How the Display of the Transaction Count Affects the Purchase Intention

Martin Poniatowski
Paderborn University, martin.poniatowski@wiwi.upb.de

How the Display of the Transaction Count Affects the Purchase Intention

Emergent Research Forum (ERF)

Martin Poniatowski
Paderborn University
martin.poniatowski@upb.de

Abstract

While the economic significance of reviews and ratings in e-commerce is undisputed, previous studies revealed that the significant underreporting bias inherent in aggregated ratings could mislead potential customers. In particular, the positively skewed distribution (i.e., J-shape) of ratings might lead to biased assumptions about a particular seller. Providing the transaction count of a seller as an additional aggregated metric might, however, mitigate the consequences of this bias by giving potential customers more information about the seller. While research suggests to display the transaction count in review systems, its effect has not yet been investigated. Therefore, we propose to conduct an experiment to understand if providing such a transaction count in a review system actually influences consumer behavior (e.g., purchase intention), and if so, how. With this research still in process, we invite a discussion of our experiment's design, and its potential for theoretical deductions.

Keywords

Review system design, underreporting bias, signaling theory, information asymmetry.

Introduction

Online review systems became an important information source for potential customers gathering information on products, services, sellers, and service providers. Such platforms enable past customers to share their experiences with potential customers by provding textual reviews and/or numerical ratings. Review systems aggregate ratings by displaying aggregated measures, like the valence, i.e., the average rating over all numerical ratings, or the distribution, i.e., the number of each numerical rating (Gutt et al., 2019). Hence, review systems send signals in form of informational cues representing quality, which help potential customers to reduce the information asymmetry between them and sellers (Pavlou et al., 2007). However, as not all past customers decide to share their experience, the review system's ability to reduce information asymmetry is impaired by the *underreporting bias*, potentially leading to distorted valences and distributions due to non-representative ratings (Hu et al., 2017). Consequently, a positively skewed rating distribution (i.e., J-shape) might mislead potential customers into believing that a particular seller is of higher quality than if it were reviewed by all customers. Indeed, Hu et al. (2017) suggest that there is a substantial difference between the rating distributions and the actual rating distribution based on all past customers' assessments. They also found that potential customers are aware of the bias but cannot fully account for it due to bounded rationality and are thus not able to correctly derive the underlying quality.

Therefore, by drawing on information signaling theory, the goal of this study is to investigate whether displaying an additional measure, in form of the count of previously completed transactions, has the potential to mitigate the consequences of the underreporting bias in online review systems. Displaying sellers' transaction counts provides a new informational cue, which reduces the information asymmetry between potential customers and sellers. It enables potential customers to derive the discrepancy between the transaction count and the rating volume, in other words, the number of missing ratings, and therefore helps them to account for the underreporting bias. In the case of a large discrepancy (e.g., only 10% of all transactions were rated by past customers) the underreporting bias is rather large and the aggregated measures might show non-representative information about the seller's quality. If the discrepancy is small (e.g., 90% of all transactions were rated), the aggregated measures come closer to representing the true quality. The transaction count acts as an informative signal enabling potential customers to better assess the sellers' quality by accounting for the underreporting bias and hence reduces the information asymmetry between them (Akerlof, 1970; Spence, 1973). Thus, this study's objective is to investigate the impact of sellers' transaction count as a way of providing more information that might influence potential customers in their purchase decision. Consequently, the following research question arises: *How does displaying the transaction count influence the purchase intention of potential customers?*

Related Literature

This study adds to the stream of literature on the perception of aggregated measures being displayed in online review systems and their effect on consumer behavior. Online review systems usually provide multiple informational cues in form of aggregated measures, like the rating valence or volume, to provide potential customer with information so they can make a well-informed decision (Gutt et al., 2019). Several studies already focused on how these **aggregated measures** influence consumer. The seminal study by Chevalier and Mayzlin (2006) provides first empirical evidence for a statistically significant association between volume, valence, and sales. Based on data on book sales on amazon.com and barnesandnoble.com they demonstrate that an increase in the volume and valence of reviews on both platforms is associated with an increase in sales. Moreover, they found that the impact of one-star reviews is greater than that of five-star reviews. Liu (2006) analyze the rating behavior of new movies released in cinemas and how they are associated with box office revenue. By distinguishing between the valence—measured in positive or negative comments on the movie instead of star ratings—and the volume of ratings they conclude that the positive association between ratings and revenue is mostly due to the volume of ratings and not its valence. Their conclusion, i.e., that the relation between volume and purchases is more robust than the relation between valence and purchases, is in line with insights from further empirical studies (Duan et al., 2008a, 2008b).

