibility-diagnosticity. This study lends partial support to their accessibility-diagnosticity explanation by showing that attributes included in the multi-dimensional rating system are indeed mentioned more in reviews in the multi-dimensional rating system. However, challenging accessibility-diagnosticity, we also show that other product-related attributes are also mentioned more in reviews.

2.7.2 Managerial Implications

Product ratings in e-commerce websites are significant influencers of sales. This study shows that implementing a multi-dimensional rating system may cause a decrease in overall ratings. Other than its direct effect as lower sales, if prospective customers mistakenly attribute lower overall ratings to the seller (for example, by assuming that the seller cannot deliver her promises), lower overall ratings caused by the multi-dimensional rating system may lead customers to other e-commerce sites to purchase the same goods.

Moreover, because overall ratings represent mainly the attribute ratings in a multi-dimensional rating system, customers may end up purchasing a product that fits worse to them just because the better match product is not evaluated high in the attributes that are part of the multi-dimensional rating system, but in fact does better in other attributes that the focal customer cares most about.

Hence, review system designers should be careful in adopting multi-dimensional rating systems to avoid such adverse consequences. At least, they should find ways to mitigate the bias in online ratings caused by the multi-dimensional rating systems (for example, by including more attributes in the multi-dimensional rating system, Schneider et al., 2021).

One direct advice of our findings is that e-commerce websites should supplement their multi-dimensional rating systems by including seller-related attributes (e.g. shipment) into the rating system, because reviewers mention substantially less about seller-related attributes in their reviews in multi-dimensional rating system. For the customers who would like to learn about how fast the product would be delivered, for example, multi-dimensional rating systems have little to offer since reviewers mostly mention about product-related attributes in their reviews. Separating product- and seller-related attributes should be a goal of the review system designers.

Although multi-dimensional rating systems cause rating biases and a decrease in the likelihood of mentioning seller-related attributes, they also encourage mentioning more about product-related attributes. Given the informational role of written reviews for consumers (Pavlou and Dimoka, 2006, Butler and X. Wang, 2012) as well as firms (Timoshenko and Hauser, 2019), multi-dimensional rating systems play an important role by directing reviewers to mention more about product-related attributes.

2.8 Conclusion

This study examines the effects of transitioning from a single-dimensional to a multi-dimensional rating system on a Turkish e-commerce platform, focusing on its impact on ratings and review content. Our findings reveal that the adoption of a multi-dimensional rating system leads to lower overall ratings, influenced by the inclusion of lower-evaluated product attributes. This result underscores the significance of attribute selection in shaping overall rating dynamics and challenges the assumption that multi-dimensional systems necessarily yield higher overall ratings.

In addition to influencing ratings, the shift to a multi-dimensional rating system substantially affects review content. While the likelihood of writing reviews decreases, the quality and focus of reviews improve, with reviewers emphasizing product-specific attributes. However, mentions of seller-related attributes decline, potentially reducing the information available to customers regarding seller performance.

These findings highlight the trade-offs inherent in implementing multi-dimensional rating systems. While such systems enhance the granularity and relevance of product-related feedback, they may introduce biases in overall ratings and reduce insights into seller-related attributes. This duality underscores the importance of careful design and implementation, ensuring that rating systems provide balanced and unbiased information for consumers and sellers alike.

Future research could extend these findings by exploring the long-term effects of multi-dimensional rating systems across different product categories and their influence on customer satisfaction and purchase decisions. By building on these insights, platforms can optimize their rating systems to enhance user experience and foster trust in online marketplaces.

References

Berger, Jonah (2014). "Word of mouth and interpersonal communication: A review and directions for future research". In: *Journal of consumer psychology* 24.4, pp. 586–607.

Bhukya, Ramulu and Justin Paul (2023). "Social influence research in consumer behavior: What we learned and what we need to learn?–A hybrid systematic literature review". In: *Journal of Business Research* 162, p. 113870.

Brandes, Leif, David Godes, and Dina Mayzlin (2019). "What drives extremity bias in online reviews? Theory and experimental evidence". In: *Theory and Experimental Evidence*.

Burtch, Gordon et al. (2018). "Stimulating online reviews by combining financial incentives and social norms". In: *Management Science* 64.5, pp. 2065–2082.

Butler, Brian S and Xiaoqing Wang (2012). "The cross-purposes of cross-posting: Boundary reshaping behavior in online discussion communities". In: *Information Systems Research* 23.3-part-2, pp. 993–1010.

Chen, Pei-Yu, Yili Hong, and Ying Liu (2018). "The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments". In: *Management Science* 64.10, pp. 4629–4647.

Chen, Yubo, Qi Wang, and Jinhong Xie (2011). "Online social interactions: A natural experiment on word of mouth versus observational learning". In: *Journal of marketing research* 48.2, pp. 238–254.

Chevalier, Judith A and Dina Mayzlin (2006). "The effect of word of mouth on sales: Online book reviews". In: *Journal of marketing research* 43.3, pp. 345–354.
Chintagunta, Pradeep K, Shyam Gopinath, and Sriram Venkataraman (2010). "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets". In: *Marketing science* 29.5, pp. 944–957.
Cho, Hallie S, Manuel E Sosa, and Sameer Hasija (2022). "Reading between the stars: Understanding the effects of online customer reviews on product demand". In: *Manufacturing & Service Operations Management* 24.4, pp. 1977–1996.
Clemons, Eric K, Guodong Gordon Gao, and Lorin M Hitt (2006). "When online reviews meet hyperdifferentiation: A study of the craft beer industry". In: *Journal of management information systems* 23.2, pp. 149–171.
Dellarocas, Chrysanthos, Xiaoquan Zhang, and Neveen F Awad (2007). "Exploring the value of online product reviews in forecasting sales: The case of motion pictures". In: *Journal of Interactive marketing* 21.4, pp. 23–45.
Duan, Wenjing, Bin Gu, and Andrew B Whinston (2008). "The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry". In: *Journal of retailing* 84.2, pp. 233–242.
Goes, Paulo B, Mingfeng Lin, and Ching-man Au Yeung (2014). ""Popularity effect" in user-generated content: Evidence from online product reviews". In: *Information Systems Research* 25.2, pp. 222–238.
Gutt, Dominik et al. (2019). "Design of review systems–A strategic instrument to shape online reviewing behavior and economic outcomes". In: *The Journal of Strategic Information Systems* 28.2, pp. 104–117.
Ho, Yi-Chun, Junjie Wu, and Yong Tan (2017). "Disconfirmation effect on online rating behavior: A structural model". In: *Information Systems Research* 28.3, pp. 626–642.
Hu, Ye and Xinxin Li (2011). "Context-dependent product evaluations: an empirical analysis of internet book reviews". In: *Journal of Interactive Marketing* 25.3, pp. 123–133.
Huang, Ni, Yili Hong, and Gordon Burtch (2017). "Social network integration and user content generation". In: *MIS quarterly* 41.4, pp. 1035–1058.
Jiang, Yabing and Hong Guo (2015). "Design of consumer review systems and product pricing". In: *Information Systems Research* 26.4, pp. 714–730.
Karaman, Hülya (2021). "Online review solicitations reduce extremity bias in online review distributions and increase their representativeness". In: *Management Science* 67.7, pp. 4420–4445.
Lee, Nah, Bryan Bollinger, and Richard Staelin (2023). "Vertical versus horizontal variance in online reviews and their impact on demand". In: *Journal of Marketing Research* 60.1, pp. 130–154.
Lee, Young-Jin, Kartik Hosanagar, and Yong Tan (2015). "Do I follow my friends or the crowd? Information cascades in online movie ratings". In: *Management Science* 61.9, pp. 2241–2258.
Lembregts, Christophe, Jeroen Schepers, and Arne De Keyser (2024). "Is it as bad as it looks? Judgments of quantitative scores depend on their presentation format". In: *Journal of Marketing Research* 61.5, pp. 937–954.
Li, Linyi, Shyam Gopinath, and Stephen J Carson (2022). "History matters: the impact of online customer reviews across product generations". In: *Management Science* 68.5, pp. 3878–3903.
Li, Xinxin and Lorin M Hitt (2008). "Self-selection and information role of online product reviews". In: *Information systems research* 19.4, pp. 456–474.

Lin, Zhijie, Ying Zhang, and Yong Tan (2019). “An empirical study of free product sampling and rating bias”. In: *Information Systems Research* 30.1, pp. 260–275.

