Cicognani, S., Figini, P., & Magnani, M. (2022). Social influence bias in ratings: A field experiment in the hospitality sector. ***Tourism Economics*****,** ***28*****(8), 2197-2218.**

https://doi.org/10.1177/13548166211034645

Social Influence Bias in Ratings:
A Field Experiment in the Hospitality Sector

Simona Cicognani

Utrecht University School of Economics, The Netherlands (s.cicognani@uu.nl)

Paolo Figini*

University of Bologna, Italy (paolo.figini@unibo.it)

Marco Magnani

University of Padua, Italy (marco.magnani@unipd.it)

Abstract

We investigate the empirical phenomenon of rating bubbles, i.e., the presence of a disproportionate number of extremely positive ratings in user-generated content websites. We test whether customers are influenced by prior ratings when evaluating their stay at a hotel through a field experiment that exogenously manipulates information disclosure. Results show the presence of (asymmetric) social influence bias: access to information on prior ratings that are above the average positively influences the consumers' rating of the hotel. In contrast, information on ratings that are below the average does not affect reviewers. Furthermore, customers who have never been to the hotel before the intervention are more susceptible to prior ratings than customers who have repeatedly been to the hotel before. Finally, customers who are not used to writing online reviews are more prone to social influence bias than customers who frequently write online reviews. Our findings suggest that online rating systems should be adjusted to mitigate this bias, especially as these platforms become more relevant and widespread in the hospitality sector.

Keywords: Online Ratings; Field Experiment; Consumer Behaviour; Social Influence Bias; Hospitality Sector

JEL Classification: C93; D83; L86; Z31

* Corresponding author: Paolo Figini, Department of Economics and CAST – Centre for Advanced Studies in Tourism, University of Bologna, Rimini Campus. E-mail: paolo.figini@unibo.it, Tel. +39 051 209 2699; Fax +39 051 209 8143. Piazza Scaravilli 2, 40126 Bologna, Italy. The authors thank Riccardo Tonielli for precious help in data collection, Alessandro Bucciol and participants to the CBTS Workshop, the IMEBESS Workshop, the SEYE workshop, and the seminar at the University of Lugano for their useful comments. The usual disclaimers apply.

Social Influence Bias in Ratings: A Field Experiment in the Hospitality Sector

Introduction

Anything can be rated online: restaurants, hotels, movies, books, flowers and any kind of product exchanged or consumed online or offline. Rating systems allow consumers to express their opinion through binary scales (like/dislike) or Likert-type scales measuring satisfaction (e.g., 1 to 5 points), often including the possibility of writing a short review. They have become more complex over time: nowadays, users can also upload pictures, rate subcategories, share and rate reviews (e.g., Amazon's "Was this review useful?"), and gain quality certifications (e.g., TripAdvisor's "Trip Collective system").

The ubiquity of online ratings has implications for management and marketing, the organisation of markets, and consumer behaviour. In travel and tourism, consumers have increasingly incorporated online reviews into product searches to make better-informed decisions (Nielsen, 2016). The interest of consumers in online reviews has been steadily growing for more than a decade. Nowadays, most customers often or occasionally read online reviews before purchasing (The BrightLocal, 2018) to fasten and improve the decision-making process and narrow down their search (OECD, 2019). Consumers trust this type of information at least as much as recommendations from family or friends because of the perceived lack of commercial self-interest, which might bias instead information from intermediaries or companies (Litvin et al., 2008).

This paper investigates the reasons behind empirical evidence often found in online rating systems: rating bubbles, i.e., the presence of a disproportionate number of extremely positive ratings in user-generated content websites. Among the factors proposed to explain such regularity, Social Influence Bias (SIB), i.e., the tendency to conform to the perceived norm in the community (Muchnik et al., 2013), stands out and represents the core of our study.

We set up a novel experimental design, enabling us to disentangle SIB from purchasing bias and under-reporting bias, other factors that can explain the skewed distribution of ratings. Unlike previous research, which focused on laboratory experiments or observational data, we conduct a field experiment in which we ask subjects (real hotel customers) to rate their experience through an online form. This way, we assess whether and to what extent prior ratings influence the customers' rating activity in an actual purchasing decision.

The novelty of our contribution is fourfold. First, we study how SIB unfolds in the specific case of an experience good (i.e., a product whose quality cannot be assessed before purchase): accommodation. This contrasts with most previous studies in which individual rating behaviour was analysed for search goods/services (whose features can be evaluated before purchase) or non-market items such as political opinions or personal comments (Muchnik et al., 2013). The distinction between search and experience goods/services has been addressed only in terms of the impact of online reviews on product sales (Cui et al., 2010; Ghose & Ipeirotis, 2011) and on the perceived review helpfulness (Mudambi & Schuff, 2010). Specifically, our experimental design allow us to detect any differential rating behaviour between repeat customers and new customers. Repeat customers are customers who have already been to the hotel before the intervention. To them, the hospitality service is arguably a search good, as they have prior and private information stemming from previous stays and on which the review is built. On the contrary, new customers (those who have not been to the hotel before the intervention) face an experience good. Through this experimental set-up, differences in how subjects rate experience and search goods are not deducted by comparing results of different experiments but are directly estimated within the same design.

Second, in contrast to previous lab experiments where experimenters artificially allocated items, we investigate subjects who self-selected into the market, therefore considering the rating activity of real customers. This is relevant since owning a good might lead to a different assessment of it than a situation in which non-owners are asked to evaluate the same good (Ong et al., 2015). Thus, our research is the first one ever conducted through a field experiment on the presence of SIB in online ratings for a real purchasing experience.

Third, our design allows us to disentangle the relative importance of SIB compared to other factors in explaining rating bubbles: purchasing bias – i.e., consumers who purchase the product hold a more favourable opinion of it; under-reporting bias - i.e., only consumers who hold strong favourable or unfavourable opinions engage in the activity of reviewing the product. Through the field experiment, we can also engage with customers who generally do not write reviews online, hence detecting any differential behaviour (for instance, in the average rating score or the susceptibility to previous reviews) compared to habitual reviewers. To the best of our knowledge, only Schlosser (2005) looked at this difference, but in an artificial laboratory set-up and at the outset of the online reviews' existence. In addition to Schlosser, we specifically look at the interaction between SIB and the condition of being a frequent reviewer, shedding light on whether rating bubbles are mainly driven by SIB or under-reporting bias.

Fourth, we contribute to a better knowledge of how the different backgrounds of customers moderate SIB in terms of review attitude and product/service adoption.

Methodologically, the use of the field experiment is key in investigating a real purchasing experience and entails a high external validity of our study (Viglia and Dolnicar, 2020). While the presence of SIB has been detected in contexts where the rated object was not purchased by subjects but randomly assigned by the experimenter (Schlosser, 2005; Krishnan et al., 2014), the topic of self-selection into the market deserves proper attention, as the behaviour of real consumers might differ from the one in the lab: it is essential that consumers feel attached to what they are asked to rate for the rating experience to be salient and, consequently, genuine preferences to be revealed (Morales et al., 2017). Moreover, the investigation of self-selection through observational data of online platforms excludes, by definition, non-reviewers (Li et al. 2019). Hence, a field experiment is the best approach to jointly consider actual purchased products and customers who currently do not review online.