However, the transaction count constitutes an aggregated measure as well, but so far no research has been conducted on how the display of the **transaction count** influences consumer behavior. One study conducted by Dellarocas and Wood (2008) utilizes information on so-called "silent transactions" (i.e., transactions, where one or both parties do not submit a rating) in bidirectional review systems[1] to develop an approach for estimating the rating distribution of the missing ratings. However, their insights cannot be applied to unidirectional review systems, nor do they allow any predictions to be made on how the transaction count, if displayed, might influence consumer behavior. Therefore, this study conducts the first step to closing this gap by contributing to the research stream on how the display the transaction count in online review systems is perceived and affects consumer behavior.

Theoretical Background and Hypothesis Development

Drawing on information signaling theory, this study aims to investigate how displaying the transaction count has the potential to mitigate the underreporting bias by reducing information asymmetry and consequently to influence the purchase intention of potential customers.

Based on Spence's **signaling theory**, Connelly et al. (2011) define three main constructs: signaler, signal, and receiver. *Signalers* are informed parties who possess information on an outcome (e.g., service quality). Signalers send *signals*, aimed at representing the (unobservable) quality of an outcome, to the *receivers*, who have less information than signalers and consider participating in the exchange based upon the signals sent upon the outcome (Spence, 1973, 2002; Stiglitz, 2002). Signals can be sent by multiple parties simultaneously and might vary in form of various information cues and the cost of sending signals (Connelly et al., 2011). The experiences with a certain seller provided by past customers (*signalers*) through online review systems are information cues representing quality, which help potential customers (*receivers*) to make well-informed decisions by reducing the information asymmetry between them and sellers. Previous studies suggest that uncertainties due to information asymmetry from a potential customer's perspective can be mitigated by signaling information cues like the rating valence, distribution or volume, so that potential customers are able to estimate a seller's quality (Dimoka et al., 2012; Pavlou et al., 2007). Hence,

[1] In this context bidirectional reviewing means that both the seller and the buyer can submit a rating on the transaction (e.g., in Ebay or Airbnb). Unidirectional reviewing means that only the buyer can submit a rating (e.g., Amazon).

signaling theory provides a theoretical perspective to understand consumer behavior when additional informational cues, like the transaction count, are displayed.

The **underreporting bias** leads to distorted information cues (i.e., aggregated measures) when a non-representative proportion of past customers share their experiences of a seller. Based solely on these non-representative measures (e.g., valence, distribution, volume) potential customers might be misled or not able to fully account for the underreporting bias (Hu et al., 2017). Consequently, when potential customers receive insufficient information cues that cannot adequately represent actual quality, providing an additional measure could substantially improve the decision situation from the perspective of potential customers by reducing information asymmetry.

According to signaling theory, displaying the transaction count would help potential customers to better distinguish high quality from low quality sellers by reducing the information asymmetry and accounting for the underreporting bias. Incorporating the new informational cue would enable them to assess the volume of ratings by knowing the transaction count, thus allowing them to receive a more comprehensive signal of sellers' quality. The smaller the discrepancy (e.g., 90% of transactions have ratings), the better the valence and distribution are in reflecting the actual experience of sellers and, consequently, the better the basis for a decision. In turn, if the discrepancy is larger (e.g., 10% of transactions have ratings), the less representative the valence and distribution, and the worse the basis for a decision. Consequently, the following hypothesis is derived: *A lower discrepancy between transaction count and volume increases the purchase intention.*

Experiment Design

To test our hypothesis, we plan to conduct an online experiment on Amazon Mechanical Turk (AMT). The experiment is designed as a situational experiment (Bendoly et al., 2006) which puts participants into the situation where they have to decide on purchasing an object of their choice from an online shop based on the information presented on its the shop's review page (see Figure 1). So far, we conducted a pretest with 100 participants. Following the suggestions from prior research (Buhrmester et al., 2018; Kees et al., 2017), access was restricted to participants who fulfilled certain requirements, i.e., an approval rate above 98%, an approved number of tasks higher than 1,000, and being located in the US. As suggested by Kees et al. (2017), participants were rewarded with 0.85 US$ each.