Liu, Yong (2006). “Word of mouth for movies: Its dynamics and impact on box office revenue”. In: *Journal of marketing* 70.3, pp. 74–89.

Moe, Wendy W and David A Schweidel (2012). “Online product opinions: Incidence, evaluation, and evolution”. In: *Marketing Science* 31.3, pp. 372–386.

Moe, Wendy W and Michael Trusov (2011). “The value of social dynamics in online product ratings forums”. In: *Journal of marketing research* 48.3, pp. 444–456.

Muchnik, Lev, Sinan Aral, and Sean J Taylor (2013). “Social influence bias: A randomized experiment”. In: *Science* 341.6146, pp. 647–651.

Naumzik, Christof, Stefan Feuerriegel, and Markus Weinmann (2022). “I will survive: Predicting business failures from customer ratings”. In: *Marketing Science* 41.1, pp. 188–207.

Office for National Statistics, UK (2021). Accessed on June 10, 2021. URL: https://www.ons.gov.uk/businessindustryandtrade/retailindu%20stry/timeseries/j4mc/drsi.

Pavlou, Paul A and Angelika Dimoka (2006). “The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation”. In: *Information systems research* 17.4, pp. 392–414.

Proserpio, Davide and Georgios Zervas (2017). “Online reputation management: Estimating the impact of management responses on consumer reviews”. In: *Marketing Science* 36.5, pp. 645–665.

Pu, Jingchuan et al. (2020). “Does identity disclosure help or hurt user content generation? Social presence, inhibition, and displacement effects”. In: *Information Systems Research* 31.2, pp. 297–322.

Qualtrics (2021). Accessed on June 10, 2021. URL: https://www.qualtrics.com/blog/online-review-stats.

Schlosser, Ann E (2005). “Posting versus lurking: Communicating in a multiple audience context”. In: *Journal of Consumer Research* 32.2, pp. 260–265.

Schneider, Christoph et al. (2021). “When the stars shine too bright: The influence of multidimensional ratings on online consumer ratings”. In: *Management Science* 67.6, pp. 3871–3898.

Timoshenko, Artem and John R Hauser (2019). “Identifying customer needs from user-generated content”. In: *Marketing Science* 38.1, pp. 1–20.

Tunc, Murat M, Huseyin Cavusoglu, and Srinivasan Raghunathan (2017). “Single-dimensional versus multi-dimensional product ratings in online marketplaces for experience goods”. In.

Wang, Chong, Xiaoquan Zhang, and Il-Horn Hann (2018). “Socially nudged: A quasi-experimental study of friends’ social influence in online product ratings”. In: *Information Systems Research* 29.3, pp. 641–655.

Wu, Fang and Bernardo A Huberman (2008). “How public opinion forms”. In: *International Workshop on Internet and Network Economics*. Springer, pp. 334–341.

Zheng, Xin et al. (2018). “Understanding the Impact of Multi-Dimensional Ratings on Product Sales: An Information-Inconsistency Perspective”. In: *Available at SSRN 3148033*.

Chapter 3

Strategic Rating Behavior

3.1 Introduction

Product ratings and reviews play significant roles in influencing consumers' purchasing decisions on online shopping platforms. The escalating proportion of internet-based sales compared to traditional retail (rising from 17% in 2018 to 34% in 2021, as per Office for National Statistics, 2021) underscores the growing importance of virtual word-of-mouth. Online product ratings are known to help brands build credibility and spread the word about the businesses (Forbes, 2022). Reflecting these shifts, numerous studies have thoroughly investigated the correlation between online word-of-mouth and sales (Chevalier and Mayzlin, 2006, Yong Liu, 2006, Dellarocas, Zhang, and Awad, 2007, Duan, Gu, and Whinston, 2008, Chintagunta, Gopinath, and Venkataraman, 2010, Clemons, Gao, and Hitt, 2006, Y. Chen, Q. Wang, and Xie, 2011, Moe and Trusov, 2011, L. Li, Gopinath, and Carson, 2022, Cho, Sosa, and Hasija, 2022, N. Lee, Bollinger, and Staelin, 2023), consistently revealing robust connections between online ratings, reviews, and sales.

Scholars have also explored the factors contributing to the generation of ratings and reviews. These factors encompass a variety of influences on the sentiment expressed in ratings and reviews, including social influence (i.e. impact of others' ratings and reviews) (Moe and Trusov, 2011, Muchnik, Aral, and Taylor, 2013, Y.-J. Lee, Hosanagar, and Tan, 2015, C. Wang, Zhang, and Hann, 2018, Burtch et al., 2018, Bhukya and Paul, 2023), visibility to others (Schlosser, 2005, Goes, Lin, and Au Yeung, 2014, Huang, Hong, and Burtch, 2017), adoption timing (F. Wu and Huberman, 2008, X. Li and Hitt, 2008), product popularity (Hu and X. Li, 2011), review context (Moe and Schweidel, 2012), disconfirmation effects (Ho, J. Wu, and Tan, 2017), and the design of the review system itself (P.-Y. Chen, Hong, and Ying Liu, 2018, Schneider et al., 2021, Lembregts, Schepers, and Keyser, 2024, Çakal, 2024). In this paper, we supplement and extend the existing research on the effects of social influence on online ratings and reviews by examining a new dimension of influence, namely **strategic rating**.

3.1.1 Positioning Strategic Rating Behavior

While prior research has explored how users react to others' ratings, we investigate whether users intentionally adjust their ratings in response to the number of prior ratings a product has received. Specifically, we hypothesize that users may engage in strategic rating by assigning lower ratings to products that have already accumulated a high number of ratings, particularly when they perceive the existing average rating as inflated. This behavior suggests that users attempt to influence the overall rating to better align with their personal evaluation of the product.

To formalize this idea, we introduce a simple model of strategic rating behavior. Users aim to influence the average rating, defined as:

$$\bar{r} = \frac{r_1 + r_2 + \ldots + r_n}{n} \tag{3.1}$$

where $\bar{r}$ represents the average rating and n is the number of ratings. If a user's true evaluation, r_t, differs significantly from $\bar{r}$, they may submit a manipulated rating, r_m, instead of r_t. The extent of this distortion is given by:

$$|r_t - r_m| \tag{3.2}$$

Observationally, as n increases, the distortion also increases, meaning users are more likely to provide exaggerated ratings to counterbalance the existing average rating. This leads to two key empirical predictions:

- **Prediction 1**: Users strategically adjust their ratings when faced with a larger number of prior ratings.
- **Prediction 2**: Users strategically adjust their ratings when the average rating deviates significantly from their own evaluation.

To test this hypothesis, we analyze user-rating data from Ratebeer.com, a well-known beer rating platform, focusing on users in California. Our dataset comprises ratings from the 340 most active users in California for the 500 most commonly reviewed beers. We track the number of prior ratings each beer had received at the time of a given rating and control for user-specific characteristics and other relevant variables. Using a fixed-effects regression model, we find a statistically significant negative relationship between the number of prior ratings and the ratings provided by users, supporting the presence of strategic rating behavior.

Further, we conduct a sentiment analysis of the textual reviews accompanying these ratings. Interestingly, we find no statistically significant relationship between the number of prior ratings and the sentiment expressed in the reviews. This divergence suggests that while users may manipulate numerical ratings to influence the aggregate score, their written reviews more accurately reflect their true opinions.

Our findings contribute to the literature by revealing a previously underexplored dimension of rating behavior: the strategic use of ratings to shape product evaluations. Understanding this behavior has important implications for the design of online rating systems, as platforms seeking to maintain fair and accurate product assessments must account for potential distortions introduced by strategic rating. By uncovering the motivations and mechanisms behind this behavior, our study provides valuable insights for platform designers and policymakers aiming to ensure more reliable and representative user-generated ratings.

3.2 Relevant Literature

3.2.1 Social Influence and Online Ratings

Studies examining how social influence impacts product evaluations revolve around two themes: first, how others' evaluations impact our evaluations, and second, how the knowledge of the visibility of our evaluations affect our evaluations. As one of the earliest studies in this domain, both Cohen and Golden, 1972 and Burnkrant and Cousineau, 1975 compare the nature of evaluations given by subjects upon social influence versus no influence. They find that subjects who were shown a corpus of uniform evaluations by others experienced more influence in their own evaluations than the subjects who were shown less uniform evaluations. Furthermore, subjects who expected their evaluations to be visible by others did not behave differently than those who expected visibility, keeping information uniformity constant.

