

Kumulative Dissertation zum Themengebiet:

**Inferring Quality with Reputation Systems – Experimental Evidence
on Elicitation Mechanisms and Aggregation Metrics**

Der Fakultät für Wirtschaftswissenschaften der
Universität Paderborn

zur Erlangung des Grades
Doktor der Wirtschaftswissenschaften
- Doctor rerum politicarum -

vorgelegt von
Dipl.-Wirt.-Ing. Dirk van Straaten
geboren am 24. März 1986 in Ostercappeln

2021

CONTENTS

1. Introduction and Synopsis

2. Paper of the Dissertation

- **Hoyer, B. and van Straaten, D. (2021):** Anonymity and Self-Expression in Online Rating Systems – An Experimental Analysis. Working Paper, Paderborn University.
- **van Straaten, D. (2021):** Incentive Schemes in Customer Rating Systems – Comparing the Effects of Unconditional and Conditional Rebates on Intrinsic Motivation. Working Paper, Paderborn University.
- **van Straaten, D., Melnikov, V., Hüllermeier, E., Mir Djawadi, B., Fahr, R. (2021):** Accounting for Heuristics in Reputation Systems: An Interdisciplinary Approach on Aggregation Processes. Working Paper, Paderborn University.
- **van Straaten, D., Fahr, R. (2021):** Fighting Fire with Fire – Overcoming Ambiguity Aversion by Introducing More Ambiguity. Working Paper, Paderborn University.

Introduction

Customer reviews are consulted by 82% of the customers before making online purchasing decisions (Smith and Anderson 2016). Given the growing number of customer reviews, reputation systems provide metrics of customer ratings as measures of valence (i.e., the level of positivity or negativity). Although these metrics of valence provide information about the satisfaction of customers with products and services, they however do not reflect objective product quality (De Langhe et al. 2016). This discrepancy can result in severe problems as the interest of customers in reading customer reviews is also due to the possibility to access information about the quality of the product (Kim 2020). Hu et al. (2017) identify that the deviation of quality and average ratings is driven by two self-selection biases, namely the acquisition bias (i.e., customers consciously choose the purchased products, resulting in higher propensities of satisfaction and thus higher ratings) and the under-reporting bias (i.e., customers show a higher propensity to publish reviews while feeling extreme satisfaction or dissatisfaction). These self-selection biases result in severe problems as this causes a lack of important experiences from customer groups that do not publish reviews, ultimately giving rise to biases in quality assessments. Addressing the self-selection of customers who provide reviews, Askalidis et al. (2017) propose eliciting reviews from customer groups that would otherwise not publish their experience, thus widening the sample to “represent the population of all buyers” (p. 25). However, instruments for eliciting customer reviews are very sophisticated as they can also cause crowding out of customers who would submit their experience without these instruments (Vilnai-Yavetz and Levina 2018; Khern-am nuai et al. 2018). Moreover, there is evidence that the quality of reviews is also negatively affected by financial incentives (Burtch et al. 2017; Khern-am nuai et al. 2018).

Aiming at a higher participation of customers with less side effects, this dissertation thus examines alternative measures for eliciting customer reviews in reputation systems. Although reducing the under-reporting helps obtaining a more representative sample, biases in metrics of customer ratings will persist due to the acquisition bias. As customers buy items online more often, they establish an inherent knowledge by gaining experience on customer reviews, and on phenomena such as the large share of 5-star ratings or the j-shape of customer rating distributions (Hu et al. 2009). Using this inherent knowledge of the customers, an additional goal of this dissertation is to adapt the metrics measuring the valence of customer ratings. Emphasizing the central aim, this dissertation aspires to *narrow the gap between objective quality and the metrics that measure the valence of customer*

ratings. To do this, the effect of self-selection biases on customers' perception of customer ratings as well as the various measures aiming at reducing these self-selection biases are investigated. The following paragraphs provide short summaries of the four papers of this dissertation, which present the research questions and the central results.

