Subproject A4: Empirical Analysis in Markets for OTF Services

Behnud Mir Djawadi[1], Alina Elrich[1], René Fahr[1], Bernd Frick[1], Daniel Kaimann[1], Dennis Kundisch[1], Michelle Müller[1], Martin Poniatowski[1], Sabrina Schäfers[1]

1 Faculty of Business Administration and Economics, Paderborn University, Paderborn, Germany

1 Introduction

In OTF markets, sophisticated combined services in mainly small quantities with predominantly experience attributes are traded. These attributes, which can hardly be known before using the service, and the many actors involved in creating these services result in unavoidable information asymmetries with corresponding adverse effects on the market outcome, including the risk of market failure [ake78]. Due to these characteristics of the OTF market, online reviews and certificates are significant in reducing information asymmetries between the service provider and the customer about the quality of the traded services. While a substantial body of literature emerged that examines to what extent online reviews are an effective measure to mitigate this problem, several research gaps are identified for the study of online reviews in OTF markets.

A) Influence of service and market characteristics on online reviews and economic outcomes

Understanding the influence of service and market characteristics on online reviews and economic outcomes is crucial for developing a successful OTF market in which all participants can generate (economic) benefits. Prior research has revealed that an increase in the average rating, as well as the total number of reviews from consumers, have a positive effect on the demand of a product or service and the price offered by a firm (e.g.,[LH08]). Moreover, several studies have estimated the to what extent reviews by professional critics impact the sales performance [HMM12]. However, a significant limitation of many of these studies is that they tend to control either only for consumer reviews or only for professional critics, even though Chintagunta et al. [CGV10] have highlighted the potentially significant differences between reviews from professional critics and users. Therefore, as a part of the CRC, Cox and Kaimann [CK15] analyze the relationship between economic outcomes and two signals of product quality: reviews from professional critics and reviews from

behnud.djawadi@wiwi.uni-paderborn.de (Behnud Mir Djawadi), alina.elrich@uni-paderborn.de (Alina Elrich), rene.fahr@wiwi.uni-paderborn.de (René Fahr), bernd.frick@wiwi.upb.de (Bernd Frick), daniel.kaimann@uni-paderborn.de (Daniel Kaimann), dennis.kundisch@wiwi.uni-paderborn.de (Dennis Kundisch), michelle.mueller@wiwi.uni-paderborn.de (Michelle Müller), martin.poniatowski@wiwi.uni-paderborn.de (Martin Poniatowski), sabrina.schaefers@uni-paderborn.de (Sabrina Schäfers)

consumers. Additionally, as the current literature has only partially examined the influence of ratings and ad expenditures separately and jointly on a product's or service's demand in online markets, Frick and Kaimann [FK17] shed light on this relationship. Moreover, what is less well understood in the online review literature are the effects of the variance of the online reviews (i.e., the distribution of the online ratings) on economic outcomes on the level of firms. To address this gap, Zimmermann et al. [ZHKN18] developed an analytical model that decomposes the variance of online ratings into two sources: *taste differences* in search and experience attributes of a product or service and *quality differences* among instances of this product or service in the form of product failure. In addition, on the more aggregated market level, it has remained unclear what impact the local market competition has on the heterogeneity of available businesses in a market. Gutt et al. [GHR19] contributed to the answer to this question by empirically investigating the relationship between the range and average of the mean rating distribution in the market and market competition.

B) Opportunistic behavior of service providers

Complex services are combined in OTF markets. Customers may not have the expertise to judge the service or product quality correctly. In OTF markets, there is the potential that customers may make mistakes in evaluating the true quality of the provided service. This inaccurate customer feedback, however, may impact the service provider's decisions about the quality. For example, the service provider might be tempted to charge a high price but deliver low quality, as the customer erroneously may rate the service as high quality. While the current literature has addressed the effect of inaccurate customer feedback on the functioning of reputation systems (e.g., [MDC14]), it only focused on intentional customer behavior. For example, customers intentionally fake transactions and provide dishonest feedback to damage the seller's or service provider's reputation. The study by Mir Djawadi et al. [MFHR18] closes this research gap by examining how inaccurate customer feedback based on unintentional mistakes influences the strategic quality choices of a service provider. As a result of this, the focus lies on the behavior of the service providers, whether they seek to maintain a good reputation by providing high-quality services, even if it is tempting that the customer might not detect low quality, or whether they milk a good reputation and exploit the customers.

C) Design of online review systems in OTF markets

For some design elements, literature already shows how design decisions influence the effect of drivers of writing a review (e.g. self-expression, altruism, or a sense of belonging) on online ratings. For example, the number of evaluation dimensions is related to the evaluation level. However, there are significant research gaps in the impact and potential of design decisions in OTF-like markets. Such design decisions include metrics, the degree of anonymity, or social proximity between trading partners. Metrics (i.e., aggregated measurements of customer feedback) enable the evaluation of product quality and the comparability of products on marketplaces. Since cognitive biases identified by psychology and behavioral economics affect people's perception and way of thinking, it is unclear whether the technically implemented aggregation functions correspond to customers' actual aggregation behavior. To address this, van Straaten et al. [SMH+21] elicit customers' aggregation heuristics and contrast these with reference functions.

Anonymous reputation systems can be implemented to avoid concerns about sharing private information by providing a customer review. However, this anonymity can potentially crowd out reviews motivated by customers' self-expression (one of the main motives for writing reviews). It is unclear how anonymity and its driving forces (such as blunting self-presentation) affect customers' propensity to write reviews [RS12]. Therefore, Hoyer and van Straaten [HS22] investigate whether anonymous reputation systems have the disadvantage of blunting self-presentation and whether altruistic attitudes moderate this effect.

A growing body of research looks at mutual evaluations and demonstrates that such situations typically result in valuations that are biased upward in magnitude. This, in turn, reduces market efficiency and, in the worst case, leads to market failures. Mir Djawadi and Wester (2022) [MW22] examine to what extent social proximity between trading partners leads to upward-biased ratings.

D) Certification to reveal service quality in OTF markets

Another essential instrument for reducing information asymmetries in markets for complex (combined) physical and digital services is a certification (such as certified consulting services or information security certification in online shopping). Certification by an external agent allows providers to signal (minimum) standards for the certified product or service characteristics, thus reducing information asymmetries.

While the effects of certifications are clear from a theoretical point of view, their empirical relevance has not been convincingly documented. Accordingly, the study by Fanasch and Frick [FF20] compares the effects of self-declaration and certification on product prices by distinguishing between certified products and service characteristics. The results of this study shed light on the unique relationship of certifications to reduce information asymmetries in OTF markets.