Schlosser, 2005 shows that upon receiving a negative review, reviewers tend to rate the product less favorably, but their evaluations are not affected by neutral or positive reviews. In support of the impact of visibility on product evaluations, Schlosser, 2005 also finds that evaluators whose ratings will be made public provide more negative ratings than the others whose ratings will be kept private.

Huang, Hong, and Burtch, 2017 make use of a natural experiment to investigate the changes in user behavior upon changes in the audience size and composition. Specifically, they exploit the integration of Yelp and TripAdvisor with Facebook to investigate how user reviews change in volume and linguistic features in response to a change in the audience mixture. Their results reveal that social integration leads to an increase in review volume, but a decrease in review quality.

In a similar vein with Schlosser, 2005 and Huang, Hong, and Burtch, 2017, Goes, Lin, and Au Yeung, 2014 study how opinion writers respond to an increase in their audience size using data from Epinions. In this website, users post product reviews, and they can 'trust' each other via the web-of-trust feature. This feature allows users to 'subscribe' other users, and those trusted users' reviews are displayed in priority in the subscribers' page. They use observational data and show that as users become more popular (i.e. more users subscribe to them), they produce more reviews and more objective reviews.

Burtch et al., 2018 investigate the impacts of financial incentives and social norms on content generation. Through randomized experiments, they show that financial incentives lead to an increase in the number of reviews generated while social norms, manipulated by displaying how many peers provided a review, cause an increase in review length.

Hu and X. Li, 2011 use book reviews on Amazon and find that late reviews tend to contradict with early ones. They attribute this difference either to self-selection bias or to expectations from previous reviews (or both). In addition to self-selection, late readers' expectations from the books might be altered by the early readers' comments (who mostly hold higher preferences for the books), hence a disconfirmation effect might have occurred as the expectancy disconfirmation paradigm predicts. Furthermore, Hu and X. Li, 2011 investigate the moderating role of book popularity on rating behavior, and conclude that differentiation in later reviews is larger for unpopular books than popular books.

Moe and Schweidel, 2012 explore the determinants of whether to post a rating as well as the nature of these ratings depending on individual characteristics. Their results show that a higher-rating environment encourages posting ratings while a lower-rating environment discourages it. Additionally, they find that less frequent raters follow the prior ratings in their ratings (bandwagon effect) while frequent raters tend to differentiate themselves by posting different ratings than the prior ones.

Similar to the current study, Muchnik, Aral, and Taylor, 2013 study how previous ratings impact subsequent ratings. In a large-scale online randomized field experiment conducted on a social news aggregation website where users can comment under news and other users can up- or down-vote these comments, Muchnik, Aral, and Taylor, 2013 show that positive social influence accumulates while negative social influence is neutralized. Specifically, the researchers randomly assign an up-vote, a down-vote, or no vote randomly to newly made comments (which received no vote after formation) and examine the changes in the subsequent voting behavior. They first investigate the immediate responses to the manipulations, i.e. voting behavior of the first voter after manipulation. They find significant herding effect for up-voted comments, voters significantly up-voted the comments receiving up-vote treatment compared to control group (no vote). They also find that voters up-voted the comments that received a down-vote due to manipulation significantly more than the control comments, leading to a neutralization of the down-vote manipulation.

Y.-J. Lee, Hosanagar, and Tan, 2015 examine the varying influence of prior ratings from strangers versus friends on a social movie reviewing platform. They find that raters tend to align their ratings with those of their friends, demonstrating a herding effect. However, they also observe that herding behavior in response to crowd ratings diminishes as the number of friends rating the same movie increases.

C. Wang, Zhang, and Hann, 2018 exploit a natural experiment in an online rating website for books, movies and songs in China to investigate the impact of friends' ratings on subsequent ratings. Their results show that, upon forming friendship on the website, friends' ratings tend to become similar, a finding consistent with Y.-J. Lee, Hosanagar, and Tan, 2015's findings. Moreover, they show that social influence is stronger for users with relatively smaller friend networks and for more popular books.

Closest to the aim of the current study, F. Wu and Huberman, 2008 investigate how the current raters are affected by prior ratings by analyzing the temporal evolution of book ratings on Amazon with a focus on the difference between early and later raters. Although at the correlational level, one of their findings pertains to the differentiation behavior of late raters. They interpret this finding by the utility gained when one can influence the average rating.

Our study extends the literature on the impact of social influence on online product ratings by demonstrating that users might be affected by number of ratings the product they are rating gathered. This study differs from F. Wu and Huberman, 2008 in that we are not interested in if late raters provide different ratings than earlier raters but instead we investigate the causal impact of others' ratings on the focal users' ratings. Among the potential herding vs. strategically behaving outcomes, we find that raters act strategically and provide lower ratings to products that gathered more ratings.

3.2.2 Consumer Expertise and Product Evaluations

Literature on consumer expertise sheds light on the possible interactions of expertise and social influence. Studies in this domain revolve around the the impact of consumer expertise on product choice and information processing. In this study, we look at how consumer expertise affects the degree of social influence in product evaluations.

In their seminal paper, Alba and Hutchinson, 1987 posit that experienced individuals are better at evaluating information because they process it at a deeper level and distinguish between relevant and irrelevant information. Given that, it is intriguing to investigate whether there is a difference in the extent to which experienced versus inexperienced consumers are affected by others' ratings when rating a product.

Interestingly, Bettman and C. W. Park, 1980 show that moderately experienced consumers appear to do more processing of the currently available information and rely on prior knowledge to a lesser extent than the more and less experienced consumers. Similarly, C. W. Park and Lessig, 1981 find that, in reducing the alternatives to a smaller set of acceptable options, the decision time of a decision-maker at a moderate level of familiarity is greater than that of the decision-makers at either low or high levels of familiarity.

In his adoption of human capital model to derive insights into implications of consumer knowledge, Ratchford, 2001 derives that due to learning by doing, repeated activities become less expensive to produce which implies that frequent raters are less likely to be affected by others' ratings than less frequent raters.

Analyzing the impact of consumer expertise on the influenceability from online reviews, D.-H. Park and S. Kim, 2008 show that the number of online consumer reviews has a stronger effect on the purchase intention of consumers with low expertise than on consumers with high expertise. Similarly, Ketelaar et al., 2015 show that expertise moderates the effect of online review valence: negative online reviews lead to weaker purchase intentions among novice consumers than expert ones, and positive online reviews lead to stronger purchase intentions among novice consumers than expert ones.

In a similar vein, Aqueveque, 2015 shows that for a low-priced wine, the effect of positive word-of-mouth on perceived quality is greater for expert consumers than for novice consumers, whereas for a high-priced wine this increase of perceived quality is observed only for novice consumers.

Ding and S. Li, 2019 study book purchase decisions and conclude that for more 'established users', the use of personal knowledge dominates the information from other users' actions in solving the quality uncertainty issue, namely choosing a book. Hence, more experienced readers rely more on their own expertise in choosing a book than others' opinions compared to less experienced readers.

Closest to the current study, Sunder, K. H. Kim, and Yorkston, 2019 study the relationship between consumer experience and influenceability from others' evaluations. Using data from an online gaming website, they show that as an individual's prior rating experience increases, the crowd's influence on subsequent ratings decreases. However, they also find that as an individual's prior rating experience increases, the friends' influence on subsequent ratings increases.

By studying the moderating effect of consumer experience in a product category on the impact of social influence on product ratings, we implement the literature on consumer expertise. Our results show that although, consistent with prior findings, users provide lower ratings to products as their experience in the product category increases, the relationship between others' evaluations and the focal user's evaluation does not appear to be moderated by consumer experience. We discuss the potential reasons and implications of this finding in the Discussion section.

3.3 Empirical Setting

3.3.1 Beer Rating Website

We collected data from a beer rating website on which users open a user account to rate beers. Almost all beer brands and models exist on the website. There is a page for each beer, and once one lands on a beer's page, she sees the name of the beer, the brand the beer belongs, the average rating of the beer, number of ratings the beer gathered, number of written reviews the beer gathered, information about the beer (qualities, country it belongs, etc.) and the latest reviews the beer received with the user's nickname who rated the beer, when she rated the beer, what rating she gave to the beer and her written review for the beer (see Figure 3.1 for an example beer page).