Our work contributes to research but is also relevant for practitioners. SIB determines the extent to which policies aimed at soliciting reviews from previous customers (e.g., through customer care departments) are effective. As the marketing policy of asking customers to rate online is becoming ubiquitous and involves new and naïve reviewers, any differential impact of SIB on these customers might become of paramount importance in the future (Donaker et al., 2019). Moreover, SIB would amplify the impact of fake reviews, leading to a cumulation of biased opinions. Finally, the presence of SIB would encourage firms to allocate funds to counteract it when it is at play, a suboptimal equilibrium compared to a bias-free benchmark.

The paper is structured as follows: the next section reviews the relevant stream of literature. Then, the theoretical reasoning on which the main hypotheses are derived is presented; next, we describe the experimental design and the procedures. We then present the main results. The last section discusses the findings and offers concluding remarks.

Literature Review

Consumers typically incorporate online reviews during the product adoption process. Heuristic cues, such as average scores (the "valence" of reviews), are interpreted as quality signals to the point that many consumers only choose businesses whose average score is above a certain threshold (The BrightLocal, 2018). There is extensive literature associating higher ratings with higher sales in many sectors, including food and hospitality (Book et al., 2016; Kim et al., 2011; Lewis and Zervas, 2019; Öğüt and Onur Taş, 2012; Phillips et al., 2017; Ye et al., 2011).

The significant impact of online reviews on sales has called for a specific investigation on why rating distributions tend to be skewed towards extreme values. This empirical regularity, often referred to as J-shaped rating distributions or rating bubbles, has been found for experience goods (Chevalier & Mayzlin, 2006; Hu et al., 2009; Luca & Zervas, 2016) and search goods alike (Melián-González et al., 2013; Lafky, 2014). Schoenmueller et al. (2020) showed that extreme ratings are common across various online platforms (e.g., Airbnb, Booking.com and TripAdvisor). As online reviews are relevant for product adoption and are trusted by consumers (Nielsen, 2016), should rating distributions be biased (there would be, hence, rating bubbles), inefficient outcomes might arise (De Langhe et al., 2016).

Three main reasons have been proposed in the literature to explain the J-shaped distribution of ratings: (i) purchasing bias: those writing a review are the ones who purchased the product. Hence they have more benevolent opinions (Hu et al., 2009); (ii) under-reporting bias: only consumers who hold strong positive or negative opinions engage in reviewing the product (*brag-or-moan* effect) (Schoenmueller et al., 2020); (iii) social and observational learning: individual rating behaviour is affected by ratings of other consumers (Muchnik et al., 2013). While (i) and (ii) refer to personal decisions of the subject, (iii) is also related to external factors, as virtually all major rating websites and rating interfaces within e-commerce websites present a summary of prior ratings (on top of other information such as the number of ratings and, occasionally, the rating distribution) before or during the review process (e.g., TripAdvisor, Yelp, Facebook, Google Reviews). In addition, solicited e-mails can also be a vehicle of information about prior ratings (Litvin & Sobel, 2018). The pervasiveness of information on prior ratings (King et al., 2014) could hence explain rating bubbles and skewed online rating distributions: although consumers rate products and services after they experienced them, and hence prior ratings should lose their role as informative signals, their effect might persist, in a typical herding fashion (Cosley et al., 2003; Çelen & Kariv, 2004; Aral, 2014).

Research on post-purchase and rating behaviour is quite limited (Magnani, 2020). Schlosser (2005) suggested the presence of a *self-presentation concern* in a laboratory setting: subjects exposed to previous negative reviews reviewed more harshly than those exposed to no reviews or positive ones because they perceived negative opinions as coming from experts and more objective people. Hu et al. (2009) conducted a controlled experiment in which subjects had to rate randomly selected products characterised by a J-shaped rating distribution on Amazon. Results showed that controlled ratings were distributed according to a unimodal distribution, with a relative majority of middle-scale scores; the authors suggested that this was due to both purchasing and under-reporting biases. Lafky (2014) further explored the latter effect and found that ratings can become signals to channel altruism towards subsequent buyers and sellers.

SIB and herding effects are the focus of a study by Lee et al. (2015), who used data of movie ratings on a popular movie social network finding that friends' ratings always affected one's ratings, while ratings from the larger cohort of respondents (the "crowd") had an effect only when the movie was very popular. Krishnan et al. (2014) also found SIB on a political platform. When consumers follow the crowd in the rating activity, distributions are formed through the cumulation of biased preferences, representing a distorted quality signal to future consumers and firms alike (Muchnik et al., 2013). The problem becomes much more relevant if the rating bubble originates from fraudulent reviews, a crucial issue for many platforms (TripAdvisor, 2019) and a hot topic of research (Luca & Zervas, 2016; Wu et al., 2020).

A summary of the literature cited in this section is presented in Table 1, recalling the conditions of each study (platform and products investigated) and a brief outline of the main results.

[TABLE 1 ABOUT HERE]

Theoretical Reasoning and Hypotheses

Reviews are based on subjective perceptions, which can be biased if individuals are highly susceptible to external cues (Cosley et al., 2003); in this context, early adopters of a product or influential opinion leaders are pivotal in spreading information and behaviours (Rogers, 1985). This process can generate SIB, with informational cascades and bias in quality perception where *"aggregate collective judgment and socialised choice could be easily manipulated, with dramatic consequences for our markets, our politics, our health"* (Muchnik et al., 2013, p. 647). The introduction of rating systems, where opinions of prior consumers are easily accessible and widely spread, has possibly amplified social influence on individual behaviour (Chevalier & Mayzlin, 2006; Muchnik et al., 2013; Krishnan et al., 2014). The empirical evaluation of SIB is not an easy task, and although herding behaviour has been observed in many contexts (Anderson & Holt, 1997; Çelen & Kariv, 2004; Muchnik et al., 2013; Krishnan et al., 2014), the extent to which individual rating can be affected by prior ratings did not receive sufficient scrutiny, especially in the tourism sector.
As previously recalled, SIB has been observed in laboratory settings (Schlosser, 2005), where subjects were asked to rate products selected by the experimenter, or with observational data, in platforms like TripAdvisor (Han and Anderson, 2018), without even knowing whether the subjects had purchased and consumed or experienced the product/service. It is essential to intercept consumers during their actual activity and post-purchase evaluation, which is feasible through a field experiment, where subjects who are generally inactive in online platforms can also be approached (Viglia and Dolnicar, 2020). This is important for hotel accommodation, a typical experience good. Unlike choices regarding standardised products or opinions, booking a room entails high risk, since in most cases, the specific service has not been experienced beforehand. In this situation, consumers might rely more heavily on different forms of word-of-mouth to obtain sufficient information on the quality and reduce the level of uncertainty (Liu & Park, 2015). In this framework, our first hypothesis is that SIB extends to experience goods, being present in hospitality rating systems, similar to what occurs for search goods (Moe and Schweidel, 2012):

Hypothesis 1.