The composition of individual services causes additional complexity when analyzing service quality certification. In particular, it is unclear whether certification of all individual services or combined services is necessary. Consequently, Fanasch and Frick [FF20] provide evidence on whether and to what extent collective reputation for combined services impacts market prices.

2 Main Contributions

In the following, the main contributions are described in greater detail, which are meant to address the research gaps regarding the mitigation of information asymmetries in OTF markets introduced before.

2.1 Influence of Service and Market Characteristics on Online Reviews and Economic Outcomes

Interaction and impact of reviews from professional critics and customer reviews on the market performance of experience goods

Due to the digitalization of retail, customers can express their product experiences via user-generated reviews in various discussion forums, assessment portals, or on retailer websites. We contend that independent information expressed in product reviews has three plausible ways in which it can influence consumer behavior. First, more positive evaluations lower uncertainty over product quality among prospective consumers. Secondly, if evaluations reach near unanimity in opinion, then customers' certainty in their purchase decision will increase. Thirdly, the more reviews submitted, the lower the evaluation insecurity among future buyers. We focus on the relationship between the different facets of customer reviews: namely, valence, volume, and variance.

Valence represents the average weighted review score from customers. The use of weighted online average scores is consistent with prior studies of experience goods, e.g., movies, books, and magazines [CGV10]. Volume measures the total number of reviews posted. We expect that more significant numbers of ratings reduce the information asymmetry and thus positively affect consumer choices. Our measure of variance is the sum of squares of the proportion of positive, negative, and mixed opinions among the total number of reviews. It is thus analogous to the Hirschman-Herfindahl index, which captures the degree of homogeneity or heterogeneity of any variable of interest. Our variance variable is therefore bounded between an upper value of 1 (perfect homogeneity) and a lower value of 0.33 (perfect heterogeneity; equal proportions of reviews in each of the three opinion categories).

Cox and Kaimann [CK15] have focused on the relationship between commercial performance and two signals of product quality: reviews from professional critics and online word-of-mouth. By analyzing the consumer behavior and sales performance in the digital platform industry of video games, we shed light on the literature by comparing the relative influences of consumer word-of-mouth with reviews from professional critics. To empirically estimate and separate the effects of the two signals, we analyze a sample of 1,480 video games and their sales figures between 2004 and 2010. Such a comparison is potentially valuable given the prevailing view that word-of-mouth and other content generated by users is becoming increasingly influential in consumer decision-making, possibly even to the exclusion of traditional reliance on the opinions of experts or professional critics. Our analysis considers the possibility that these signals affect consumer behavior jointly and separately through a more detailed examination of interactions between signals and the subsequent effect on sales performance.

The findings of this study reinforce the hypothesis that the reviews of professional critics associate strongly with commercial success. After taking steps to control for endogeneity using a generalized method of moments (GMM) estimator, we find evidence that reviews from professional critics influence sales instead of merely predicting them, suggesting that their independence and reputation serve as a credible signal that helps consumers support the decision-making process by minimizing uncertainty. We also find only limited evidence to suggest that the valence of consumer word-of-mouth affects product sales once we control reviews from professional critics and interaction terms. Consequently, our results are contradictory to a commonly held belief in the value of consumer word-of-mouth and emphasize the greater importance of reviews from professional critics in the digital market context.

Regarding placing the study and findings within the context of the existing literature, the results reinforce some well-established findings while presenting alternative and sometimes contradictory evidence on others. We arrive at contradictory results in comparison with Chintagunta et al. [CGV10], given that they highlight the importance of the valence of word-of-mouth and find neither evidence for the significant explanatory power of reviews from professional critics nor the volume and valence of user reviews. We essentially arrive at the opposite conclusion, potentially due to our aggregation to the national level versus their regional-level analysis, an acknowledgment the authors make in their paper. Instead, our findings highlight the importance of simultaneously controlling for reviews from critics and consumer word-of-mouth.

Consequently, consumers likely assess the credibility and reliability of the interaction of distinct types and similar types of signals jointly. Although there is a shortage of studies in marketing on the interaction of signals, the economics literature has produced several studies on the value of multiple signals of product quality, most notably concerning the moral hazard associated with agency theory [Hol79] and most often considered concerning issues of corporate governance and performance-related pay [FS11]. However, the effect of additional signals may diminish at the margin as the total number of available signals increases. Basuroy et al. [BDT06] are among the first to empirically study the interaction between quality signals. Other authors, such as Kirmani and Rao [KR00], also account for the interaction between a limited number of independent types of signals in their theoretical framework. Following the studies and principles of signaling theory, we consider the importance of additional signals and their interactions.

Frick and Kaimann [FK17] have used real-world data to study the impact of customer reviews on market demand in electronic markets for mobile applications. Using data from the Apple App Store, we analyzed a sample of 32 applications with 5,792 daily observations and their number of installations during the first six months of 2015. The applications were randomly selected from each category in the App Store. The findings extend the current literature, which has only partially examined the influence of ratings and ad expenditures separately and jointly on downloads. The results show a positive interaction between valence and variance. In addition, the interaction between valence and variance has a more significant positive effect on quality perceptions and, thus, a more considerable impact on application downloads.

Furthermore, the empirical findings also support that customer reviews and marketing efforts boost installations. However, if they co-occur, the influence of both effects will be diminished. Thus, our findings advance reputation analyses by explicitly considering the possibility that these signals affect consumer behavior jointly and separately by conducting a more detailed analysis of interactions in electronic markets.

Effect of different sources of online review variance on product prices and demand at the

level of firms
Consumer ratings enable prospective consumers to learn from other consumers' experiences. This means that the reporting of experiences within an online rating enables new consumers to assess experience attributes of a product or service (i.e., those attributes that can hardly be observed prior to purchase but only after the purchase) better than before the emergence of consumer ratings. Thus, consumer ratings transform the product or service's experience attributes into attributes that can be searched within the reviews of this product or service. For instance, by reading online reviews, potential customers can learn about past customers' experiences concerning the ease of navigating over an application's menu. Some customers might like a simple menu, whereas others might prefer a more complex menu with highly adjustable settings. Customers' disagreements resulting from opposing opinions are thus caused by *taste differences* and might not necessarily imply a bad product or service for all consumers. However, potential customers are able to learn not only about the different taste-related experience attributes associated with an application but also about (potential) service failures, i.e., *quality differences*. For instance, if an application fails to work on a specific data set, consumers are also likely to report this failure within their online review. Both types of experience attributes (taste-related and quality-related) are likely to induce additional variance of a product. The key difference between both sources of variance is that all potential customers would agree about their dislike of variance caused by *quality differences* but not necessarily by the variance caused by *taste differences*.