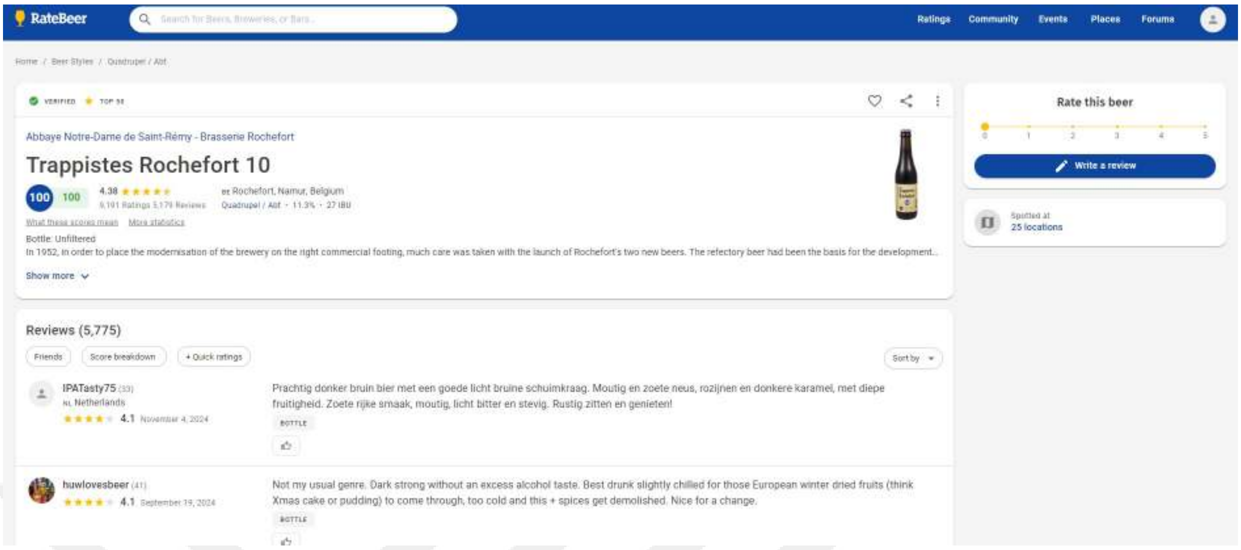


FIGURE 3.1: Page of a Beer on the Beer Rating Website

We also can access the users' pages where all the beers they rated are listed. On a user page, we can access the user's nickname, where she lives, how many beers she rated, and the list of beers she rated that shows the name of the beer rated, rating provided to that beer, and the date at when the beer is rated (see Figure 3.2 for an example).

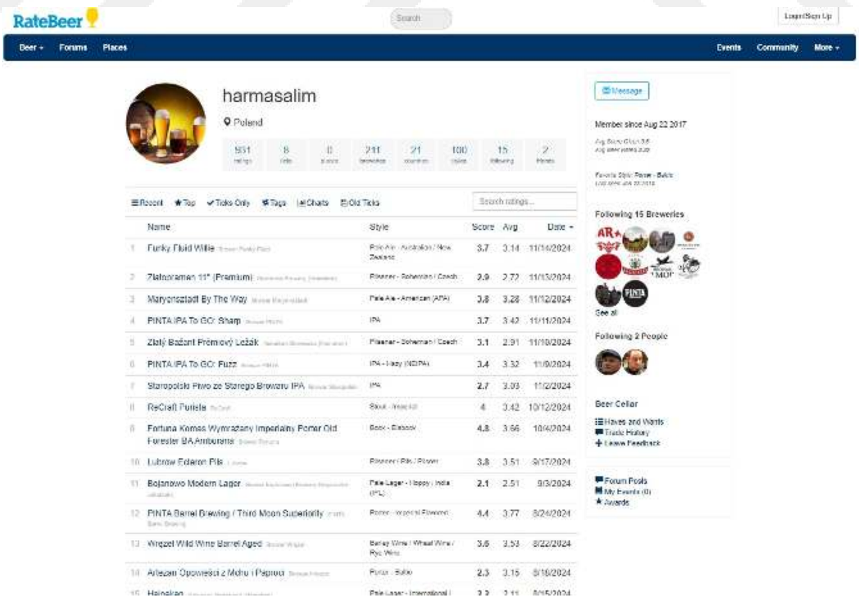


FIGURE 3.2: Page of a User on the Beer Rating Website

To rate a beer, a user needs to land on the beer's page, and click on the 'Write a review' button that is located on the top-right of the page (Figure 3.1). Upon clicking on the button, the user sees that a box opens on which review and rating areas are displayed. For rating, the user provides ratings for the 'Aroma', 'Appearance', 'Taste', 'Mouthfeel' qualities as well as an 'Overall' rating as can be seen on Figure 3.3. The average of these ratings are calculated immediately and displayed on the bottom of the box. If the user changes her rating for one of these attributes, the average rating also changes immediately.

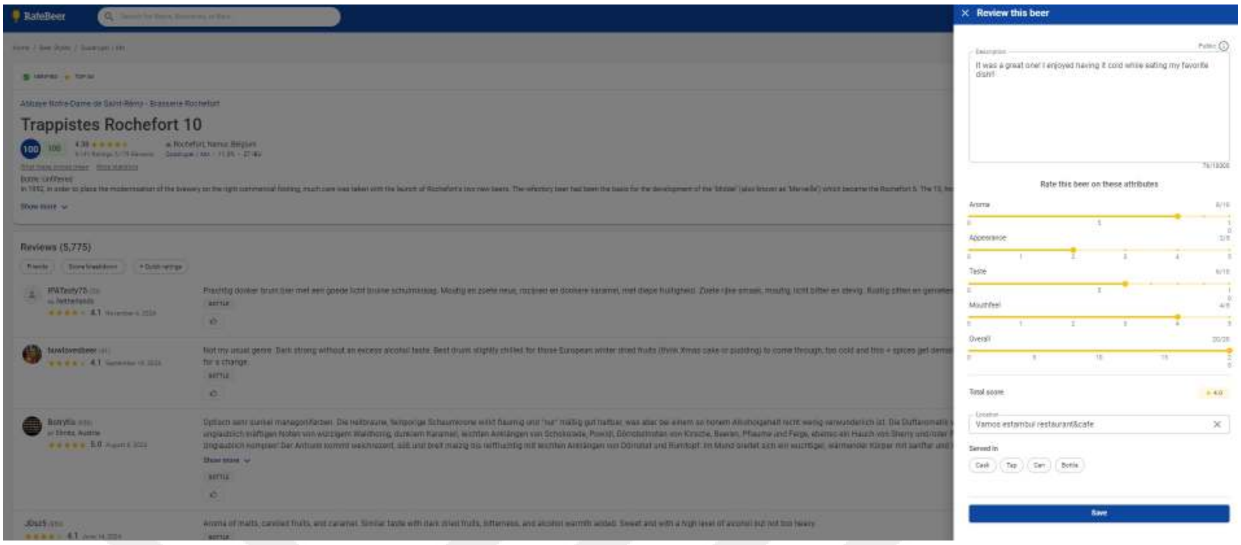


FIGURE 3.3: Rating a Beer on the Beer Rating Website

The beer and user rating information is ideal to study the causal impact of number of ratings a beer gathered on the rating the user provides to that beer, because we can calculate the number of ratings a beer gathered before the focal user rates that beer by counting the number of ratings posted for that beer before the focal user rated it. In addition, we can calculate the average rating of the beer while the focal user were rating it by taking the average of the ratings that were given before the focal user. This is important because it is likely that the focal user is affected by the average rating of the beer while she is rating that beer as extant research put forth (Muchnik, Aral, and Taylor, 2013, Y.-J. Lee, Hosanagar, and Tan, 2015, C. Wang, Zhang, and Hann, 2018) and as we hypothesized.

Additionally, we can compute the consumer experience by counting the number of beers the user rated before rating the focal beer. This is important since the influenceability of the user from the number of ratings gathered by the beer may differ for experienced versus inexperienced consumers (Ding and S. Li, 2019, Sunder, K. H. Kim, and Yorkston, 2019).

Finally, the average ratings of the beers at the time of data collection serves as a proxy for beer quality. Because we collected the rating data for the most-rated beers among Californian users, at the time of data collection, these beers had accumulated many ratings, allowing us to use their averages as proxies for product quality in our final analyses.