Consumer rating activity in the hospitality industry is affected by prior ratings.

The effect of prior ratings might be asymmetric, as the polarity of rating distributions might suggest (Schoenmeuller et al., 2020). According to Muchnik et al. (2013), when rating a movie, positive social influence accumulates and creates a tendency towards rating bubbles, whereas negative social influence inspires users to correct such negative ratings (Coker, 2012). Schlosser (2005) also found that less positive opinions lead to milder downward adjustments. Indeed, when the baseline opinion is already favourable, we expect prior positive ratings to reinforce consumer attitude and positively affect ratings more than the negative reinforcement triggered by prior negative ratings. Consistently, we test the asymmetry of SIB in rating:

Hypothesis 2.

Social Influence Bias is asymmetric; excellent (i.e., above-average) prior ratings have a stronger effect on rating activity than low (i.e., below-average) prior ratings.

A novel aspect of this work is the assessment of SIB when comparing an experience with a search good. Accordingly, a relevant dimension through which the customer base is segmented concerns repeat purchasing activity (Mittal et al., 1998; Frank et al., 2014); thus, repeat visitors (customers who previously stayed in the same hotel during a different visit) can be compared with non-repeat visitors (customers who stayed in the hotel for the first time during the experiment). Consistently with previous seminal works (Stigler, 1961; Nelson, 1970), we argue that repeat customers have already internalised their search costs, or, put differently, having purchased the service previously, they face a search good. In the post-experience phase, this is important because repeat visitors have more private information about previous stays on which to build their evaluation. Further, satisfaction with hotel services positively affects the likelihood of returning (Choi and Chu, 2001). In contrast, the service represents an experience good for new customers, and they might be more susceptible to heuristic cues such as prior ratings available on online systems (Filieri and McLeay, 2014). Hence:

Hypothesis 3.

Repeat customers are less susceptible to Social Influence Bias stemming from prior ratings than new customers.

Most of the previous studies build upon the vast availability of online and networked population-based data sets (Muchnik et al., 2013, Li et al., 2019). Our study is different because the experiment was conducted online, but prospective reviewers were contacted offline. This way, we could investigate the behaviour of customers who were not accustomed to reading, writing, or posting reviews online (non-reviewers). With the integration of online reviews on social networks, it becomes important to understand whether those who are not accustomed to writing reviews (but could start doing so) might carry a different attitude (and behavioural biases) than expert reviewers.

On this topic, we could test whether rating bubbles also stem from under-reporting bias, that is, the fact that those who choose to post a review have more extreme opinions, as suggested by Hu et al.

(2009). If under-reporting bias exists (due to the inactivity of customers with moderate opinions), we expect that the average rating of non-reviewers is lower than the one of reviewers:

Hypothesis 4.

Non-reviewers' rating presents a lower average score than the rating of frequent reviewers.

Finally, we posit that, even though they might have more moderate views, as postulated in the previous Hypothesis, non-reviewers are more susceptible to SIB; this is in line with previous research (Moe and Schweidel, 2012), where less active reviewers were found to imitate prior reviewers more than frequent ones. Observational learning drives consumers to get accustomed to online review systems and recognise their limitations and possible biases. Hence, the rating behaviour of frequent reviewers is expected to be less susceptible than the rating of non-reviewers; indeed, previous research has shown that review experience (in terms of the number of reviews posted online) moderates the impact of prior ratings on individual rating behaviour (Ma et al., 2013). On the contrary, non-reviewers, being unfamiliar with online ratings, are more insecure about the environment and prone to internalise external information (such as prior ratings) during the rating process (Foster, 2005; Filieri and McLeay, 2014). Hence:

Hypothesis 5.

Non-reviewers are more susceptible to Social Influence Bias stemming from prior ratings than frequent reviewers.

Hypotheses 1 and 2 test the robustness of previous findings, extending evidence to the case of the hospitality industry and when real purchasing decisions are considered. Hypotheses 3 to 5 shed light on the interaction between SIB and being a repeat customer or a frequent reviewer, issues that have received scant attention so far in the literature, especially when evaluating real purchasing decisions.

Experimental Design and Procedures

The goal of the field experiment was to investigate the consequences of being exposed to different information sets (i.e., data about ratings of prior consumers) when rating the stay at a hotel. The experiment was conducted in August and September 2015 in a 3-star superior hotel located in the Riviera of Rimini, an important Italian seaside destination. It consisted of an online questionnaire (developed using Google Forms) that customers could access through a URL listed in a flyer (displayed in the Appendix, Figure A1) handled by the hotel manager when they were checking out. This way, we reached all hotel customers during the experiment period since they had to proceed through the checkout. Customers were informed that they could fill in the questionnaire within two weeks using their computers, smartphones, or tablets.

The questionnaire (reported in the Appendix) included three parts: in the first part, customers were asked to rate the hotel, as well as its characteristics, on a 5-point scale. This part was aimed at mimicking the rating scale used by popular rating platforms. The second part included socio-demographic questions, while the third part included questions about customers' previous experience in the same hotel and destination and their attitude towards online reviews. The experiment did not offer any reward based on performance since subjects simply expressed their opinion about the service. Furthermore, it is worth highlighting that the focus of our study was the presence of SIB in online ratings and not the average rating or the rating distribution per se. Therefore, even if external factors (confounding variables) might be at play in how people rated their stay at the hotel, we can exclude that these factors biased the SIB results since external factors were orthogonal to treatments. What is relevant for detecting SIB was the difference in average ratings across treatments (Viglia and Dolnicar, 2020).

The experiment included three treatments, differing for the disclosed information set; the questionnaires were the same, except for an intervention sentence written above the overall rating question and informing subjects on previous customers' ratings (see Figure 1). In the control treatment, customers were asked to rate the property without any information about prior ratings. In the 3-point treatment, subjects were informed that at least 17 prior hotel customers had rated 3 on a 5-point scale. In the 5-point treatment, the questionnaire disclosed that at least 17 previous customers had rated 5 on a 5-point scale. 17 was chosen as the reference number for both treatments to avoid deception (i.e., conveying false information to experimental subjects) since, at the time of the experiment, the hotel had an average score of 4.5/5 across 232 reviews on TripAdvisor and, among these reviews, 17 rated the hotel 3/5, and 130 rated it 5/5. Given the high average rating of the hotel, we pooled together all ratings equal to or lower than three as "low". The choice to disclose only the absolute number of ratings, without referencing the total number of reviews or the total distribution of ratings, was central to the aim of the experiment: we did not want to provide information on the rating distribution to test how a simple normative message could be read differently across treatments.