In the first article, *Hoyer and van Straaten (2021)*, we investigate how the motive of self-expression drives rating behavior in customer review systems. We therefore develop an experimental market in which subjects act as customers, purchasing products and afterwards choosing whether to provide a costly customer rating to signal the quality of the purchased product. Thereby, the treatment variation of whether subjects act anonymously or under a self-determined pseudonym allows us to investigate the impact of anonymity or self-expression, respectively, on the propensity to provide customer ratings.

Controlling for the altruistic attitudes of the subjects as a measure of intrinsic motivation, our results show that self-expression is indeed a driver of customer ratings, even in artificial markets. However, intrinsically motivated subjects are not affected by the introduction of anonymity, as they publish similarly under both treatment conditions. In contrast, the lack of self-expression under anonymity drops the ratings of subjects that are not intrinsically motivated.

This study has two major contributions. First, we find supporting evidence that self-expression indeed drives rating behavior in customer review systems. Contrary to preceding articles identifying self-expression as a motive of customer reviews (Hennig-Thurau et al. 2004; Cheung and Lee 2012), we thereby employ an incentivized set-up and contribute with our methodological approach, ruling out the doubts about hypothetical set-ups. Second, market providers might consider increasing the degree of self-expression as a motive to elicit more reviews. However, as intrinsically motivated subjects are not affected by the introduction of anonymity, decreasing the degree of self-expression in reputation systems might also be considered when only intrinsically motivated customer ratings are desired due to higher helpfulness (cf. Peddibhotla and Subramani 2007).

In the second article, *van Straaten (2021)*, I compare the effects of different incentive schemes on the propensity to publish customer ratings, concentrating in particular on unconditional rebates and conditional rebates. To do this, I present a novel experimental design in which the motives for customer ratings are restricted to indirect peer reciprocity, altruism (i.e., intrinsic motivation), seller reciprocity, and economic incentives.

The results of this economic laboratory experiment indicate that economic incentives have the strongest effect on the propensity to publish a customer rating. Moreover, the propensity to publish ratings drops for intrinsically motivated subjects. In comparison, unconditional rebates have a weaker positive effect on the propensity to publish a rating, showing the benefit of not crowding out intrinsically motivated customer ratings.

This study contributes to the growing body of literature on the design of customer review systems (cf. Gutt et al. 2019). In particular, results of this experimental study give insights into the motives behind publishing customer ratings and the effect of extrinsic motivational factors on the intrinsic motivation. Given the higher helpfulness of intrinsically motivated customer reviews, the major contribution of this study is that unconditional rebates do not crowd out intrinsic motivation and are therefore a valid alternative for eliciting additional feedback about products and services in reputation systems. Even though conditional rebates elicit more additional feedback, this accompanies crowding out customer ratings driven by intrinsic motivation.

In the third article, *van Straaten et al. (2021)*, we investigate how customers assess customer rating distributions and whether the arithmetic mean or other aggregation functions match their preferences. By conducting an economic experiment, participants rank customer rating distributions of underlying real products that subjects have the chance to win. The decisions are thus incentivized, which allows us to estimate the category weights of the subjects and contrast these with reference functions.

The results provide evidence that the behavior of the subjects can be explained best by employing the arithmetic mean. However, the analysis on an individual level reveals an additional trend for the majority of overweighting moderate ratings (i.e., 2-star and 4-star) and underweighting extreme ratings (i.e., 1-star and 5-star), resulting in the binary bias. In addition, minor clusters show various strategies such as minimizing 1-star ratings or minimizing negative ratings (i.e., 1-star and 2-star), which highlights the heterogeneity across subjects. Contrary to our predictions, category weights do not differ with regard to changes of the available information, that is, whether numerical information is provided.

This article contributes to a deeper understanding of aggregation heuristics of customers with regard to the employed aggregation metrics in customer review systems. Our results indicate a systematic binary bias of subjects in the evaluation of customer ratings, which has important implications for practitioners as the aggregation metrics provided by the market makers do not consider

these biases so far. Doing so can increase market efficiency. As minor clusters strongly deviate from the arithmetic mean and focus on different aspects (e.g., minimizing 1-star ratings), this heterogeneity across customers provides a second contribution by giving rise to the question whether measures of valence should be implemented for each customer individually.