3.3.2 Descriptive Statistics

We collected rating history data for the most-rating 340 Californian users. We limited the users to California to control for geographic variations in product availability and distribution. The dataset comprises ratings from the most active 340 users in California, who collectively rated the 500 most commonly reviewed beers. These 500 beers belong to 138 different beer brands.

The key statistics of the data are displayed in Table 3.1. The average of all ratings provided by the users in our dataset to the beers they rated is 3.67, representing a lower average rating than can be found in other rating/e-commerce websites. 4.00

is the rating that is given most out of a 0-5 rating scale. The ratings display a good variation with a standard deviation of 0.66, a minimum of 0.50 and a maximum of 5.00.

The users in the dataset rated an average number of 382 beers, and number of beers rated most frequently is 22. The number of beers rated by users in the dataset display a good variation with a standard deviation of 461.8, a minimum of 1 and a maximum of 3712.

	Average	*Mode*	*Median*	*Std Deviation*	*Minimum*	*Maximum*
Beer Ratings	3.67	4.00	3.88	0.66	0.5	5.00
Experience	382.67	22	204	461.8	1	3712

TABLE 3.1: Descriptive Statistics of Beer Ratings

Figure 3.4 displays the relationship between the number of ratings gathered by a beer before the focal user rates that beer and the rating provided by the focal user. As a model-free evidence, we see that as number of ratings the beer gathered increases, users provide lower ratings. To put it roughly, a 1000 increase in the number of ratings is associated with a 0.50 decrease in the rating provided to that beer.

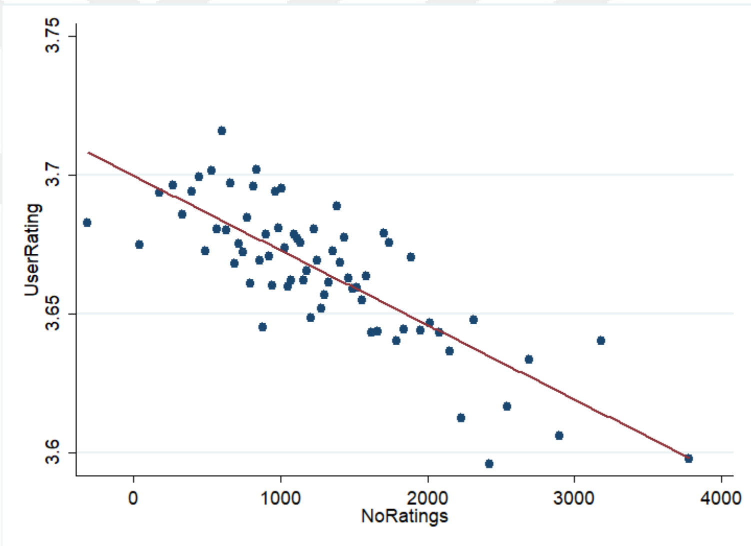


FIGURE 3.4: Relationship Between Number of Ratings a Beer Gathered Before the Focal User Rates that Beer and the Rating Provided by the Focal User

3.4 Empirical Analysis

We investigate the relationship between the number of ratings gathered by a beer and the rating provided by the user for that beer. Our base model is:

$$UserRating_{ij} = \beta_0 + \beta_1 NoRatings_{ij} + \beta_2 ExposedAverage_{ij} + \beta_3 FinalAverage_j + \alpha_i + \epsilon_{ij} \tag{3.3}$$

where where i and j index the user and the beer, respectively. $UserRating_{ij}$ is the dependent variable denoting the rating provided by user i for beer j.

$NoRatings_{ij}$ is the main variable of interest and it denotes the number of ratings the beer j gathered before the user i rates that beer. $ExposedAverage_{ij}$ denotes the average rating of the beer j while user i rates that beer.

$FinalAverage_j$ denotes the average rating of the beer j at the time of data collection. Because the beers in our dataset are the ones that are most commonly rated by the users, they gathered many ratings at the time of data collection, hence we use the average ratings of them as a proxy for the beer quality.

Finally, α_i is the user-fixed effects. We include user-fixed effects to control for potential variations in rating behavior of users that are not dependent on time.

In our further studies, we also include user experience, $Experience_{ij}$, which denotes the experience gained by user i until she rates beer j. Following Mitchell and Dacin, 1996 and Bondi, Rossi, and Stevens, 2024, we operationalize user experience as the number of beers rated before rating the focal beer.

Including user experience is beneficial for two reasons: first, experience changes over time, hence it is not captured by user-fixed effects. Second, the relationship between experience and the rating provided to a beer, keeping else constant, is interesting itself. Additionally, it is intriguing to see if the relationship between the number of ratings and the rating provided is moderated by user experience.

In a separate group of analyses, we employ the same models but change the dependent variable to either number of words used in the written review, $NoWords_{ij}$, or sentiment of the written review, $Sentiment_{ij}$. Looking at the relationship between the number of ratings a beer gathered and the number of words/sentiment used in the review written for that beer and comparing the results with the results from the main analysis would allow us to see if the users react differently to observing the number of ratings provided to a beer in their ratings and written reviews. We calculate the sentiment of written reviews by using the vaderSentimentScores package of MATLAB, where sentiment score varies between -1 and 1.

Next, we present the results from our base model as well as other models in which we include user experience and interactions between various variables and discuss the implications of these findings.

3.5 Results

3.5.1 Impact of Exposed Number of Ratings on User Ratings

The results from the main and extended models examining the impact of the number of ratings a beer received on the rating provided by the user are presented in Table 3.2. The first column presents our estimate of the base model (Equation 3.3). On average, ratings decrease by 0.027 points as number of ratings the product gathered increases by 1000. The negative sign of the change in ratings is in accordance with the model-free evidence presented in Figure 3.4.

In Columns 2-4, we include additional control variables that may influence rating behavior. In Column 2, we include $Experience$, in Column 3 the interaction of $NoRatings$ and $FinalAverage$, and in Column 4, the interaction of $NoRatings$ and $Experience$. Across all specifications, the estimate of the focal variable, $NoRatings$, is statistically significant, and the magnitude varies between 0.027 and 0.1 per 1000 ratings, indicating a negative relationship between number of ratings gathered by the beer and the rating provided to that beer.

The coefficient of $ExposedAverage$ is positive and statistically significant in all regressions, implying a positive social influence effect. In general, users provide similar ratings to the average ratings of the beers they are rating, a finding consistent

VARIABLES	(1) UserRating	(2) UserRating	(3) UserRating	(4) UserRating
NoRatings	-2.68e-05***	-3.53e-05***	-0.000106***	-0.000104***
	(4.44e-06)	(4.15e-06)	(3.57e-05)	(3.56e-05)
ExposedAverage	0.712***	0.696***	0.640***	0.638***
	(0.0790)	(0.0780)	(0.0709)	(0.0711)
FinalAverage	0.399***	0.423***	0.454***	0.455***
	(0.0826)	(0.0818)	(0.0762)	(0.0763)
Experience		-5.93e-05***	-5.95e-05***	-4.49e-05***
		(1.37e-05)	(1.37e-05)	(1.52e-05)
NoRatings * FinalAverage			1.84e-05*	1.91e-05**
			(9.49e-06)	(9.59e-06)
NoRatings * Experience				-1.58e-08
				(1.04e-08)
Constant	-0.375***	-0.372***	-0.274***	-0.276***
	(0.0508)	(0.0508)	(0.0811)	(0.0811)
Observations	48,523	48,523	48,523	48,523
R-squared	0.614	0.615	0.615	0.615
User Fixed-Effects	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 3.2: Relationship Between the Number of Ratings and the Rating Provided

with the findings of previous studies (Muchnik, Aral, and Taylor, 2013, Y.-J. Lee, Hosanagar, and Tan, 2015, C. Wang, Zhang, and Hann, 2018).

FinalAverage, our proxy for beer quality, also appears to be positively related to the ratings provided by the users across all specifications. However, it is interesting to note that the magnitude of the relationship between *FinalAverage* and *NoRatings* is smaller than the magnitude of the relationship between *ExposedAverage* and *NoRatings* in all regressions. This observation implies that, while rating a product, the product's current rating is a more dominant determinant of the user's rating than its quality.