After giving an overview of the task and stressing the importance of reading very carefully as one cannot return to previous pages once the next-button is pressed, participants were asked to fill in information about their age and gender, and if they have ever made an online purchase. Next, the situation was given as: "Please imagine you would want to buy a product offered by this online store" before continuing to the review page of an online shop, which is an adapted form of a real review page (Figure 1).

Participants were randomly assigned to one of four groups (i.e., control, treatment 1 to 3). The control group represents the base case with no information on the transaction count of the online shop. Treatment 1 participants were presented with a review page with a large discrepancy between the volume and the transaction count (i.e., about 10% of transactions that resulted in a rating), either for the last 12 months or in total. Participants of treatment groups 2 and 3 were presented with a mediocre and small discrepancy, and the only information we changed here was the display of the transaction count. The page that followed the review page included attention checks, asking for information provided on the review page, like the color of the online shop logo, the numerical average star rating, the number of ratings and purchases for the last 12 months and in total. Hence, depending on the condition of the participants, they had to answer four or six attention-check questions. Numerical questions were classified as correct when the answer did not deviate by more than 20% from the correct answer. Next, the participants were asked if they would purchase from the presented online shop and provided a 7-point Likert-scale from 1 ("no, definitely not") to 7 ("yes, definitely"). Moreover, we provided a free text field for participants to explain, "what made you decide?", "which information led you to this decision?", and "what were your reasons?". After that, they were asked "did you find the information about the online shop useful?" and, again, provided a 7-point Likert-scale from 1 ("no, not useful at all") to 7 ("yes, very useful"). Here, the participants were asked to answer the question "what information was especially useful to you, and why?" in a text field.

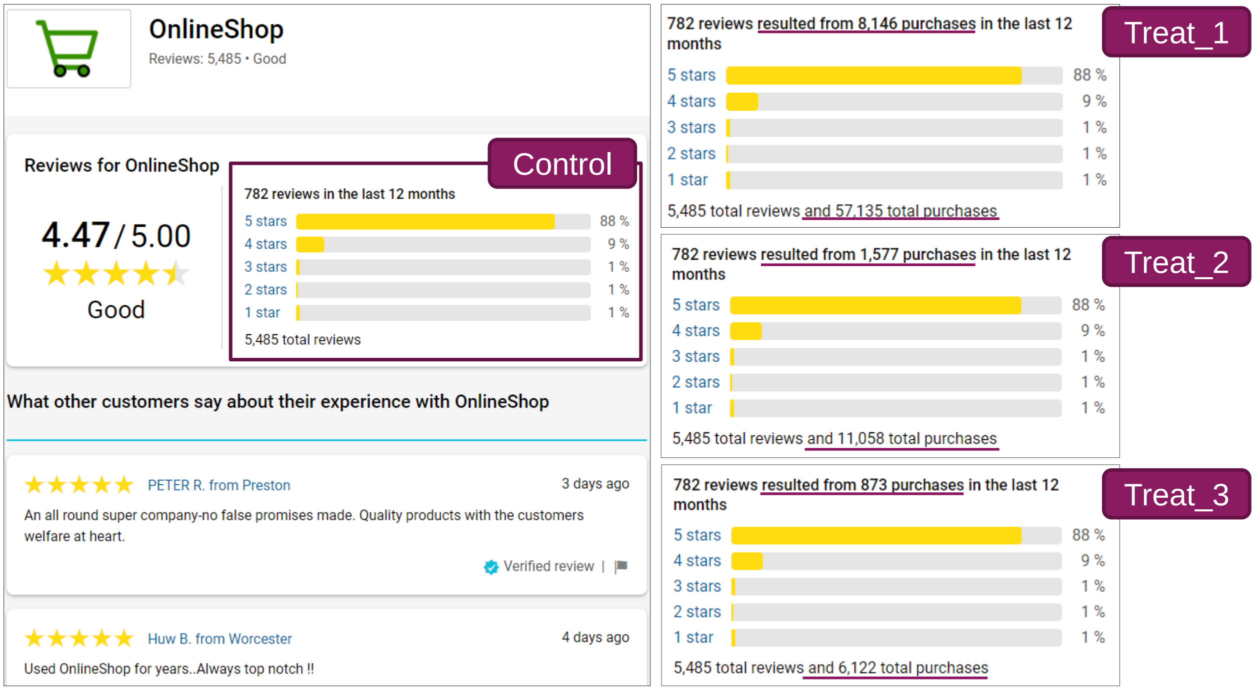

Figure 1: Review Pages of the Online Shop Depending on the Group

Preliminary Results and Further Steps

About 43% of the participants were female, and the average age was 39 years. Only 35 participants out of 100 answered all attention checks correctly. Therefore, the analysis is restricted to these participants. Table 1 presents selected variables for those participants.