In the fourth article of this dissertation, *van Straaten and Fahr (2021)*, we investigate the effect of multiple sources of ambiguity on decision making. Defining ambiguity in accordance with Einhorn and Hogarth (1985) as uncertainty emerging from unknown information about the output generating process, we widen the scope of the well-established ambiguity literature, which has its origin in the urn experiment of Ellsberg (1961) and focus on uncertainty about probabilities (Camerer and Weber 1992).

Using the experimental design of van Straaten et al. (2021) and changing the decision domain by letting subjects rank lotteries (i.e., risky prospects) instead of customer rating distributions, we identify the effect of ambiguity due to an unknown source credibility of reviewers and the ordinal star scale on decision making. By also varying the provided information (i.e., only visual information or visual enriched with numerical information), we include a second source of ambiguity.

Controlling for risk preferences, we estimate category weights of each subject and find differences between customer rating distributions and lotteries, indicating ambiguity aversion. Contrary to our predictions on the amplifying effect of multiple sources of ambiguity, we find that the introduction of a second source of ambiguity has different effects on decision making under risk and ambiguity. That is, instead of constant differences between both decision domains, there is evidence for convergence of category weights when no numerical information is provided.

This article contributes to the literature on decision making under ambiguity as we provide a more efficient approach on how to deal with situations in which people have to choose between risky and ambiguous prospects. Instead of minimizing ambiguity, a second source of ambiguity might be introduced for both alternatives to compare alternatives with less biases due to ambiguity aversion.

Summarizing the four articles, this dissertation points out approaches to address the challenges emerging by the self-selection bias of customer review systems. The results of Hoyer and van Straaten (2021) provide evidence that addressing self-expression as a motive in customer review systems is a valid approach to reduce the under-reporting bias. In van Straaten (2021), I find evi-

dence that unconditional rebate mechanisms increase the propensity of subjects to submit customer ratings. Both self-expression and unconditional rebates show the advantage of not crowding out intrinsically motivated reviewers. Although most additional ratings are elicited by conditional rebates, this approach shows the drawback of crowding out intrinsically motivated reviewers.

The acquisition bias leads to a large share of positive ratings in customer rating systems, resulting in positively biased aggregation metrics (Hu et al. 2017). The rising inherent knowledge of the customers might also effect the manner in which they assess the information about the valence of customer rating distributions. In van Straaten et al. (2021) we identify that customers indeed show the tendency to underweight extreme ratings and overweight moderate ratings (in comparison with the arithmetic mean), which indicates that the acquisition bias might be taken into account inherently by customers. Moreover, the results in van Straaten and Fahr (2021) show that the deviations from the arithmetic mean are stronger in the domain of customer ratings (in comparison with decisions under risk) and thus seem to be driven by an unknown source credibility of the customer ratings and the ordinal stars-scale in customer review systems.

All the articles of this dissertation apply the methods of experimental economics with its benefits of controlling the environment and manipulating only the aspects of interest. For instance, in van Straaten et al. (2021) the aggregation of customer ratings are taken into account by implementing real customer rating distributions in the laboratory. Excluding thereby other information such as the prices, the specifications, the product names or pictures in the decision process, we eliminate much noise and, additionally, minimize a random decision making of the subjects by implementing incentives that are linked to the decision making. Knowing the cause (i.e., the information of customer rating distributions) and the effect (i.e., the ranking decisions of the subjects), while eliminating other impact factors (e.g., the price) and minimizing random choices (due to the implemented incentives), our experiment provides results with a sufficiently high internal validity that other methods would not have achieved. It is thus an example of empirical research, that can be realized best with methods of experimental economics.