The interaction between *NoRatings* and *FinalAverage* appears to be statistically significant (at 10% in second regression and at 5% in third regression) and slightly positive. According to this result, users provide higher ratings to beers that are of higher quality, as the number of ratings that beer gathered increases. In contrast to the current average rating of a beer, the final average rating of a beer appears to induce higher ratings as the beer gathers more ratings (see also Section 3.5.2).

Finally, the coefficient of *Experience* is negative and statistically significant in all regressions, a finding consistent with extant research that suggests that more experienced consumers tend to elaborate more (Alba and Hutchinson, 1987). Hence, as can be expected, more experienced users provide lower ratings to beers, keeping else constant. However, more interestingly, experience does not appear to moderate the relationship between *UserRating* and *NoRatings*. This result implies that users' tendency to provide lower ratings to beers that gathered more ratings does not change with experience.

3.5.2 Impact of Exposed Average Ratings on User Ratings

VARIABLES	(1) UserRating
NoRatings	-0.000276***
	(4.86e-05)
FinalAverage	1.057***
	(0.0302)
Experience	-4.51e-05***
	(1.55e-05)
NoRatings * FinalAverage	6.55e-05***
	(1.43e-05)
NoRatings * Experience	-1.63e-08
	(1.07e-08)
2. ExposedAverage	-0.00801
	(0.0133)
3. ExposedAverage	-0.00421
	(0.0191)
2. NoRatings * ExposedAverage	-7.48e-06
	(9.26e-06)
3. NoRatings * ExposedAverage	-2.97e-05**
	(1.20e-05)
Constant	-0.150
	(0.105)
Observations	48,523
R-squared	0.614
User Fixed-Effects	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.3: *ExposedAverage* as a Categorical Variable

Table 3.3 presents a more nuanced analysis of the relationship between the interaction of *NoRatings* and *ExposedAverage*, and the rating provided by the user. Specifically, we grouped *ExposedAverage* into three such that the first group consists of the lowest ratings, the second group consists of ratings that are higher than the ratings in the first group and the third group consists of the highest ratings, and the categorical *ExposedAverage* variable is created accordingly.

In this regression, we add the the second and third categories of the categorical *ExposedAverage* variable as well as the interactions of these categories with *NoRatings* as regressors. Results reveal that the statistical and economic significance of *NoRatings* is preserved, indicating a robust relationship between the number of ratings a beer gathered and the rating provided to that beer. Similarly, the coefficient of *Experience* appears to be negative and significant, as was found previously.

Keeping else constant, the coefficients of the second and third groups of *ExposedAverage* appear not significant, meaning that users do not assign different ratings to beers that have lower or higher average ratings. However, the interaction of the third group of *ExposedAverage* (higher average ratings group) with *NoRatings* is negative and statistically significant at 1%. This finding implies that, keeping the number of ratings a beer gathered constant, as the average rating of a beer increases, users give lower ratings to the beers.

Taken together with the negative coefficient of *NoRatings*, this analysis shows that, keeping else constant, users give lower ratings to beers that gathered more and higher ratings. In the next section, we examine whether users' written reviews are related to the number of ratings a beer has gathered, aiming to shed more light on the mechanism behind the observed effect.

3.5.3 Mechanism Check: Impact of Exposed Number of Ratings on Review Length and Sentiment

VARIABLES	(1) NoWords	(2) NoWords	(3) Sentiment	(4) Sentiment
NoRatings	0.00310	0.00256	-1.35e-05	-1.54e-05
	(0.00182)	(0.00181)	(2.76e-05)	(2.74e-05)
ExposedAverage	16.63***	17.15***	0.239***	0.241***
	(4.374)	(4.386)	(0.0450)	(0.0449)
FinalAverage	-4.542	-4.816	-0.0418	-0.0428
	(4.372)	(4.399)	(0.0417)	(0.0416)
Experience	0.00707**	0.00242	-7.85e-06	-2.45e-05**
	(0.00280)	(0.00253)	(1.34e-05)	(1.22e-05)
NoRatings * FinalAverage	-0.000559 (0.000464)	-0.000798* (0.000466)	5.18e-06 (7.06e-06)	4.33e-06 (7.11e-06)
NoRatings * Experience		5.01e-06*** (1.51e-06)		1.79e-08** (7.04e-09)
Constant	8.843**	9.451**	-0.141***	-0.139***
	(4.427)	(4.431)	(0.0509)	(0.0508)
Observations	48,523	48,523	48,523	48,523
R-squared	0.510	0.512	0.150	0.150
User Fixed-Effects	Yes	Yes	Yes	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.4: Number of Words in a Review and the Sentiment of a Review as Dependent Variables

So far, we have established that users post lower ratings to beers with more and higher ratings. This finding is consistent with both differentiation of users from others (Moe and Schweidel, 2012) and strategic rating behavior. To further disentangle these effects, we also examine how users respond to prior number of ratings in their written reviews.

Cheema and Kaikati, 2010 show that reviewers aiming to stand out tend to carry lower sentiment toward products and are less likely to recommend them. In a series of lab experiments, they show that high need for uniqueness decreases consumers' willingness to provide positive word-of-mouth. Furthermore, consumers with higher need for uniqueness discuss product details more. Hence, if differentiation is at play, we might expect users to write longer reviews with lower sentiment, deliberately diverging from the majority. Conversely, if strategic rating behavior is at play, we may expect users to focus their strategic intentions on numeric ratings rather than reflecting them in the sentiment of their written reviews.

Table 3.4 presents results from the regressions where the dependent variable *UserRating* of the main model is replaced with either number of words used in written reviews *NoWords* or the sentiment of the review *Sentiment*. In these regressions, we analyze if number of ratings a beer gathered affects the number of words used in written reviews as well as the sentiment of the reviews.

Curiously, in all regressions, the coefficient of *NoRatings* does not appear to be statistically significant, in contrast to the case where user rating is the dependent variable. Hence, we can conclude that, keeping else constant, number of ratings a beer gathered is not related to the number of words used in a review or the sentiment of the review.

In Columns 1 and 2, we see that the coefficient of *ExposedAverage* is positive and statistically significant. On average, users use 16-17 more words in their reviews for beers that garnered higher average ratings. Similarly, as can be seen in Columns 3 and 4, users provide more positive reviews to the same beers.

As can be expected, users appear to provide more lengthy and less positive reviews to beers as their experience increases. In only the third regression, the coefficient of *Experience* is not statistically significant.

Taken together, these results suggest that users' written reviews do not systematically diverge in sentiment or length in response to the number or average rating of prior reviews. This finding supports the hypothesis that the observed decrease in ratings for beers with more and higher ratings is driven by strategic rating behavior rather than differentiation. Specifically, users appear to focus their strategic intentions on numeric ratings, while their written reviews remain unaffected in sentiment or length, underscoring the deliberate nature of their rating adjustments.

3.5.4 Further Mechanism Check: Users' Reaction to Variation in Ratings

Our analyses show that users provide lower ratings to beers that gathered more and higher ratings (see Tables 3.2 and 3.3). This finding is consistent with our initial hypothesis that users may opt to rate lower these beers in a strategic way in an attempt to affect their ratings. Our finding that the number of ratings a beer gathered does not affect the length or the sentiment of the reviews (see Table 3.4) is also consistent with strategic rating hypothesis since it is cognitively more difficult to change the content of a written review than a rating and the ratings are more salient than reviews. Here, we provide additional evidence that the lower ratings to products with more and higher ratings is not due to differentiation but instead due to strategic rating behavior.

Extant research indicates that consumers with a high need for uniqueness tend to differentiate themselves more as unanimity among other consumers increases (Tepper, Bearden, and Hunter, 2001, Tepper, 1997). In our context, if differentiation is driving the impact of others' ratings on user rating, we would expect users to respond to variation in prior ratings. Specifically, as the variation in prior ratings decreases, users would post lower ratings. In contrast, if strategic rating is at play, users would respond only to the number of ratings and the average rating rather than to the variation in prior ratings.