[FIGURE 1 ABOUT HERE]

Three main lines of reasoning drove the choice of 3/5 and 5/5 as reference ratings in our treatments: i) the great majority of scores in online rating systems (including this hotel) range between 4/5 and 5/5, and scores of 1, 2, or 3 are relatively uncommon (Schoenmeuller et al., 2020); therefore, a hypothetical 1-point treatment might have been interpreted as suspicious by subjects; ii) the 1-point treatment would have been impossible to implement without engaging in deception because of the absence of 1/5 ratings on the hotel's TripAdvisor page at the time of the experiment; iii) given the TripAdvisor rating of the hotel, 3-point and 5-point treatments allowed us to preserve symmetry in treatments: in fact, the symmetry concerns the mean/median of the score distribution rather than the absolute values.

Regarding the choice of treatments, we were inspired by Schlosser (2005), but unlike that study, we exposed subjects to ratings rather than textual reviews since this is the most common and comprehensive criterion for comparing service quality on online rating platforms (Filieri & McLeay, 2014; Gupta & Harris, 2010). Moreover, we used a fruitful background provided by Viglia et al. (2014).

Assignment to each of the treatments was randomised: flyers containing the URLs related to the different treatments were handled by the hotel manager in sequence at the time of checkout, independently of customers' characteristics. Out of 400 flyers distributed, 75 questionnaires were completed, which entails a response rate of 19%, in line with the typical response rates of online surveys (Daikeler et al., 2019; Kaplowitz et al., 2004). After checking for the integrity of the flyers' codes and removing invalid questionnaires, 67 observations were used in the analysis: 21 for the control treatment; 22 and 24 for the 5-point and 3-point treatments.

It is important to highlight three aspects to reject a possible non-response bias criticism. First, our sample's socio-demographic characteristics represent the hotel's tourist population and the destination where the experiment took place. According to Istat (2005) and Brau et al. (2009), the only two studies that, to the best of our knowledge, encompass socio-demographic characteristics of summer tourists in Rimini, the sample well represents this popular mass tourism destination (see Table 3, last column).

Second, our aim was not to describe the online behaviour of this population, as most customers do not write reviews online. Anderson and Simester (2014) found that only 1.5% of the customers of the investigated firm had written a review. In our case, since reviewers were 48% of the sample, with a response rate of 19%, the overall rate was around 9%. Considering that the share of reviewers can be higher in hospitality and tourism than what found by Anderson and Simester (2014), we likely caught a sizable number of reviewers who stayed in the hotel in the period under investigation. Moreover, we were able to engage with a similar number of non-reviewers (52% of our sample), representing the silent majority not writing reviews in the real world. This is important, not only because it allowed the comparison of behaviours but also because these are likely the subjects that in the future might become reviewers, thus unfolding managerial implications when comparing findings.

Third, a non-response bias problem may arise should the behaviour of non-respondents differ from that of respondents. In this respect, we assume that those not responding were like the many non-reviewers intercepted by the experiment. Since these subjects were influenced by prior ratings (Result 5 in the Empirical Analysis Section), we are confident that results were not biased by non-respondents (in fact, their presence might have reinforced the SIB result).

It is worth recalling that the questionnaire was built in such a way as to mimic popular rating websites. The introductory sentences and the presentation of ratings (e.g., the terrible/excellent scale) were drawn based on a standard TripAdvisor rating page at the time of the experiment. Finally, no prior textual reviews were shown during treatments. While recognising the relevance of textual reviews, we preferred a simple design to assess SIB through rating scores, the most common heuristics on rating platforms (Forman et al., 2008).

Empirical Analysis

A detailed description of the variables is provided in Table 2, while descriptive statistics of the respondents' characteristics are displayed in Table 3. The average respondent was about 46-year-old, male, Italian, with upper-secondary education. He mainly travelled with a partner or with family and spent 7–10 days at the hotel. Customers were well acquainted with the destination and the hotel: on average, they had visited Rimini five times and the hotel four times. Overall, 73% of the subjects declared that they had been to the hotel at least once before the experiment, meaning they were "Repeat customers" to the hotel.

[TABLE 2 AND 3 ABOUT HERE]

As for their rating behaviour, 52% of the respondents had never written an online review. Moreover, one third had read reviews about the hotel before booking their stay, and 75% of them declared that they had been affected by the reviews they had read. Table 4 reports the average rating score given by the respondents for the hotel (4.61) and its characteristics. Table 4 shows that the average rating was very close to the outcome on TripAdvisor at the time of the experiment (4.50).

[TABLE 4 ABOUT HERE]

We then moved to test the five hypotheses and estimate the impact of treatments on the rating attitude through non-parametric tests and regression analyses, controlling for a set of subjects' characteristics. A first hint of whether rating behaviour differed across treatments is in Table 5. The average rating score was lowest in the 3-point treatment and highest in the 5-point treatment, with the control treatment lying between the two. Thus, evidence reported in Table 5 suggests that being exposed to information concerning prior ratings may have changed subjects' ratings, in line with Hypothesis 1. This is corroborated by Figure 2, where the rating density of the 5-point treatment is strikingly more skewed towards 5/5 than the 3-point and control treatments.

[TABLE 5 ABOUT HERE]

[FIGURE 2 ABOUT HERE]

This *prima facie* evidence was supported by non-parametric testing.[1] When comparing 5-point and control treatments, the difference in rating was statistically significant at the 1% level ($p = 0.007$, two-sample Wilcoxon rank-sum test, $n_1 = 22$ and $n_2 = 21$, two-sided). Similarly, the difference between the 3-point and 5-point treatments was statistically significant at the 1% level ($p = 0.002$,

[1] As reported in Fagerland (2012), non-parametric tests should be used when the sample size is small and there is evidence of non-normality in the distribution of the variable of interest. Both conditions apply to our study (however t-test results are very similar). Non-parametric testing is commonly used in behavioural science for these reasons (Siegel and Castellan, 1988; Kraska-Miller, 2013).

two-sample Wilcoxon rank-sum test, $n_1 = 24$ and $n_2 = 22$, two-sided). These tests allowed us to validate Hypothesis 1:

Result 1. *Consumers' rating is affected by information on prior ratings.*

We then analysed the asymmetry of the influence of prior ratings (Hypothesis 2). We tested whether excellent prior ratings had a different impact than moderate ratings: the 5-point treatment was expected to yield a stronger SIB effect than the 3-point treatment under Hypothesis 2 (as the control treatment distribution tends towards more positive values itself). We compared the average rating in 3-point vs control and 5-point vs control. Only the latter comparison was statistically significant ($p = 0.007$, two-sample Wilcoxon rank-sum test, $n_1 = 22$ and $n_2 = 21$, two-sided), whereas the former was not ($p = 0.520$, two-sample Wilcoxon rank-sum test, $n_1 = 24$ and $n_2 = 21$, two-sided). This result allowed us to validate Hypothesis 2:

Result 2. *Social Influence Bias is asymmetric: being exposed to excellent (above-average) prior ratings generates a significant positive bias in ratings not mirrored by a significant bias when exposed to low or moderate (below-average) prior ratings.*

Hence, we found evidence that rating bubbles stem from asymmetric herding even in the hospitality sector. To corroborate Results 1 and 2, we estimated Models 1 and 2 (Table 6) using two alternative dependent variables, respectively: a dummy variable called *Excellent rating*, which took the value 1 if the rating was 5, and 0 otherwise; and *Overall rating*, which is a categorical variable with three values: equal to either 1, 2, or 3; equal to 4; and equal to 5. The rationale for this partition of *Overall rating* stemmed from the small number of observations with ratings below 4: to have an efficient estimation, alternatives that are rarely chosen must be aggregated (Cameron & Trivedi, 2013). Results, however, were also robust to the use of the full-scale overall rating as a dependent variable (not shown for brevity). We estimated through maximum likelihood (Model 1 is a logit, while Model 2 is an ordered logit).