To examine the relationship between the variance caused by *taste* or *quality differences* and a product's or service's prices and demand, the authors develop a two-period analytical model featuring a monopoly retailer and consumers that differ in taste and risk aversion. The first period explains the review-generation process, whereas, in the second period, the effects of how new consumers use the generated reviews from the first period in their decision-making process are examined. In the first period, where the product or service has no online reviews yet, a set of innovators enter the market. The retailer sets a profit-maximizing price based on its expectations about the product's or service's characteristics as well as the expected utility for the consumer. Afterward, the innovators decide whether to purchase the product or service based on the price and their expectations about the characteristics of the product or service. Then, the purchasing innovators (or at least those with extreme experiences) publish honest ratings about the product. In the second period, where online reviews for the product or service are present from the first period, a set of imitators enter the market. The retailer observes the consumer ratings of innovators and sets a profit-maximizing price for the product based on the observed consumer ratings. Afterward, imitators observe the consumer ratings and derive the product's or service's characteristics through these reviews. This means that imitators have no remaining uncertainty about the experience attributes of the product or service. Next, they decide whether or not to purchase the product or service based on the price and the observed consumer ratings. The authors analyze two types of goods: (1) consistent quality goods, where the variance of consumer ratings is solely caused by *taste differences* (model based on Sun [Sun12] to connect with prior literature), and (2) inconsistent quality goods, where the variance of consumer ratings is caused by *taste differences* and *quality differences* in the form of product failure.

Concerning the case of consistent quality goods, the theoretical model predicts that price and demand both increase with the average rating, as a higher average rating is a credible signal of high product quality. This result represents a theoretical confirmation of prior empirical findings [LH08]. With an increasing variance of ratings (which is solely caused by *taste differences*), the price increases while the demand for the good or service decreases. The intuition behind this relationship is as follows: An imitator with a taste that closely matches the product or service enjoys this product or service more than a product with a low variance of ratings. The retailer charges a higher price to all imitators to skim the higher willingness to pay of imitators with tastes that closely match the product or service. This higher price deters imitators with tastes that do not closely match the product or service, resulting in lower demand. Figure 6 illustrates the relationship between the products or services online rating variance and the price and demand of the product's or service's, respectively.

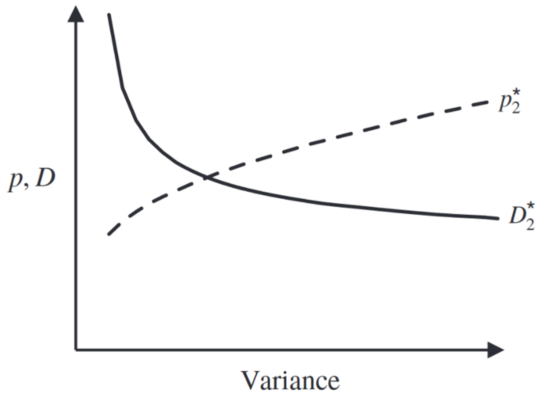

Figure 7: *Optimal price and demand for consistent quality goods. (Source: Own illustration)*

Note: For reasons of simplicity, the price and demand are plotted on the same axis. p_2^* represents the optimal price for the product or service in the second period, and D_2^* represents the optimal demand in the second period.

Regarding the analysis of a product or service that can be characterized as an inconsistent quality good (i.e., a product or service that can fail), the model differs in two major ways. First, the model now also takes the consumer's individual risk aversion (i.e., the negative utility caused by the risk that the product or service fails) into account, which in turn is associated with the consumer's purchase intentions. Second, if a consumer in the first period buys a product or service that failed, the consumer will publish the lowest rating possible. In contrast to consistent quality goods, the expected enjoyment of inconsistent quality goods depends not only on two but on three product characteristics: the average rating, the variance caused by *taste differences*, and the variance caused by *quality differences*. Similar to the case of consistent quality goods, the model predicts that price and demand will both increase with the average rating of an inconsistent quality good, as a higher rating acts as a credible signal of high product quality. Further, if the variance caused by taste differences increases, the price increases, and the demand decreases for an inconsistent quality good. The intuition behind this result is the same as in the consistent quality goods case. However, for inconsistent quality goods, if the variance caused by *quality differences* increases, the price of the product or service will decrease. This is because a higher variance caused by quality differences indicates a high failure rate of the product or service, which represents a signal for low quality and will thus reduce the price.

To examine the effects on the demand for the product or service, there are two distinct effects that need to be taken into consideration: (1) the price effect, which states that a reduced price will increase the demand, and (2) the failure effect, which models the relationship that a higher probability of product failure will decrease the demand. If the variance caused by *quality differences* is sufficiently low and the variance caused by *quality differences* increases, then the model predicts that the demand effect will be smaller than the failure effect, resulting in a decreased demand. A somewhat counterintuitive case occurs if the variance caused by *quality differences* is sufficiently high and the variance caused by *taste differences* is sufficiently low. Then, the price effect will be greater than the failure effect, meaning that the demand will increase with an increasing variance caused by *quality differences*.

Bearing these results in mind, if the total variance is constant and the decomposition of the source of variances changes to a higher relative share of variance caused by *taste differences*, then the price will increase. This is because the probability of failure due to *quality differences* will become smaller, and the retailer will have more power to raise the product's or service's prices. If the total variance is sufficiently low, then demand increases with an increasing share of variance caused by *taste differences* (see Figure 7, left side). On the contrary, if the total variance is sufficiently high, the demand will decrease with an increasing share of variance caused by *taste differences* (see Figure 7, right side).