To test whether users respond to variation in prior ratings, we calculated the standard deviation of ratings for a beer prior to the focal user's rating, referred to as *ExposedVariation*. We then estimated our base model (Equation 3.3) with *ExposedVariation* included as an additional regressor. Table 3.5 presents the results of this estimation. In contrast to the prediction of differentiation hypothesis, we find that the coefficient of *ExposedVariation* is not statistically significant at desired

VARIABLES	(1) UserRating
NoRatings	-2.85e-05***
	(4.42e-06)
ExposedAverage	0.715***
	(0.0802)
FinalAverage	0.397***
	(0.0838)
ExposedVariation	-0.125
	(0.130)
Constant	-0.368***
	(0.0519)
Observations	48,523
R-squared	0.614
User Fixed-Effects	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.5: Relationship Between Variation in Prior Ratings and User Rating

levels. This finding lends further support to the hypothesis that strategic rating behavior is driving the influence of prior ratings on user ratings.

3.6 Discussion

3.6.1 Theoretical Contributions

The drivers of online ratings that are identified in extant research can be classified as self-selection, social influence and the rating system. By tapping into a different dimension of influenceability from others' evaluations, current study extends the set of drivers of the online ratings. Specifically, we present evidence that others' evaluations not only leads to herding or differentiation effects, but may also trigger strategic rating behavior.

Extant research put forth that users may herd after others' ratings (Y.-J. Lee, Hosanagar, and Tan, 2015, C. Wang, Zhang, and Hann, 2018) or differentiate their ratings from others' (Moe and Schweidel, 2012). In this study, we show that other than these effects, users may act strategically and give lower ratings to products that gathered more and higher ratings.

The results of our study confirms the findings of previous studies in regard to the impact of consumer experience on product evaluations. In line with extant research, we find that more experienced users provide lower ratings, on average. However, interestingly, we do not find that the strategic rating is moderated by experience. Specifically, regardless of their experience, users provide lower ratings to products that garnered more and higher ratings.

Our analysis of the impact of others' ratings on review length and review sentiment shows intriguing results. In contrast to ratings, users do not adjust either the length or the sentiment of their reviews upon viewing the number of ratings a product gathered. The strategic behaving only occurs in ratings. Yet, again in line with extant research, we find that users provide more lengthy and more negative reviews as their experience increases.

All in all, our study explores and finds evidence in favor of strategic rating behavior in online settings. By documenting strategic rating behavior, current study extends and implements the literature on the drivers of online ratings.

3.6.2 Managerial Implications

Online ratings and reviews have been shown to significantly influence sales (Chevalier and Mayzlin, 2006, Clemons, Gao, and Hitt, 2006, Chintagunta, Gopinath, and Venkataraman, 2010). As consumers increasingly rely on online reviews to gather product information, their importance continues to grow. This underscores the need for rating systems that encourage truthful and unbiased evaluations of products.

This study provides evidence that users tend to assign lower ratings to products with a higher number of positive prior ratings, offering valuable insights for rating system design. Since users' ratings are influenced by the number and valence of previous ratings, system designers might consider reducing the visibility or salience of this information during the rating process. Concealing such information can encourage users to provide more independent assessments, minimizing strategic distortions. If concealing rating information is not feasible, platforms could implement algorithms to adjust the calculation of average ratings, giving greater weight to earlier ratings and less weight to later ones. This approach could mitigate the downward bias introduced by subsequent strategically lower ratings.

Additionally, platforms should incorporate mechanisms to identify and address strategic distortions in user ratings. Algorithms that detect anomalous rating patterns or weight ratings based on the number of prior ratings could further enhance system robustness. Transparency in how ratings are aggregated and displayed is also critical to building user trust. Providing contextual information, such as the distribution of ratings or trends over time, can empower users to make informed decisions while reducing the likelihood of strategic behavior.

Strategic rating behavior can also serve as a valuable signal for marketing purposes. Patterns of lower ratings may indicate product misalignment with user expectations, providing actionable insights for improving product offerings or positioning. Platforms and businesses can leverage these insights to align their offerings more closely with consumer preferences, ultimately improving customer satisfaction and sales performance.

Importantly, while numerical ratings are susceptible to strategic distortions, this study finds that written reviews are not significantly affected by the number or valence of prior ratings. As such, written reviews remain a reliable source of consumer feedback. Platforms should highlight these reviews to complement numerical ratings, leveraging their authenticity to enhance consumer trust and decision-making.

Finally, improving platform fairness is essential for maintaining the integrity of rating systems. Policies discouraging manipulative behavior, combined with user education about the value of unbiased evaluations, can foster a culture of honest feedback. Educating users on the broader benefits of accurate ratings for the community can further encourage fairness and trust in the system. By addressing these considerations, platforms can create rating systems that are more robust, trustworthy, and fair, ultimately benefiting both consumers and businesses.

3.7 Conclusion

In conclusion, our study provides compelling evidence of strategic rating behavior in online environments, revealing that users tend to give lower ratings to products that have accumulated a higher number of prior ratings. This behavior is distinct from other forms of social influence and suggests deliberate efforts by users to influence overall ratings. Interestingly, while numerical ratings are subject to this strategic behavior, the sentiment and length of written reviews remain unaffected, indicating that reviews may offer a more authentic reflection of user experiences.

Our findings highlight the complex dynamics of user behavior in online rating systems, where social influence not only drives conformity or differentiation but also inspires strategic actions. This nuanced understanding adds a new dimension to the existing literature on the drivers of online ratings and underscores the importance of accounting for strategic behaviors when analyzing user-generated content.

For e-commerce platforms, these results emphasize the need to carefully design rating systems that minimize strategic distortions. Approaches such as anonymizing cumulative ratings or introducing mechanisms to counteract biases could help ensure that aggregated ratings more accurately reflect product quality. At the same time, our findings underscore the value of written reviews as an alternative source of genuine consumer feedback, offering a richer and less strategically manipulated perspective on user experiences.

By documenting the prevalence of strategic rating behavior, this study advances the understanding of online consumer behavior and provides actionable insights for platforms seeking to improve the reliability and integrity of their rating systems.

References

Alba, Joseph W and J Wesley Hutchinson (1987). “Dimensions of consumer expertise”. In: *Journal of consumer research* 13.4, pp. 411–454.

Aqueveque, Claudio (2015). “The influence of experts’ positive word-of-mouth on a wine’s perceived quality and value: the moderator role of consumers’ expertise”. In: *Journal of Wine Research* 26.3, pp. 181–191.

Bettman, James R and C Whan Park (1980). “Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis”. In: *Journal of consumer research* 7.3, pp. 234–248.

Bhukya, Ramulu and Justin Paul (2023). “Social influence research in consumer behavior: What we learned and what we need to learn?–A hybrid systematic literature review”. In: *Journal of Business Research* 162, p. 113870.

Bondi, Tommaso, Michelangelo Rossi, and Ryan L Stevens (2024). “The Good, the Bad and the Picky: Consumer Heterogeneity and the Reversal of Product Ratings”. In: *Management Science*.

Burnkrant, Robert E and Alain Cousineau (1975). “Informational and normative social influence in buyer behavior”. In: *Journal of Consumer research* 2.3, pp. 206–215.

Burtch, Gordon et al. (2018). “Stimulating online reviews by combining financial incentives and social norms”. In: *Management Science* 64.5, pp. 2065–2082.

Çakal, Ali (June 2024). “How Do Multi-Dimensional Rating Systems Affect Ratings and Reviews? An Empirical Investigation”. PhD thesis. Koç University.

Cheema, Amar and Andrew M Kaikati (2010). “The effect of need for uniqueness on word of mouth”. In: *Journal of Marketing research* 47.3, pp. 553–563.