[TABLE 6 ABOUT HERE]

Models 1 and 2 in Table 6 include the following control variables: gender, age, stay period, being a new customer at the hotel, having read a review of the hotel before the stay, and being a reviewer.

The positive and significant coefficient of the 5-point treatment in Table 6, Model 1, confirmed that being exposed to information about excellent prior ratings is associated with a higher rating score.

In contrast, the statistical insignificance of the coefficient of the 3-point treatment showed that being exposed to information about low prior ratings does not significantly affect the rating. The ordered logit model shown in Table 6, Model 2, with *Overall rating* as the dependent variable, confirmed the results robustness: the 5-point coefficient was positive and statistically significant; the 3-point coefficient was not statistically significant, thus confirming the asymmetry of SIB. Results from OLS regressions (not shown for brevity) entirely confirm the results of Table 6 and can provide an estimate of the marginal effect: by falling into the 5-point treatment, the probability of giving a 5/5 score increased by about 32%.

As for repeat customers, the behavioural prediction was that they would be less likely to be influenced by prior ratings than new customers since the former had more established and sound private information about the hotel's characteristics than the latter (Hypothesis 3). We then ran an ordered logit model[2] with *New customer* (*Repeat customer*) being a dummy variable, taking the value 1 if the subject had never (had already) stayed at the hotel before the experiment and with four interaction terms between *New customer*/*Repeat customer* and *5-point/3-point* treatments. Results are shown in Table 6, Model 3.

This moderation analysis shows that repeat customers did not rate differently than new customers, as the coefficient of *Repeat customer* was not statistically significant. In contrast, the interaction terms were statistically significant. Both types of customers were influenced by the 5-point treatment, although repeat customers to a lesser extent than new customers. It is crucial to notice that new customers were also (negatively) influenced by the 3-point treatment, differently from repeat customers. Table 7 reports the predicted probabilities of rating 5/5 in the 5-point treatment versus the other groups across this customer's characteristic (3-point and control treatments have been aggregated for the sake of the comparison since previous non-parametric testing and regression analysis showed that the two groups were not statistically different). The joint reading of Table 6, Model 3, and Table 7 shows that new customers were more susceptible to SIB because the difference in predicted probabilities was more pronounced for them than for repeat customers: new customers have a stronger SIB when exposed to low prior ratings. This result led us to confirm Hypothesis 3:

Result 3. *New customers are more susceptible to Social Influence Bias than repeat customers.*

[2] The ordered logit was chosen over the logit model since it holds more information; however, results were robust to the use of the logit model with *Excellent* as a dependent variable. We did not conduct non-parametric testing for this moderation analysis because of the low numerosity of some subsegments of the subject pool.

As for the comparison of reviewers and non-reviewers, understanding whether the ratings of these two groups differ would help shed light on one of the proposed explanations for rating bubbles: under-reporting bias, which states that those who write online reviews have more extreme (and typically more positive) preferences than those who do not bother to express and share their opinion. Should the behaviour of the two types of subjects not be statistically different, this type of bias would be excluded from the reasons driving J-shaped rating distributions on UGC platforms. The subject pool was almost equally divided between non-reviewers (52% of the sample) and reviewers (48%). The average overall rating for non-reviewers was 4.59/5, versus 4.63/5 for reviewers, a difference that was not statistically significant (p = 0.557, two-sample Wilcoxon rank-sum tests, $n_1 = 35$ and $n_2 = 32$, two-sided).

[TABLE 7 AND 8 ABOUT HERE]

Regression analysis s presented in Table 6, Model 4, where the dummy variable *Not reviewer (Reviewer)*, which takes the value 1 if the subject had never (already) posted a rating online, was included in the model. Four interaction terms between *Not reviewer*/*Reviewer* and *5-point/3-point* treatments were also included. Results show that having experience in posting online reviews did not exert per se a significant effect on the rating behaviour of subjects: the coefficient of *Reviewer* was not statistically significant. This result suggests that the rating of non-reviewers and reviewers did not differ per se, leading us to the important result of rejecting Hypothesis 4:

Result 4. *Social Influence Bias is not confounded by under-reporting bias; non-reviewers' and reviewers' average rating scores do not differ.*

However, we noted that the coefficient of *Reviewer* was not statistically significant in the 5-point treatment when interacting with the treatment variables, while it was only weakly significant in the 3-point treatment. The opposite result was found for *Not reviewer*, suggesting that non-reviewers were affected when exposed to excellent prior ratings but not to low prior ratings. Table 8 reports the predicted probabilities of rating 5/5 in the 5-point treatment compared to the other groups across reviewers' characteristics. The joint reading of Table 6, Model 4, and Table 8 shows that non-reviewers were more susceptible to SIB because the difference in predicted probabilities was more pronounced for them than for frequent reviewers. It also suggests that non-reviewers were more easily influenced by the 5-point treatment than reviewers. This result leads us to report the importance of SIB, particularly for new reviewers, and to confirm Hypothesis 5:

Result 5. *Non-reviewers are more susceptible to information about excellent prior ratings than reviewers.*

Overall, the participation in the experiment of subjects who never wrote on online rating systems worked to exacerbate the overall degree of SIB in the study. This has interesting managerial implications for rating platforms, which will be discussed in the concluding section.

Discussion and Conclusions

This study fits into a recent stream of research addressing the reliability of reviews and assessing biases in rating scores. We designed a field experiment aimed at three goals: one, assessing if and how prior ratings influence individual rating behaviour in the case of hotel accommodation, a typical experience good; two, proposing a novel approach to identify SIB and separate it from under-reporting bias, through the exposition of non-reviewers to the rating system and the evaluation of their activity; three, disentangling and comparing the rating of repeat customers with that of first-time visitors, this way testing whether SIB plays a different role depending on the service falling more in the realm of experience rather than search goods.