Comparing the method of this dissertation with other methods in empirical research, latter often collect more observations and more variables of interest, thus having for instance more power in explaining interdependencies between variables and a higher external validity by analyzing data generated in the field. In this respect, experimental economics and other empirical methods are complementary as the implications of this dissertation point out directions for further research, calling

for investigations on the robustness and impact of the results outside the laboratory. For instance, the question to which degree the technical implementation of the identified aggregation patterns in van Straaten et al. (2021) effects the purchasing behavior and the outcomes in online marketplaces remains for future research.

Synopsis

Publication	Hoyer, B. and van Straaten, D. (2021): Anonymity and Self-Expression in Online Rating Systems – An Experimental Analysis. Working Paper, Paderborn University.
Contribution of joint work with co-authors	<ul style="list-style-type: none"> • Co-authorship with Dr. Britta Hoyer (B. Hoyer: 50%, D. van Straaten: 50%) • Idea by B. Hoyer • Development of experimental design jointly • Experimental procedure jointly • Analysis by D. van Straaten • Write-up of paper jointly
Conferences / Workshops	<p>The paper was presented by D. van Straaten at the following conferences and workshops:</p> <ul style="list-style-type: none"> • 09/2018: Annual Meeting 2018 of “Gesellschaft für experimentelle Wirtschaftsforschung e.V.” (GfeW), Paderborn, Germany • 07/2018: SABE-IAREP conference 2018, London, England
Scientific dissemination	<ul style="list-style-type: none"> • Start of work: December 2017 • First draft: March 2019 • Draft submitted to <i>Journal of Experimental Economics</i>: November 2020 • Current draft under major revision; scheduled resubmission: July 2021

Publication	van Straaten, D. (2021): Incentive Schemes in Customer Review Systems – Comparing the Effects of Unconditional and Conditional Rebates on Intrinsic Motivation. Working Paper, Paderborn University.
Contribution of joint work with co-authors	This work is single-authored.
Conferences / Workshops	<p>The paper was presented by D. van Straaten at the following conferences and workshops:</p> <ul style="list-style-type: none"> • 09/2019: Poster presentation at “Faculty Research Workshop”, Paderborn University, Melle, Germany. <p>The current draft is submitted to “IAREP-SABE Conference 2021”.</p>
Scientific dissemination	<ul style="list-style-type: none"> • Start of work: September 2019 • First draft: December 2020 • Current draft: April 2021

Publication	van Straaten, D., Melnikov, V., Hüllermeier, E., Mir Djawadi, B., Fahr, R. (2021): Accounting for Heuristics in Reputation Systems: An Interdisciplinary Approach on Aggregation Processes. Working Paper, Paderborn University.
Contribution of joint work with co-authors	<ul style="list-style-type: none"> Co-authorship with Vitalik Melnikov, Prof. Dr. Eyke Hüllermeier, Dr. Behnud Mir Djawadi, Prof. Dr. René Fahr (D. van Straaten: 40%, V. Melnikov: 20%, E. Hüllermeier: 10%, B. Mir Djawadi: 15%, R. Fahr: 15%) Idea and development of experimental design jointly Experimental procedure by D. van Straaten and B. Mir Djawadi Analysis by V. Melnikov and D. van Straaten Write-up of paper by D. van Straaten (Introduction, Theory, Experimental Design, Results, Conclusion) and V. Melnikov (Data and Statistical Model, Results); feedback, comments, and corrections by R. Fahr
Conferences / Workshops	<p>The paper was presented by D. van Straaten at the following conferences and workshops:</p> <ul style="list-style-type: none"> 07/2017: SABE 2017 conference, Newcastle, Australia 09/2016: Annual Meeting 2016 of "Gesellschaft für experimentelle Wirtschaftsforschung e.V." (GfeW), Gießen, Germany
Scientific dissemination	<ul style="list-style-type: none"> Start of work: February 2016 Pre-tests at "Wissenschaftstage 2016" in 06/2016 and in the laboratory 12/2016; Conduction 12/2018 First draft: November 2019 Current draft: April 2021 Submission to <i>Decision Support Systems</i> planned in 2021 Note: Small parts of this research project could also feed into the dissertation of my co-author Vitalik Melnikov.