Variable	N	Mean	Median	SD	Min	Max
Number of Purchases	34	7.85	5	8.95	1	50
Submitting Ratings/Reviews	34	0.47	0	0.51	0	1
Considering Ratings/Reviews (5-Point Likert)	34	3.88	4	0.96	2	5
Information Useful (7-Point Likert)	35	6.23	7	1.31	1	7
Purchase Intention (7-Point Likert)	35	6.01	6	1.17	2	7

Table 1: Descriptive Data of the Pretest

Despite the randomized assignment, our participants were evidently not evenly distributed to the groups (see Table 2). Many participants probably dropped the task and one of the reasons why they dropped the task could be the attention checks. Worried about not being able to answer the numerical questions on volume and transaction count correctly, and thus invalidating the task, which would potentially decrease their approval rate on AMT, they left the task even though non-completion meant missing out on the financial compensation. Unfortunately, the count on how many actually left is not provided by AMT.

Conducting the Kruskal-Wallis test to test for statistically significant differences between the groups shows no differences. This is not surprising, given the small sample size. However, the pretest was conducted to ensure that the task could be technically accomplished and the results are gathered correctly. To mitigate the low rate of correctly answered attention checks, we plan to add a further attention check at the beginning of the experiment, as suggested by Oppenheimer et al. (2009), who found that some attention checks tend to increase participants' attentiveness. Moreover, as suggested by participants' feedback on the task, the review text under the aggregated measures was a distraction (see Figure 1 lower left side), which is why we consider removing the reviews to enable participants to focus solely on the aggregated measures.

Group	N	Mean	Median	SD	Min	Max
Control	14	6	6.5	1.47	2	7
Treatment 1	7	5.86	6	1.35	3	7
Treatment 2	8	6.5	6.5	0.53	6	7
Treatment 3	6	6	6	0.89	5	7

Table 2: Descriptive Data of Purchase Intention (7-Point Likert)

Acknowledgements

This work was partially supported by the German Research Foundation (DFG) within the Collaborative Research Centre On-The-Fly Computing (GZ: SFB 901/3) under the project number 160364472.

References

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, *84*(3), 488.

Bendoly, E., Donohue, K., & Schultz, K. L. (2006). Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management*, *24*(6), 737–752.

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An Evaluation of Amazon's Mechanical Turk, Its Rapid Rise, and Its Effective Use. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *13*(2), 149–154.

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, *43*(3), 345–354.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management*, *37*(1), 39–67.

Dellarocas, C., & Wood, C. A. (2008). The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. *Management Science*, *54*(3), 460–476.

Dimoka, Hong, & Pavlou (2012). On Product Uncertainty in Online Markets: Theory and Evidence. *MIS Quarterly*, *36*(2), 395.

Duan, W., Gu, B., & Whinston, A. B. (2008a). Do online reviews matter? — An empirical investigation of panel data. *Decision Support Systems*, *45*(4), 1007–1016.

Duan, W., Gu, B., & Whinston, A. B. (2008b). The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *Journal of Retailing*, *84*(2), 233–242.

Gutt, D., Neumann, J., Zimmermann, S., Kundisch, D., & Chen, J. (2019). Design of review systems – A strategic instrument to shape online reviewing behavior and economic outcomes. *The Journal of Strategic Information Systems*, *28*(2), 104–117.

Hu, N., Pavlou, P. A., & Zhang, J. (2017). On Self-Selection Biases in Online Product Reviews. *MIS Quarterly*, *41*(2), 449–471.

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141–155.

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

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872.

Pavlou, Liang, & Xue (2007). Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal-Agent Perspective. *MIS Quarterly*, *31*(1), 105.

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

Spence, M. (2002). Signaling in Retrospect and the Informational Structure of Markets. *American Economic Review*, *92*(3), 434–459.

Stiglitz, J. E. (2002). Information and the Change in the Paradigm in Economics. *American Economic Review*, *92*(3), 460–501.