A relevant contribution of this paper pertains to the identification and separation of SIB from under-reporting bias, possible because of the investigation of subjects who were not active on review platforms. In this regard, we discuss two relevant findings: first, under-reporting bias does not contribute to rating bubbles. In fact, average rating scores of reviewers and non-reviewers were not statistically different, in line with a recent study by Smironva et al. (2020) and contrary to what was suggested, among others, by Hu et al. (2009) and Lafky (2014). Second, reviewers and non-reviewers strongly differ in their susceptibility to SIB: information about excellent prior ratings specifically affected subjects who were not accustomed to posting ratings on online platforms. This result is in line with Moe and Schweidel (2012), who found that less frequent posters exhibit bandwagon behaviour and opens important implications for platforms. The joint reading of these two results suggests that, when real purchasing decisions are considered, rating bubbles are mainly determined by SIB, not by under-reporting bias. Hence bubbles could amplify in the future, because of the growing relevance and pervasiveness of online review platforms, due to their ability to attract new and naïve customers.

A second contribution concerns the characteristics of the experimental subjects, who had self-selected in the service market to be rated, this way overcoming important caveats of existing literature. Previous studies either assessed real rating attitudes through observation of data on online platforms (this way excluding subjects who were not willing to use or were not at ease with these systems) or set up laboratory experiments where subjects were asked to rate items randomly and exogenously assigned by experimenters (this way abstracting from real purchase decisions). In this regard, it is essential to state that our findings are in line with most of the related literature (Schlosser, 2005; Sridhar and Srinivasan, 2012) and, although based on a single case study, confirm that SIB is a relevant issue in rating systems also when controlling for self-selection and real purchase decisions. We also confirm an asymmetric effect in SIB, consistently with Muchnik et al. (2013): subjects tend to herd to the display of information about excellent ratings (higher than the average), but they are not significantly influenced by information about low ratings (lower than the average).

A third contribution sheds evidence, for the first time clearly and coherently, on how SIB is moderated by the nature of the good, either an experience or a search good. Theoretically, SIB is expected to play a stronger role for experience goods. Empirically, previous experiments were either considering experience (Schlosser, 2005; Lee et al., 2015) or search goods (Hu et al., 2009) and, with different designs and control conditions, a comparison of findings was challenging. The customer base segmentation investigated in this experiment (repeat/first-time visitors) is instead associated with the different perception of the product as a search or an experience good, thus allowing to say a word on this moderation. The finding that new customers were more heavily affected by prior ratings than repeat customers suggests that SIB is particularly relevant when experience goods are rated. This difference is in line with theoretical expectations: repeat visitors have more private information (stemming from their previous stays at the hotel) to build their evaluation than first-time visitors. On the contrary, new visitors with no previous experience of the hotel are more susceptible to the opinions of previous customers and are hence more subject to herding behaviours.

Our results have implications for online platforms and hotel operators. As regards platforms, we ring a warning bell about the consequences of those attempts to expand the volume of ratings by attracting new reviewers, a collateral effect of the growing integration of rating platforms with social networks. Since SIB is particularly relevant for new customers and inexperienced reviewers, SIB will likely play an increasing role in the future, thus jeopardising the informative role that review systems have recently gained. Actionable prescriptions for reforming the functioning of online rating systems go in two directions.

One, interfaces should be designed to provide users with as little information as possible about prior ratings when asked to review. The average score and other hints that might influence reviewers should be hidden. In contrast, TripAdvisor and Google currently show the number, the content, and the score of recent reviews on the same page where users are asked to rate the service. Hiding information might be ineffective, as users can always open different browsers to find it. However, the mechanism used by Booking.com (where the review activity starts from a solicitation email and is not connected to the review page of the hotel) is probably enough to discourage most users from checking previous ratings. This action partially contrasts with the emerging literature about UGC enjoyment (Park & Nicolau, 2015), which calls for rich layouts and very informative websites. The trade-off between the amount of information minimising rating distortions and increasing content enjoyment is a future challenge for both research and website design.

Two, experimentation with machine-learning algorithms to estimate and automatically correct biases might partially correct SIB. This is something that TripAdvisor is probably implementing in its algorithm, as the average score considers the quantity, the quality, and the recency of reviews. Similarly, the implementation of time windows or thresholds on the minimum number of reviews, under which scores are collected but not shown yet to other customers, might lead to a less biased distribution of ratings before becoming public. Time windows when data are collected but not posted can be implemented when a new service or product is rated and throughout its online cycle: for example, Booking.com currently shows only less than two years old reviews.

Such reforms often contrast with the needs of hospitality management and, more in general, of service and products providers. Asking customers to rate online is a popular marketing strategy, and it is common for firms to show average scores (or paste excellent reviews) on their websites and within promotional messages. Given the asymmetric impact of SIB and the ever-growing involvement of new reviewers (who are found to be more susceptible to this bias), these policies are likely to exacerbate SIB in the future, hence contributing to rating bubbles.

To hospitality management, what is relevant is the computation of the costs and benefits of SIB in the long run. Such evaluation is a cumbersome task, given the complexity of the relationships at play. On the one hand, rating bubbles benefit firms because of the positive impact on prices and sales. On the other hand, these dynamics might bear long-run costs in two respects. First, as rating systems are upper-bounded, a continuous improvement in this performance indicator becomes impossible. Second, as rating bubbles become the norm, the informative content and the reliability of rating platforms would diminish. What is winning in user-generated content is the ability to discern good from bad quality. An excessive presence of excellent ratings stemming from SIB might quickly move the market towards different ways to assess quality. For example, social media marketing conducted by influencers can be considered in competition with online rating platforms. Since influencers are compensated for their sponsorship, one challenge that hotel managers face is how to convey authenticity and reliability regarding the comments posted by influencers on social media (Gretzel, 2018).

Investigating these matters is likely to continue in the future, as the present work has many limitations to tackle, primarily related to the small sample size and the specific case study under investigation. These limitations are also opportunities for future extensions of this line of investigation based on field experiments. Given the relevance of UGC and online ratings in the service market, the expansion of the research scope might work through the replication of the experiment across and within different tourism businesses (restaurants, amusement parks, and cultural activities) in order to assess how rating platforms affect the overall experience at the destination. This would also say a firm word on the different impacts of SIB for experience and search goods, on which we provide a clear result.

Similarly, as the experiment took place in a mass-tourism leisure destination during peak season, it is fundamental to assess the robustness of findings to different types of hotels (e.g., located in cultural or business destinations) and to reviews written in different seasons: the possibility of a "seasonal component" in SIB stems from the specific characteristics and preferences of tourists pursuing leisure activities in different times of the year: tourists in the low season have different motivations, are usually less sensitive to mass tourism behaviours, and less prone to follow the crowd (Candela & Figini, 2012; Figini & Vici, 2012). Hence, their experience at the destination might be of different content compared to customers in the high season, and it would not be surprising to find a lower susceptibility to SIB.

Regarding the sample size, the collection of more observations or the repetition of the same experiment on a larger scale would allow undertaking within-group analysis with a higher statistical power. Moreover, implementing new sets of treatments, such as public vs private posting of reviews, would enable us to investigate another interesting phenomenon of online rating platforms, the multiple-audience effect. Specifically, ratings might differ if the subject knows that the rating will be displayed to the large public or kept secret for the provider of the service/product. In general, the experimental design will have to increase its sophistication, possibly using the laboratory for robustness checks and further insights on the sample characteristics, to evaluate better the role played by non-respondents. In this line, the laboratory should be considered a valuable complement, not an alternative, to field experiments (Viglia & Dolnicar, 2020).