Publication	van Straaten, D., Fahr, R. (2021): Fighting Fire with Fire – Overcoming Ambiguity Aversion by Introducing More Ambiguity. Working Paper, Paderborn University.
Contribution of joint work with co-authors	<ul style="list-style-type: none"> • Co-authorship with Prof. Dr. René Fahr (90% D. van Straaten; 10% R. Fahr) • Idea and experimental design by D. van Straaten • Programming of experiment in <i>z-tree</i> and experimental procedure by D. van Straaten • Analysis by D. van Straaten • Write-up of paper by D. van Straaten • Feedback and comments by R. Fahr
Conferences / Workshops	<p>The paper was presented by D. van Straaten at the following conferences and workshops:</p> <ul style="list-style-type: none"> • 09/2017: Annual Meeting 2017 of „Gesellschaft für experimentelle Wirtschaftsforschung e.V.“ (GfeW), Kassel, Germany <p>The current draft is submitted to the “VfS 2021 Annual Conference”.</p>
Scientific dissemination	<ul style="list-style-type: none"> • Start of work: July 2017 • First draft: August 2020 • Current draft: April 2021

References

Askalidis, G., Kim, S. J., and Malthouse, E. C. (2017). Understanding and overcoming biases in online review systems. *Decision Support Systems*, 97:23–30.

Burtch, G., Hong, Y., Bapna, R., and Griskevicius, V. (2017). Stimulating online reviews by combining financial incentives and social norms. *Management Science*, 64(5):2065–2082.

Camerer, C. and Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of risk and uncertainty*, 5(4):325–370.

Cheung, C. M. and Lee, M. K. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision support systems*, 53(1):218–225.

De Langhe, B., Fernbach, P. M., and Lichtenstein, D. R. (2016). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *Journal of Consumer Research*, 42(6):817–833.

Einhorn, H. J. and Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological review*, 92(4):433.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The quarterly journal of economics*, pages 643–669.

Gutt, D., Neumann, J., Zimmermann, S., Kundisch, D., and 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.

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

Hoyer, B. and van Straaten, D. (2021). Anonymity and Self-Expression in Online Rating Systems - An Experimental Analysis. Working Papers Dissertations 70, Paderborn University, Faculty of Business Administration and Economics.

Hu, N., Pavlou, P. A., and Zhang, J. J. (2017). On self-selection biases in online product reviews. *MIS Q.*, 41(2):449–471.

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

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

Kim, R. Y. (2020). When does online review matter to consumers? the effect of product quality information cues. *Electronic Commerce Research*, pages 1–20.

Peddibhotla, N. B. and Subramani, M. R. (2007). Contributing to public document repositories: A critical mass theory perspective. *Organization Studies*, 28(3):327–346.

Smith, A. and Anderson, M. (2016). Online Shopping and E-Commerce. Online Reviews. Accessed April 30, 2021. <https://www.pewresearch.org/internet/2016/12/19/online-reviews>.

van Straaten, D. (2021). Incentive Schemes in Customer Rating Systems - Comparing the Effects of Unconditional and Conditional Rebates on Intrinsic Motivation. Working Papers Dissertations 71, Paderborn University, Faculty of Business Administration and Economics.

van Straaten, D. and Fahr, R. (2021). Fighting fire with fire - overcoming ambiguity aversion by introducing more ambiguity. Working Papers Dissertations 73, Paderborn University, Faculty of Business Administration and Economics.

van Straaten, D., Melnikov, V., Hüllermeier, E., Mir Djawadi, B., and Fahr, R. (2021). Accounting for heuristics in reputation systems: An interdisciplinary approach on aggregation processes. Working Papers Dissertations 72, Paderborn University, Faculty of Business Administration and Economics.

Vilnai-Yavetz, I. and Levina, O. (2018). Motivating social sharing of e-business content: Intrinsic motivation, extrinsic motivation, or crowding-out effect? *Computers in Human Behavior*, 79:181–191.