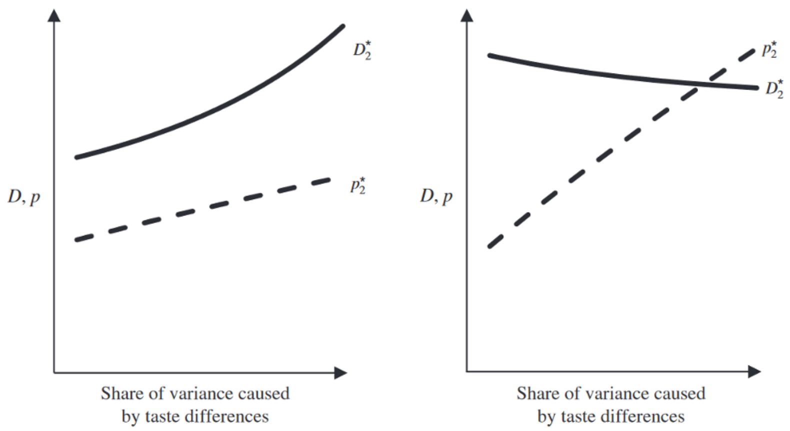

Figure 8: *Optimal price and demand for inconsistent quality goods—Changes in the composition of the variance. (Source: Own illustration)*

Note: For reasons of simplicity, the price and demand are plotted on the same axis. The left graph depicts the case where the total variance is sufficiently low, whereas the right graph depicts the case where the total variance is sufficiently high. p_2^* *represents the optimal price for the product or service in the second period, and* D_2^* *represents the optimal demand in the second period.*

The results of the paper have important managerial implications for the development of an OTF market. If service providers on OTF markets were to consider the composition of the variance of consumer ratings, then they could improve their sales forecasts and increase profits by adjusting their inventories accordingly to satisfy demand or by charging higher prices for those products or services for which a relatively larger share of the variance is caused by *taste differences*. Additionally, service providers on OTF markets could implement mechanisms to explicitly communicate information about the decomposition of the variance to allow more consumers to use this important information in their decision-making.

Selected propositions from the analytical model of this research were also empirically confirmed within Subproject A4 [Gut+18]. Within this study, the author employs a machine learning approach to decompose the variance caused by *quality differences* (i.e., product failure) and *taste differences* for digital cameras on Amazon.com and estimates its impact on the product's price and demand. In line with the predictions of the theoretical model, the author finds that variance caused by *quality differences* is negatively associated with price and demand and that the variance caused by *taste differences* positively affects the product's price and demand. Moreover, the theoretical insights by Zimmermann et al. [ZHKN18] have also informed and inspired marketing research [LBS22] to further extend the theoretical model developed within this paper and establish cumulative knowledge on this topic.

Impact of local market competition on the heterogeneity of available businesses on the market level

However, an unresolved question on the more aggregated market level remained. For instance, it has remains unclear what impact the local market competition has on the heterogeneity of available businesses in a market. Even though prior work has investigated the formation of mean ratings for businesses or the nature of particular reviews [LH08], these studies do not inform about the relationship between competition in a local market and the properties of a market's mean rating distribution. When mean rating distributions change with competition, this means that a business's mean value has to be evaluated differently in markets with the different competition. The relative position in the market of a 3.5-star business will be different in a market of low competition compared to a market with high competition. Therefore, it is crucial to better understand that ratings must be considered within a bigger picture, i.e., taking into consideration boundary conditions in the process of decision-making. The article by Gutt et al. [GHR19] contributes to this notion and provides empirical evidence for this presumption.

While the theoretical literature in industrial organizations has analyzed equilibrium market outcomes for vertically differentiated industries where quality provisioning is primarily tied to marginal costs (e.g., restaurants), the empirical evidence was still scant. Theory suggests that larger markets can support a greater number of firms that cover a larger quality spectrum than smaller markets, such that both tails of the distribution of available qualities grow. So far, the empirical evidence was limited to the growth in the higher end of the quality distribution [BW10]. By studying how local market competition affects the dispersion in both tails of a market's mean rating distribution, Gutt et al. [GHR19] were able to also investigate the dynamics in the lower end of the service quality distribution and evaluate the impact on the average of the distribution.

What we see as an important lesson learned from this work is that the range and average of a market's mean rating distribution may vary depending on the competition level in the particular markets. Here, the difference between the very best and very worst of individual mean ratings in a particular market is described by the range, and the average consists of all individual mean rating valences within that market. To compare different markets, Gutt et al. [GHR19] identified markets (i.e., cities) that do not overlap (i.e., markets which are isolated). The competition level within a market is determined by the total number of businesses. Based on a comprehensive data analysis on a combined data set from *Yelp.com* and *city-data.com*, the authors found that the range of markets' mean rating distributions increases with competition (i.e., number of businesses) but that the average of the markets' mean distributions decreases (see Figure 8). This means a 4.5-star business competes with more 4.5-star businesses in a small market than in a larger market (arrow (a)), whereas a 2.5-star business competes with relatively fewer comparable businesses—in terms of mean ratings-in a small market than in a large one (arrows (b)). With regard to the range of the mean rating distribution, a 4.5-star rating business might be among the upper businesses in a small market but might be considered less elite, being in a larger market due to the wider range on large markets (arrow (c)). While a 2.5-star business might be among the worst businesses in a small market, the worst choices in larger markets have even lower mean ratings (arrow (d)).

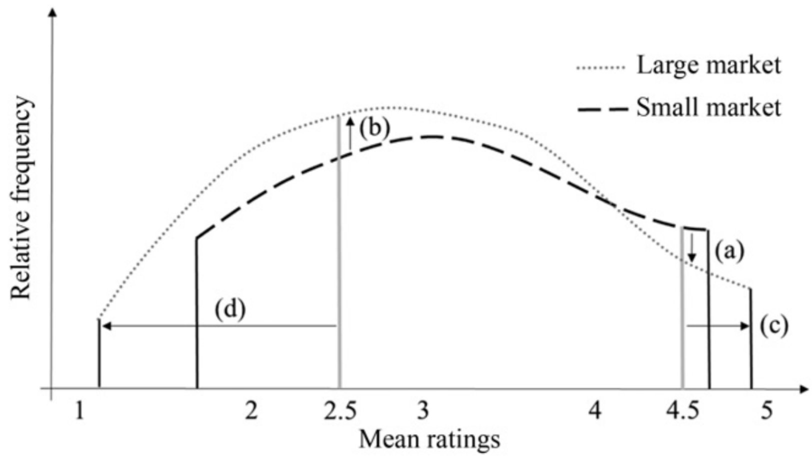

Figure 9: *Competition faced by 2.5- and 4.5-star businesses in two different markets.*

By conducting multiple regression estimations (e.g., cross-sectional estimation, ordinary least squares, and two-stage least square, including instrumental variables, fixed and random effects) on the combined data set from Yelp.com and city-data.com, it was possible to show that local market competition is a driver of the heterogeneity of available mean ratings in a market [GHR19].

One conclusion from these results is that a larger market has proportionately more lower-rated businesses. In contrast, higher-rated businesses have relatively fewer comparable businesses and face less competition in such a market. As market size increases, firms proliferate such that many different quality levels become available, assuming that ratings reflect the quality levels approximately. In these markets, firms compete on the quality of the service they sell through a combination of fixed and variable costs. A larger market can thus support more low-quality firms and firms that offer higher qualities than the highest-quality firm in a small market.