In this last respect, we were able to intercept a relevant share of "non-reviewers," i.e., customers who do not regularly use online platforms but agreed to participate in the experiment. However, the overall response rate of the experiment (19%) does not exclude specific non-response bias stemming from non-observable characteristics that might affect online behaviour and susceptibility to SIB. Given the growing involvement of new reviewers in rating platforms, this is the most exciting area in which to design future field and laboratory experiments.

References

Anderson, E. T., & Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, *51*(3), 249–269.

Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *American Economic Review*, *87*(5), 847–862.

Aral, S. (2014). The problem with online ratings. *MIT Sloan Management Review* 55(2): 47–52.

Book, L. A., Tanford, S., & Chen, Y.-S. (2016). Understanding the impact of negative and positive traveler reviews: Social influence and price anchoring effects. *Journal of Travel Research*, *55*(8), 993–1007.

Brau, R., Scorcu, A. E., & Vici, L. (2009). Assessing visitor satisfaction with tourism rejuvenation policies: The case of Rimini, Italy. *Journal of Environmental Planning and Management*, *52*(1), 25–42.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge University Press.

Candela, G., & Figini, P. (2012). *The Economics of Tourism Destinations*, Springer, Heidelberg.

Çelen, B., & Kariv, S. (2004). Distinguishing informational cascades from herd behavior in the laboratory. *American Economic Review*, *94*(4), 484–498.

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.

Choi, Tat Y., & Raymond Chu. (2001). Determinants of hotel guests' satisfaction and repeat patronage in the Hong Kong hotel industry. International Journal of Hospitality Management, 20(3), 277–297.

Coker, B. L. (2012). Seeking the opinions of others online: Evidence of evaluation overshoot. *Journal of Economic Psychology*, *33*(6), 1033–1042.

Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., & Riedl, J. (2003). Is seeing believing?: How recommender system interfaces affect users' opinions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 585–592). ACM.

Cui, G., Lui, H.-k., & Guo, X. (2010). Online reviews as a driver of new product sales. In *2010 International Conference on Management of e-Commerce and e-Government* (pp. 20–25). IEEE.

Daikeler, J., Bošnjak, M., & Lozar Manfreda, K. (2019). Web versus other survey modes: An updated and extended meta-analysis comparing response rates. *Journal of Survey Statistics and Methodology,* 8(3), 513-539.

De Langhe, B., Fernbach, P. M., & 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.

Donaker, G., Kim, H., & Luca, M. (2019). Designing Better Online Review Systems. *Harvard Business Review*, November-December 2019 Issue.

Fagerland, M. W. (2012). t-tests, non-parametric tests, and large studies—a paradox of statistical practice?. *BMC Medical Research Methodology*, *12*(1), 1-7.

Figini, P. & Vici, L. (2012). Off-season Tourists and the Cultural Offer of a Mass-Tourism Destination: the Case of Rimini, *Tourism Management*, 33(4):825-839.

Filieri, R., & McLeay, F. (2014). E-wom and accommodation: An analysis of the factors that influence travelers adoption of information from online reviews. *Journal of Travel Research*, *53*(1), 44–57.

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

Foster, A. (2005). A Non-Linear Model of Information Seeking Behaviour. Information Research: *An International Electronic Journal*, 10(2).

Frank, B., Torrico, B. H., Enkawa, T., & Schvaneveldt, S. J. (2014). Affect versus cognition in the chain from perceived quality to customer loyalty: The roles of product beliefs and experience. Journal of Retailing, 90(4), 567-586.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498–1512.

Gupta, P., & Harris, J. (2010). How e-wom recommendations influence product consideration and quality of choice: A motivation to process information perspective. *Journal of Business Research*, *63*(9), 1041–1049.

Han, S., & Anderson, C. K. (2018). Estimating the Effect of Social Influence on Subsequent Reviews. In *INFORMS International Conference on Service Science* (pp. 231-238). Springer, Cham.

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

ISTAT (2005). Il turismo nel 2004, Statistiche in breve. Roma.

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly*, *68*(1), 94–101.

Kim, E. E. K., Mattila, A. S., & Baloglu, S. (2011). Effects of gender and expertise on consumers' motivation to read online hotel reviews. *Cornell Hospitality Quarterly*, *52*(4), 399–406.

King, R.A., Racherla, P. & Bush, V.D. (2014). What we know and don't know about online word-of-mouth: a review and synthesis of the literature. *Journal of Interactive Marketing* 28(3): 167–183.

Kraska-Miller, M. (2013). *Nonparametric statistics for social and behavioral sciences*. CRC Press.

Krishnan, S., Patel, J., Franklin, M. J., & Goldberg, K. (2014). A methodology for learning, analysing, and mitigating social influence bias in recommender systems. In *Proceedings of the 8th ACM Conference on Recommender systems* (pp. 137–144). ACM.

Lafky, J. (2014). Why do people rate? Theory and evidence on online ratings. *Games and Economic Behavior*, *87*, 554–570.

Lee, Y. J., Hosanagar, K., & Tan, Y. (2015). Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Science*, *61*(9), 2241–2258.

Lewis, G., & Zervas, G., (2019). The supply and demand effects of review platforms. *Unpublished manuscript.*

Li, H., Zhang, Z., Meng, F., & Zhang, Z. (2019). "When you write review" matters. *International Journal of Contemporary Hospitality Management*.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*(3), 458–468.

Litvin, S. W., & Sobel, R. N. (2018). Organic versus solicited hotel TripAdvisor reviews: Measuring their respective characteristics. *Cornell Hospitality Quarterly*, *60*(4), 370–377.

Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, *47*, 140–151.

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, *62*(12), 3412–3427.

Ma, X., Khansa, L., Deng, Y., & Kim, S. S. (2013). Impact of prior reviews on the subsequent review process in reputation systems. Journal of Management Information Systems, 30(3), 279-310. Chicago

Magnani, M. (2020). The economic and behavioral consequences of online user reviews. *Journal of Economic Surveys*, 34(2), 263-292.

Melián-González, S., Bulchand-Gidumal, J., & González López-Valcárcel, B. (2013). Online customer reviews of hotels: As participation increases, better evaluation is obtained. *Cornell Hospitality Quarterly*, *54*(3), 274–283.

Mittal, V., Ross Jr., W.T., Baldasare, P.M., 1998. The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. Journal of Marketing, 62 (1), 33–47.

Moe, W. W., & Schweidel, D. A. (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, *31*(3), 372–386.

Morales, A. C., Amir, O., & Lee, L. (2017). Keeping it real in experimental research—Understanding when, where, and how to enhance realism and measure consumer behavior. Journal of Consumer Research, 44(2), 465-476.

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A srandomised experiment. *Science*, *341*(6146), 647–651.

Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185–200.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, *78*(2), 311–329.