Chen, Pei-Yu, Yili Hong, and Ying Liu (2018). "The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments". In: *Management Science* 64.10, pp. 4629–4647.
Chen, Yubo, Qi Wang, and Jinhong Xie (2011). "Online social interactions: A natural experiment on word of mouth versus observational learning". In: *Journal of marketing research* 48.2, pp. 238–254.
Chevalier, Judith A and Dina Mayzlin (2006). "The effect of word of mouth on sales: Online book reviews". In: *Journal of marketing research* 43.3, pp. 345–354.
Chintagunta, Pradeep K, Shyam Gopinath, and Sriram Venkataraman (2010). "The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets". In: *Marketing science* 29.5, pp. 944–957.
Cho, Hallie S, Manuel E Sosa, and Sameer Hasija (2022). "Reading between the stars: Understanding the effects of online customer reviews on product demand". In: *Manufacturing & Service Operations Management* 24.4, pp. 1977–1996.
Clemons, Eric K, Guodong Gordon Gao, and Lorin M Hitt (2006). "When online reviews meet hyperdifferentiation: A study of the craft beer industry". In: *Journal of management information systems* 23.2, pp. 149–171.
Cohen, Joel B and Ellen Golden (1972). "Informational social influence and product evaluation." In: *Journal of applied Psychology* 56.1, p. 54.
Dellarocas, Chrysanthos, Xiaoquan Zhang, and Neveen F Awad (2007). "Exploring the value of online product reviews in forecasting sales: The case of motion pictures". In: *Journal of Interactive marketing* 21.4, pp. 23–45.
Ding, Amy Wenxuan and Shibo Li (2019). "Herding in the consumption and purchase of digital goods and moderators of the herding bias". In: *Journal of the Academy of Marketing Science* 47, pp. 460–478.
Duan, Wenjing, Bin Gu, and Andrew B Whinston (2008). "The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry". In: *Journal of retailing* 84.2, pp. 233–242.
Forbes (2022). Accessed on December 23, 2023. URL: https://www.forbes.com/sites/forbesbusinesscouncil/2022/10/13/how-online-reviews-can-help-stores-grow/?sh=659bb49962d6.
Goes, Paulo B, Mingfeng Lin, and Ching-man Au Yeung (2014). ""Popularity effect" in user-generated content: Evidence from online product reviews". In: *Information Systems Research* 25.2, pp. 222–238.
Ho, Yi-Chun, Junjie Wu, and Yong Tan (2017). "Disconfirmation effect on online rating behavior: A structural model". In: *Information Systems Research* 28.3, pp. 626–642.
Hu, Ye and Xinxin Li (2011). "Context-dependent product evaluations: an empirical analysis of internet book reviews". In: *Journal of Interactive Marketing* 25.3, pp. 123–133.
Huang, Ni, Yili Hong, and Gordon Burtch (2017). "Social network integration and user content generation". In: *MIS quarterly* 41.4, pp. 1035–1058.
Ketelaar, Paul E et al. (2015). "The good, the bad, and the expert: how consumer expertise affects review valence effects on purchase intentions in online product reviews". In: *Journal of Computer-Mediated Communication* 20.6, pp. 649–666.
Lee, Nah, Bryan Bollinger, and Richard Staelin (2023). "Vertical versus horizontal variance in online reviews and their impact on demand". In: *Journal of Marketing Research* 60.1, pp. 130–154.

Lee, Young-Jin, Kartik Hosanagar, and Yong Tan (2015). “Do I follow my friends or the crowd? Information cascades in online movie ratings”. In: *Management Science* 61.9, pp. 2241–2258.

Lembregts, Christophe, Jeroen Schepers, and Arne De Keyser (2024). “Is it as bad as it looks? Judgments of quantitative scores depend on their presentation format”. In: *Journal of Marketing Research* 61.5, pp. 937–954.

Li, Linyi, Shyam Gopinath, and Stephen J Carson (2022). “History matters: the impact of online customer reviews across product generations”. In: *Management Science* 68.5, pp. 3878–3903.

Li, Xinxin and Lorin M Hitt (2008). “Self-selection and information role of online product reviews”. In: *Information systems research* 19.4, pp. 456–474.

Liu, Yong (2006). “Word of mouth for movies: Its dynamics and impact on box office revenue”. In: *Journal of marketing* 70.3, pp. 74–89.

Mitchell, Andrew A and Peter A Dacin (1996). “The assessment of alternative measures of consumer expertise”. In: *Journal of consumer research* 23.3, pp. 219–239.

Moe, Wendy W and David A Schweidel (2012). “Online product opinions: Incidence, evaluation, and evolution”. In: *Marketing Science* 31.3, pp. 372–386.

Moe, Wendy W and Michael Trusov (2011). “The value of social dynamics in online product ratings forums”. In: *Journal of marketing research* 48.3, pp. 444–456.

Muchnik, Lev, Sinan Aral, and Sean J Taylor (2013). “Social influence bias: A randomized experiment”. In: *Science* 341.6146, pp. 647–651.

Office for National Statistics, UK (2021). Accessed on June 10, 2021. URL: https://www.ons.gov.uk/businessindustryandtrade/retailindu%20stry/timeseries/j4mc/drsi.

Park, C Whan and V Parker Lessig (1981). “Familiarity and its impact on consumer decision biases and heuristics”. In: *Journal of consumer research* 8.2, pp. 223–230.

Park, Do-Hyung and Sara Kim (2008). “The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews”. In: *Electronic commerce research and applications* 7.4, pp. 399–410.

Ratchford, Brian T (2001). “The economics of consumer knowledge”. In: *Journal of Consumer Research* 27.4, pp. 397–411.

Schlosser, Ann E (2005). “Posting versus lurking: Communicating in a multiple audience context”. In: *Journal of Consumer Research* 32.2, pp. 260–265.

Schneider, Christoph et al. (2021). “When the stars shine too bright: The influence of multidimensional ratings on online consumer ratings”. In: *Management Science* 67.6, pp. 3871–3898.

Sunder, Sarang, Kihyun Hannah Kim, and Eric A Yorkston (2019). “What drives herding behavior in online ratings? The role of rater experience, product portfolio, and diverging opinions”. In: *Journal of Marketing* 83.6, pp. 93–112.

Tepper, Kelly (1997). “Categories, contexts, and conflicts of consumers’ nonconformity experiences”. In.

Tepper, Kelly, William O Bearden, and Gary L Hunter (2001). “Consumers’ need for uniqueness: Scale development and validation”. In: *Journal of consumer research* 28.1, pp. 50–66.

Wang, Chong, Xiaoquan Zhang, and Il-Horn Hann (2018). “Socially nudged: A quasi-experimental study of friends’ social influence in online product ratings”. In: *Information Systems Research* 29.3, pp. 641–655.

Wu, Fang and Bernardo A Huberman (2008). “How public opinion forms”. In: *International Workshop on Internet and Network Economics*. Springer, pp. 334–341.

Chapter 4

Conclusion

Online product ratings are a cornerstone of e-commerce, influencing consumer decisions and shaping market dynamics. This dissertation contributes to the growing body of literature on online product ratings by investigating the structural and behavioral factors that drive rating behaviors. Specifically, it highlights how rating system design and user dynamics influence the quality and reliability of product evaluations, providing valuable insights for both researchers and practitioners.

Chapter 2 examines the impact of transitioning from a single-dimensional to a multi-dimensional rating system on overall ratings and review content. By analyzing data from a major e-commerce platform, the study finds that multi-dimensional systems lead to lower overall ratings and higher-quality reviews. Importantly, the study demonstrates that multi-dimensional rating systems encourage users to focus more on product-related attributes while reducing attention to seller-related aspects. This shift enhances review quality and provides richer insights for potential buyers. These results extend existing literature on the influence of rating systems and underscore the importance of thoughtful system design in fostering authentic and actionable consumer feedback.

Chapter 3 explores a novel dimension of social influence—strategic rating behavior. Using data from a beer-rating platform, the study uncovers a tendency among users to provide lower ratings to products with a higher number of previous ratings, driven by strategic motives to affect overall ratings. Unlike other forms of social influence, such as herding or differentiation, strategic rating behavior represents a deliberate effort to skew aggregate evaluations. Interestingly, while ratings are influenced by such behavior, the sentiment and length of reviews remain unaffected, reinforcing the authenticity of textual feedback. These findings emphasize the need to account for strategic behavior when designing rating systems and contribute to the broader literature on user dynamics in online platforms.

Together, Chapters 2 and 3 provide a comprehensive view of the drivers of online product ratings. While Chapter 2 highlights the structural effects of rating system design on user behavior and review quality, Chapter 3 explores the behavioral dynamics that shape rating patterns in response to social influences. Both chapters underscore the importance of minimizing biases and enhancing the reliability of ratings to support better decision-making for consumers and businesses alike.

This dissertation not only advances theoretical understanding of online product ratings but also offers practical recommendations for e-commerce platforms. By designing systems that mitigate biases and promote transparency, platforms can ensure that ratings more accurately reflect user experiences. Additionally, the insights into strategic behavior suggest opportunities to develop mechanisms that discourage manipulative practices, safeguarding the credibility of user-generated content.

In conclusion, this research deepens our understanding of online product rating dynamics, bridging gaps in the literature and opening avenues for future work on consumer behavior in digital marketplaces. By focusing on both system-level and user-level influences, it provides a holistic perspective on the challenges and opportunities in managing online ratings.