Consequently, as competition increases, so does the range of different qualities available in the market. The change in the average quality of the market level depends on the ratio of low- and high-quality firms entering the market. In other words, it decreases if the increase in competition is mainly driven by low-quality rather than by high-quality firms. Previous empirical studies from industrial organizations have documented that increasing competition leads to an increase in the dispersion in the higher end of the service quality distribution [BW10]. Yet, the effect on the dispersion in the lower end of the distribution has been neglected. Gutt et al. [GHR19] extend this result by finding that the dispersion in the lower end of the service quality distribution exceeds that of the higher end. Moreover, literature on business intelligence and analytics (as suggested by Chen et al. [CCS12]) can build on this evidence—in particular, Yelp's mean ratings form an internally consistent data source for conducting competitive intelligence activities.

Furthermore, the insights of this work lead to the implication for requesters interested in assessing applications with the same functionality (e.g., navigation via online maps) based on electronic word-of-mouth, such as ratings. This means that applications across OTF markets with similar mean ratings should be assessed differently by considering the competition level. Different actors such as requesters (i.e., an entity requesting the creation of an application consisting of components), component providers (i.e., an entity offering access to a component), or infrastructure providers (i.e., an entity offering access to infrastructure for the execution of services) participate in OTF markets. These entities might only have access to a particular set of OTF markets due to restrictions or strategic decisions (e.g., internet censorship or ensuring independence). Moreover, the market provider (i.e., an entity operating and providing access to an OTF market) might decide to restrict access for certain actors. For example, Apple or Google can be seen as a market provider in the case of the App Store and Google Play, where they can decide who is allowed to provide apps and install them. Thus, OTF markets can be seen as isolated markets, for example, geographically or concerning an application domain that is specifically separated. Thus, some entities might exist that act across different OTF markets. Therefore, OTF markets can have different sizes regarding the number of component or infrastructure providers. Requesters might compare applications (i.e., compositions of components) providing this functionality across different OTF markets [KKMW20]. In these cases, the rating distributions for the applications might change with competition within individual OTF markets due to the competition between component providers. Thus, requesters might want to evaluate the mean values of applications differently between OTF markets with different levels of competition. In general, requesters who prefer applications with very high mean ratings will find more suitable matches in larger OTF markets. However, requesters should not randomly choose an application because larger markets are home to disproportionally more applications with low mean ratings than smaller markets. In other words, the broader range of available mean ratings comes at the expense of lower averages of the mean rating distributions in larger markets. This result emphasizes the importance of online review systems and other mechanisms to prescreen applications in large OTF markets to serve requester preferences better.

2.2 Opportunistic Behavior of Service Providers

Strategic decisions of service providers when customer feedback is inaccurate
Mir Djawadi, Fahr, Haake, and Recker [MFHR18] have focused on reputation systems with inaccurate customer feedback. Since customers may not have the expertise to judge the respective service quality correctly, customer feedback in reputation systems is not always accurate. Consequently, sellers may engage in strategic behavior in quality choices, i.e., they may deliver low quality after having achieved a high sales price due to sufficiently many positive ratings.

The research strategy involves a theoretical model and a laboratory experiment. Both components are the result of a collaboration between the Subprojects A3 and A4. In the theoretical analysis, a service provider repeatedly sells a service to short-lived customers, i.e., customers interact only once with the service provider). The rating behavior of the customer is modeled as a random variable, and the service provider's strategic quality choice is modeled as a stochastic optimization problem using a discounted Markovian decision process. The service provider chooses to produce either high- or low-quality service (at high and low costs, respectively) and receives a sales price depending on the current reputation profile. Profit changes are modeled based on reputation profiles by keeping the demand constant for all service providers and changing the price a single service provider may charge. After purchase, a customer either positively or negatively evaluates the service according to predefined error probabilities, reflecting the customer's limited ability to judge the provided service quality correctly.

The objective of the laboratory experiment is to investigate how subjects in the role of service providers make their quality choices and compare observed with optimal behavior derived from the theoretical model. In the experiment, all student subjects play the service provider role and choose between producing a high- or low-quality service (associated with high and low costs, respectively). Four treatments are implemented with exogenous variation in the customers' evaluation abilities to rate the service quality accurately. All other parameters of the theoretical model are held constant. The consideration of exogenous rating behavior depending on different evaluation abilities of the customer and a sales price resulting directly from the reputation profile allows for identifying the reputation system's causal effects on the service provider's quality choice.

Theoretically, a profit-maximizing service provider should always produce high-quality or low-quality services, irrespective of the current quality profile. This means that a good reputation is maintained optimally, or any reputation-building process is wholly neglected. However, according to the results of the experiment, not all subjects follow the optimal strategy in the treatment. Instead, the higher the propensity to choose high quality, the more accurately customers can rate the quality. Moreover, the chosen qualities are conditional on the current reputation profile. More precisely, many subjects use milking strategies and exploit a good reputation. Thus, low quality is delivered if the sales price is high until the price drops below a certain threshold due to negative ratings. High quality is chosen until the price has significantly increased.

2.3 Design of Online Review Systems in OTF Markets

How customers aggregate information to assess product quality
Van Straaten, Melnikov, Hüllermeier, Mir Djawadi, and Fahr [SMH+21] have focused on aggregation patterns of customers on selling platforms. Selling platforms usually provide aggregated measurements of customer feedback in which the numerical part of customer reviews is processed to single index values representing the valence of the product, most often implemented by the arithmetic mean. However, literature in psychology and behavioral economics has identified behavioral biases driven, e.g., by bounded rationality or heuristics (e.g., [TK74]). Therefore, it is unclear whether the technically implemented aggregation functions represent the actual aggregation behavior of customers. Van Straaten et al. [SMH+21] conduct a laboratory experiment to elicit subjects' aggregation patterns and compare these to reference aggregation functions. Thus, they investigate to which degree inherent heuristics of customers affect the aggregation of customer rating distributions and whether they result in systematic biases that should be addressed in the implemented aggregation metrics in reputation systems.

In the experiment, student subjects rank various customer rating distributions to infer the subjects' aggregation heuristics. This allows for finding the optimal aggregation function and comparing it with different alternative aggregation functions. More precisely, subjects receive customer ratings of three products and are asked to rank them according to their preferences. These customer ratings differ concerning their relative frequencies and their arithmetic means. Subjects only see the products' aggregated customer ratings and know that they are from the same product category and have similar prices and specifications, but not the products' names or detailed specifications. Aggregated ratings that allow separating rankings based on different decision heuristics are used, such as minimization of negative ratings or maximization of positive ratings from rankings that favor the arithmetic mean. Subjects choose rankings for a total of 12 categories, whereby distributions are partly artificial and partly based on aggregated customer ratings from the Amazon marketplace. To elicit actual preferences, the entire ranking decision is incentivized: subjects receive a USB flash drive that they rank first or second as payment and have the chance to win another product they choose from one of two other product categories (a tablet computer or tablet holder). They have a higher chance to win the products when they rank a product better and do not know which decision determines their payoff. We implement two treatments investigating the impact of additional numerical information on aggregation heuristics. Subjects are only given the aggregated customer ratings in the control treatment without additional information. In contrast, in the information treatment, subjects also see the relative frequency of each rating category and the arithmetic mean value associated with the distribution. To analyze subjects' rankings, the authors employ a Maximum-Likelihood approach and a Placket-Luce model, which is a model of rank data that is parametrized by quantitative preference degrees for individual choice alternatives.

The results of the experiment show that the arithmetic mean is an appropriate aggregation function as it best explains subjects' average behavior. However, the findings also reveal a tendency to overweight moderate ratings and underweight extreme one, thus indicating a binary bias that is not sufficiently considered in practice. Moreover, minor subject clusters focus on customer rating distributions with the least 1-star ratings or the least negative (i.e., 1- and 2-star) ratings. Contrary to predictions, individual characteristics, risk attitudes, online shopping experience, and additional numerical information do not affect the employed aggregation heuristics. As the findings indicate heterogeneity in aggregation behavior, marketplaces could enhance reputation systems by calculating personalized valence values.

Influence of anonymity and self-expression on the propensity of writing online reviews
Hoyer and van Straaten [HS22] have dealt with anonymity and self-expression in online rating systems. Since providing a customer review costs time and effort without an obvious benefit for the reviewer, drivers of writing a review have been analyzed frequently in the literature. Studies show that the main motives for writing customer reviews are self-expression, altruism, a sense of belonging, and economic incentives (e.g., [CL12]). On the other hand, sharing experiences about products and services might be accompanied by concerns about sharing private information and the threat of being involuntarily analyzed and clustered by third parties. To avoid this privacy issue, anonymous reputation systems can be implemented. However, this anonymity can potentially crowd out reviews that are motivated by the self-expression motive of the customers. To date, the impact of anonymity and its driving forces (such as the blunting of self-expression) on the propensity of customers to write reviews is unclear [RS12]. Therefore, Hoyer and van Straaten [HS22] conduct an experimental study to answer the following research question: Do anonymous reputation systems have a drawback of blunting self-expression as a motive of customer reviews and do altruistic attitudes moderate this effect?

The laboratory experiment consists of a stylized marketplace in which all participants are customers facing computerized sellers. The participants must decide which seller to buy a product from in 10 separate periods. After choosing a seller and learning how satisfied they are with the product, they are given the choice of whether they want to publish their satisfaction with the product, which will be visible to all participants. The two treatments enable to vary the degree of anonymity–from pseudonymity (chosen pseudonym, published when satisfaction is revealed) to anonymity (assigned buyer ID, not published when satisfaction is revealed)–and analyze whether this treatment influences the self-expression motive and hence the propensity of customers to leave reviews. Addressing the self-expression motive, in the pseudonymity treatment, a ranking of the participants' number of published ratings is shown at the end of each period, enabling them to build up a reputation as frequent reviewers. This is not the case in the anonymity treatment, as ratings are anonymous. After the market experiment, a dictator game analyzes participants' attitudes toward altruism. Conducting an economic experiment allows drawing inferences about the effect of anonymity on self-expression with high internal validity as it is controlled for other elements such as prices, product groups, visualization effects, etc.

The results indicate that self-expression is indeed blunted by anonymity. This result is driven by less-altruistic subjects, as more-altruistic subjects are not affected by the blunting of self-expression. In line with the literature on signaling theory, in the experimental setup with a focus on intrinsic motivation and self-expression, excluding the self-expression motive indeed decreases welfare in terms of monetary payoffs. The study thus carries substantial implications, as it identifies the necessity to consider the drawbacks of anonymous reputation systems that must be considered and balanced with the advantages of anonymity (i.e., reduced privacy issues), which might depend on the actual market environment and the products traded on these markets

Connection between social proximity among trading partners and upward-biased ratings

Mir Djawadi and Wester (2022) [MW22] have examined to what extent social proximity between trading partners leads to upward-biased ratings. Peer-to-peer platforms placed under the 'sharing economy', such as Airbnb or Couchsurfing, Uber, or BlaBlaCar, have recently stirred the electronic commerce landscape. Thereby, social interaction between customers and service providers is not only an inherent part, but many people participate for this reason. Due to the intimate nature of the arrangements offered (e.g., homestay or car share), sharing-economy markets depend on trust. Where considerable heterogeneity in service quality and information asymmetries among geographically dispersed and informally connected users exist, credible reputation-based feedback systems become essential to support the emergence of trust. However, many reputation systems have given rise to implausible rating distributions in such a way that ratings cannot be assumed to reflect the actual quality, calling the functionality and effectiveness of reputation systems based on user-generated feedback into question. An examination of the literature suggests diverse reasons for overwhelmingly skewed ratings, among them being non- or under-reporting of negative experiences due to fear of retaliation or social interactions between market participants (e.g., [DW08]). While there are already several suggestions for the design of reputation systems to counteract upward-biased reviews due to fear of retaliation, examining the influence of social interaction on feedback giving has only recently gained momentum.

Through the exchange of personal information, familiarity and liking between individuals are established, uncertainty about others' intents can be reduced, and generalized positive beliefs or impressions may be evoked. Knowledge about one another is found to be linked to a multitude of pro-social behaviors. A well-documented and robust finding is that interpersonal similarity has a reinforcing effect. People perceived to be like themselves are more likely to be regarded as more familiar and socially proximate. Along this line, recent research consistently finds that people are more willing to help, more lenient toward misbehavior, and trust in and feel morally obliged to interact with partners who are socially proximate (e.g., [BWD14]). Common explanations suggest that people derive greater personal satisfaction or utility from helping socially proximate individuals and/or have generalized expectations that those being similar are also more likely to be well-disposed toward them. Hence, Mir Djawadi and Wester (2022) [MW22] propose that feedback giving differs depending on the availability of personal information and social proximity between transaction partners. It is expected that disclosing personal information per se leads to better feedback than having no information at hand. Second, it is hypothesized that socially proximate interaction partners receive better feedback than socially distant interaction partners.

Given the lack of either directly observing social interaction (frequency and intensity) or assessing the degree of social proximity between transaction partners, laboratory experiments can help complement observational data derived from peer-to-peer platforms. Experiments allow the comparison of individuals' feedback-giving behavior with varying degrees of social proximity between transaction partners potentially arising from personal interactions. Therefore, the experiment simulates a sharing economy market framework where a provider offers service and is remunerated and evaluated by a customer taking up the offered service. Since a crucial feature of the sharing economy is the potential for collective action taken by both sides to enhance the experience, the quality of the service is modeled as a function of the joint effort exerted by providers and customers. Whereas these elements are static across treatments, the extent of personal information presented to transaction partners varies.

2.4 Certification to Reveal Service Quality in OTF Markets

The impact of self-declaration and certification on price premiums for experience goods
Fanasch and Frick [FF20] have used the wine industry as an example, as many producers have adopted organic or biodynamic certifications to reveal their production quality. In principle, certifications can be granted for organic and biodynamic practices. However, the few available studies either concentrate on organic practices or consider organic and biodynamic production jointly [AGG17]. However, it is essential to investigate the effect of each practice separately, as organic production is usually considered "serious." In contrast, biodynamic production is often considered "bizarre" or "somewhat strange," partly due to the methods used.

While some of these wineries have successfully applied for third-party certification, others follow strict guidelines without being certified and self-declare themselves to be eco-friendly. Using a large sample of 55,500 wines produced by 1,514 German wineries between 2010 and 2017, this study estimates a series of hedonic price models across different price quantiles. The results indicate a statistically significant price premium for organic and biodynamic certified experience goods, the magnitude of which is, however, far smaller than the effects usually identified in surveys and laboratory experiments. While self-declaration is only a credible signal for organic practices generating a price premium of 8.6 percent, biodynamic practices require certification for a price premium of 4.1 percent. The results also suggest that the interaction of collective reputation and biodynamic practices positively impacts prices.

Overall, this study contributes to current research by closing several research gaps. First, this study compares the effects of self-declaration and certification on product prices by distinguishing between organic and biodynamic practices. Second, this study conjectures that the impact of self-declaration and certification is likely to differ across bottle price distribution, as Abraben et al. [AGG17] suggested. This provides a more detailed understanding of the relationship between prices and self-declared and certified sustainability practices. Third, this paper provides new evidence on whether and to what extent collective reputation impacts the prices that organic and biodynamic wineries can charge. Given the recent changes in consumer preferences, this study uses the latest data available to identify customers' willingness to pay for organic and biodynamic wines.

Thus, this study contributes to information asymmetry and signaling theory by showing that certification costs can be avoided since, in some instances, self-declaration is sufficient to enable a price premium. As a result, certifications are essential for decreasing information asymmetries in markets for composed services. By receiving an external certification, providers can reliably indicate that they meet particular (minimum) standards.

3 Concluding Remarks

The empirical research of Subproject A4 has significantly contributed to the area of online reviews and certification in reducing information asymmetries in OTF markets.

By using econometric analysis of existing markets with comparable characteristics of the OTF market, various solution concepts to signal true service or product quality have been thoroughly investigated. Beyond the benefits of regular customer ratings, research by Cox and Kaimann [CK15] has shown that ratings from professionals or experts can even enhance the customer's confidence in the quality of a service/product which, in turn, can increase demand and reduce market inefficiencies. OTF market providers might therefore consider allowing different types of signals in the reputation system to help customers make the best selection among the offered services. Further research addressed the question whether the seemingly strongest solution concept of certification should be part of a reputation system as well. As Fanasch and Frick [FF20] have shown, there are costless substitutes for certification. Especially if service providers already built a high reputation, certifying their services or processes might only incur unnecessary costs. As costless self-declaration can lead to comparable levels of trust, services can be offered at a lower price which, would not only benefit current customers but may also attract new customers on the respective OTF market.

With regard to online reviews and economic outcomes, Zimmermann et al. [ZHKN18] were able to investigate analytically how different sources of variance of online reviews affect product prices and demand on the level of firms. It led to further research, such as Gutt [Gut+18], who empirically confirmed selected propositions from the analytical model provided by Zimmermann et al. [ZHKN18], or Lee et al. [LBS22], who extended the model and thereby introduced it into marketing research. An implication of Zimmermann et al. [ZHKN18] for service providers on OTF markets is that if they consider the composition of the variance of consumer ratings, then these providers could improve their sales forecasts and increase profits by adjusting their inventories accordingly to satisfy demand or by charging higher prices for those products or services for which a relatively larger share of the variance is caused by taste differences. Additionally, service providers on OTF markets could implement mechanisms to explicitly communicate information about the decomposition of the variance to allow more consumers to use this important information in their decision-making.

The interplay between the design of the market infrastructure (regarding the possibilities of quality assurance) and individual decisions (supply, demand, price, and quality of the traded service) has been systematically investigated. In this context, Gutt et al. [GHR19] were able to investigate empirically what impact the local market competition has on the heterogeneity of available businesses on a market level. Gutt et al. [GHR19] also provided input for further research in highly reputable journals such as the Information

Systems Research (e.g., [AR22]). In the OTF context, an implication of the findings made by Gutt et al. [GHR19] is that OTF market providers might want to consider the competition level within the review system for requesters interested in assessing applications with the same functionality across OTF markets. As those applications with similar mean ratings should be assessed differently by considering the competition level, the market providers might want to add an overview similar to Figure 8, where the mean ratings and their distribution for different OTF markets can be observed or corrected directly by providing adjusted measures depending on the competition levels when comparing between markets.

With the help of economic experiments that have been created in studies like Mir Djawadi et al. [MFHR18], causal for example, analysis of designs of reputation systems or measures (certificates, contract structures, etc.) was systematically conducted that would be otherwise difficult to achieve with field data–especially when features that do not yet exist and have no comparable counterpart in existing markets have to be considered. The results from these experiments have been valuable for revealing behavioral patterns that partly deviate from traditional economic theory and the assumption of rational actors who are only interested in maximizing individual payoffs. For example, the research revealed that a substantial share of service providers milk their good reputation, which represents a non-rational oscillating strategy, an economically irrelevant factor of social proximity seems to influence the rating behavior of market participants, and customers tend to overweight moderate and underweight extreme ratings. This excerpt of empirical findings suggest that theories of Behavioral Economics that relax the assumptions of perfect rationality and selfishness should be considered as alternatives for explaining behavior on OTF markets and designing interventions that prevent market failure. Further, the results of experiments on the acceptance of anonymous online rating systems can be used to develop rating systems that holistically reflect product or service quality and can thus serve to reduce information asymmetries. For example, research on the metrics of online valuations can be used to embed the aggregated valuations intentionally. These suitable rating systems ensure the necessary security and acceptance and, thus, the use of the rating systems. These results can be used in the conceptual development of rating systems as an element of the business model of a market provider in OTF markets.

Bibliography

[AGG17] Abraben, L. A.; Grogan, K. A.; Gao, Z.: Organic price premium or penalty? A comparative market analysis of organic wines from Tuscany. In: *Food policy* 69 (2017), pp. 154–165

[ake78] akerlof, G. A.: The market for "lemons": Quality uncertainty and the market mechanism. In: *Uncertainty in economics*. Elsevier, 1978, pp. 235–251

[AR22] Alyakoob, M.; Rahman, M. S.: Shared prosperity (or lack thereof) in the sharing economy. In: *Information Systems Research* (2022)

[BDT06] Basuroy, S.; Desai, K. K.; Talukdar, D.: An empirical investigation of signaling in the motion picture industry. In: *Journal of marketing research* 43 (2006), no. 2, pp. 287–295

[BW10] Berry, S.; Waldfogel, J.: Product quality and market size. In: *The Journal of Industrial Economics* 58 (2010), no. 1, pp. 1–31

[BWD14] Balliet, D.; Wu, J.; De Dreu, C. K.: Ingroup favoritism in cooperation: a meta-analysis. In: *Psychological bulletin* 140 (2014), no. 6, p. 1556

[CCS12] Chen, H.; Chiang, R. H.; Storey, V. C.: Business intelligence and analytics: From big data to big impact. In: *MIS quarterly* (2012), pp. 1165–1188

[CGV10] Chintagunta, P. K.; Gopinath, S.; Venkataraman, S.: The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. In: *Marketing science* 29 (2010), no. 5, pp. 944–957

[CK15] Cox, J.; Kaimann, D.: How do reviews from professional critics interact with other signals of product quality? Evidence from the video game industry. In: *Journal of Consumer Behaviour* 14 (2015), no. 6, pp. 366–377

[CL12] Cheung, C. M.; Lee, M. K.: What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. In: *Decision support systems* 53 (2012), no. 1, pp. 218–225

[DW08] Dellarocas, C.; Wood, C. A.: The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. In: *Management science* 54 (2008), no. 3, pp. 460–476

[FF20] Fanasch, P.; Frick, B.: The value of signals: Do self-declaration and certification generate price premiums for organic and biodynamic wines? In: *Journal of cleaner production* 249 (2020), p. 119415

[FK17] Frick, B.; Kaimann, D.: The impact of customer reviews and advertisement efforts on the performance of experience goods in electronic markets. In: *Applied Economics Letters* 24 (2017), no. 17, pp. 1237–1240

[FS11] Fahlenbrach, R.; Stulz, R. M.: Bank CEO incentives and the credit crisis. In: *Journal of financial economics* 99 (2011), no. 1, pp. 11–26

[GHR19] Gutt, D.; Herrmann, P.; Rahman, M. S.: Crowd-driven competitive intelligence: Understanding the relationship between local market competition and online rating distributions. In: *Information Systems Research* 30 (2019), no. 3, pp. 980–994

[Gut+18] Gutt, D. et al.: *In the Eye of the Beholder? Empirically Decomposing Different Economic Implications of the Online Rating Variance*. Tech. rep. Paderborn University, Faculty of Business Administration and Economics, 2018

[HMM12] Hennig-Thurau, T.; Marchand, A.; Marx, P.: Can automated group recommender systems help consumers make better choices? In: *Journal of Marketing* 76 (2012), no. 5, pp. 89–109

[Hol79] Holmström, B.: Moral hazard and observability. In: *The Bell journal of economics* (1979), pp. 74–91

[HS22] Hoyer, B.; Straaten, D. van: Anonymity and self-expression in online rating systems—An experimental analysis. In: *Journal of Behavioral and Experimental Economics* 98 (2022), p. 101869

[KKMW20] Karl, H.; Kundisch, D.; Meyer auf der Heide, F.; Wehrheim, H.: A case for a new IT ecosystem: On-The-Fly computing. In: *Business & Information Systems Engineering* 62 (2020), no. 6, pp. 467–481

[KR00] Kirmani, A.; Rao, A. R.: No pain, no gain: A critical review of the literature on signaling unobservable product quality. In: *Journal of marketing* 64 (2000), no. 2, pp. 66–79

[LBS22] Lee, N.; Bollinger, B.; Staelin, R.: Vertical Versus Horizontal Variance in Online Reviews and Their Impact on Demand. In: *Journal of Marketing Research (JMR)* (2022)

[LH08] Li, X.; Hitt, L. M.: Self-selection and information role of online product reviews. In: *Information Systems Research* 19 (2008), no. 4, pp. 456–474

[MDC14] Mayzlin, D.; Dover, Y.; Chevalier, J.: Promotional reviews: An empirical investigation of online review manipulation. In: *American Economic Review* 104 (2014), no. 8, pp. 2421–55

[MFHR18] Mir Djawadi, B.; Fahr, R.; Haake, C.-J.; Recker, S.: Maintaining vs. milking good reputation when customer feedback is inaccurate. In: *Plos one* 13 (2018), no. 11, e0207172

[MW22] Mir Djawadi, B.; Wester, L.: Social Proximity and Feedback-Giving Behavior - A study on how social interaction influences feedback-giving behavior on peer-to-peer platforms. In: *Mimeo* (2022)

[RS12] Rockenbach, B.; Sadrieh, A.: Sharing information. In: *Journal of Economic Behavior & Organization* 81 (2012), no. 2, pp. 689–698

[SMH+21] Straaten, D. van; Melnikov, V.; Hüllermeier, E.; Djawadi, B. M.; Fahr, R., et al.: *Accounting for heuristics in reputation systems: An interdisciplinary approach on aggregation processes*. Tech. rep. Paderborn University, Faculty of Business Administration and Economics, 2021

[Sun12] Sun, M.: How does the variance of product ratings matter? In: *Management Science* 58 (2012), no. 4, pp. 696–707

[TK74] Tversky, A.; Kahneman, D.: Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. In: *science* 185 (1974), no. 4157, pp. 1124–1131

[ZHKN18] Zimmermann, S.; Herrmann, P.; Kundisch, D.; Nault, B. R.: Decomposing the variance of consumer ratings and the impact on price and demand. In: *Information Systems Research* 29 (2018), no. 4, pp. 984–1002