Nielsen (2016). *Global connected commerce*. Nielsen.com. http://www.nielsen.com/us/en/insights/reports/2016/

Öğüt, H., & Onur Taş, B. K. (2012). The influence of internet customer reviews on the online sales and prices in hotel industry. *The Service Industries Journal*, *32*(2), 197–214.

Ong, L. S., & Tan, J. H. (2015). Sense and sensibility of ownership: Type of ownership experience and valuation of goods. *Journal of Behavioral and Experimental Economics*, *58*, 171–177.

Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, *50*, 67–83.

Phillips, P., Barnes, S., Zigan, K., & Schegg, R. (2017). Understanding the impact of online reviews on hotel performance: An empirical analysis. *Journal of Travel Research*, *56*(2), 235–249.

Rogers, E. M. (1985). *The diffusions of innovations*. The Free Press.

Schlosser, A. E. (2005). Posting versus lurking: Communicating in a multiple audience context. *Journal of Consumer Research*, *32*(2), 260–265.

Schoenmueller, V., Netzer, O., & Stahl, F. (2020). The polarity of online reviews: Prevalence, drivers and implications. *Journal of Marketing Research*, 57(5), 853-877.

Siegel, S., & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences (2nd ed.).* McGraw-Hill.

Smironva, E., Kiatkawsin, K., Lee, S.K., Kim, J., & Lee, C.H. (2020) Self-selection and non-response biases in customers' hotel ratings – a comparison of online and offline ratings, *Current Issues in Tourism*, 23(10):1191-1204.

Sridhar, S. and Srinivasan, R. (2012) Social influence effects in online product ratings. *Journal of Marketing, 76*(5): 70–88.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, *69*(3), 213–225.

The BrightLocal (2018). *Local consumer review survey*. https://www.brightlocal.com/research/local-consumer-review-survey/

TripAdvisor (2019). *2019 TripAdvisor review transparency report.* https://www.tripadvisor.com/TripAdvisorInsights/wp-content/uploads/2019/09/2147_PR_Content_Transparency_Report_6SEP19_US.pdf

Viglia, G., & Dolnicar, S. (2020). A review of experiments in tourism and hospitality. *Annals of Tourism Research*, *80*, 102858.

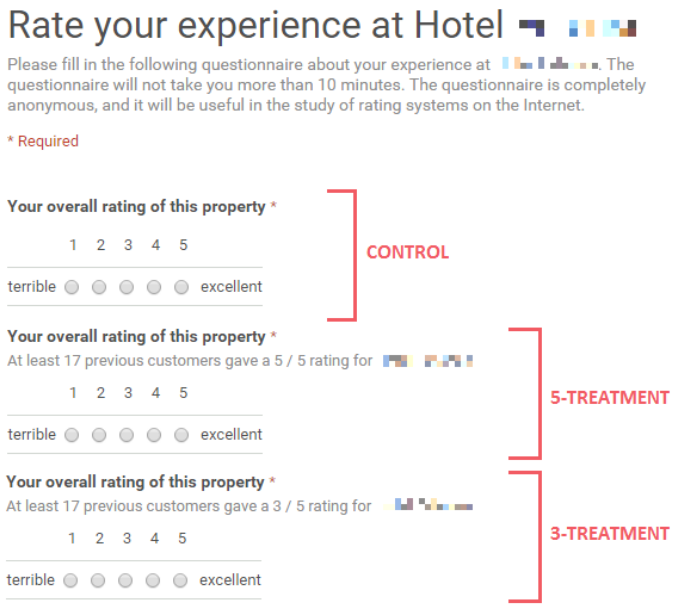
Viglia, G., Furlan, R., & Ladron-de Guevara, A. (2014). Please, talk about it! When hotel popularity boosts preferences. *International Journal of Hospitality Management*, *42*,155–164.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, *27*(2), 634–639.

Wu, Y., Ngai, E. W., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, 113280.

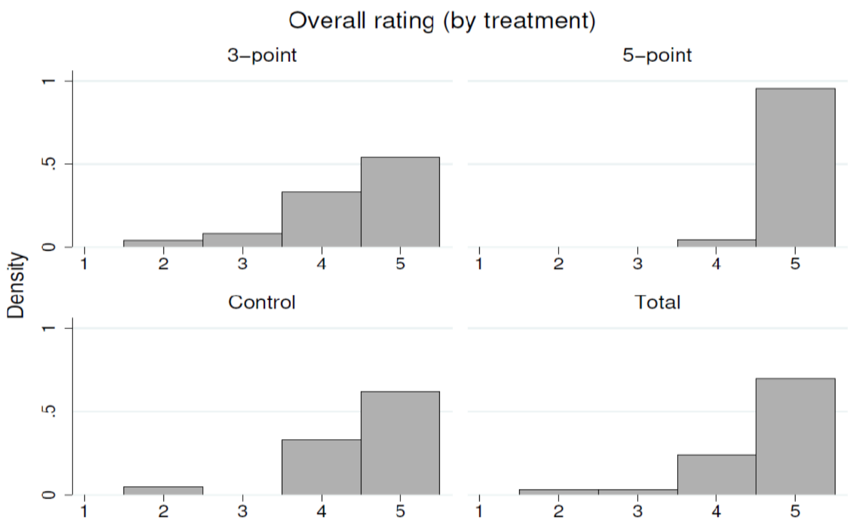
Figures

Figure 1. Different informational sets across treatments



Note. The logo and name of the hotel have been blurred for protecting anonymity.

Figure 2. Frequency distribution of overall rating by treatment



Tables

Table 1. Summary of the relevant literature

Study	**Industry/good**	**Data**	**Results**
		Effect of online reviews on product sales	
Chevalier & Mayzlin, 2006	Book	Observational	Positive effect
Ye et al., 2011	Hospitality	Observational	Positive effect
Öğüt and Onur Taş, 2012	Hospitality	Observational	Positive effect
Book et al., 2016	Hospitality	Experimental (online)	No effect of positive ratings / negative effect of negative ratings
Philips et al., 2017	Hospitality	Observational	Positive effect of positive reviews
Lewis and Zervas, 2019	Hospitality	Observational	Positive effect
		Effect of online reviews on rating behaviour	
Cosley et al., 2003	Movie	Experimental (online)	Positive effect
Schlosser, 2005	Movie	Experimental (laboratory)	No effect of positive ratings / negative effect of negative ratings
Hu et al., 2009	Book/movie	Experimental (laboratory)	Purchasing bias/underreporting bias in online reviews
Moe and Schweidel, 2012	Bath and beauty	Observational	Positive (negative) ratings increase (discourage) posting
Sridhar and Srinivasan, 2012	Hospitality	Observational	Positive effect
Muchnik et al., 2013	Opinions	Experimental (online)	Positive effect of positive ratings / positive effect of negative ratings
Krishnan et al., 2014	Opinions	Experimental (online)	Positive effect
Lee et al., 2015	Movie	Observational	Positive effect of rating from friends
Han and Anderson, 2018	Hospitality	Observational	Positive effect with a diminishing effect across review pages
Li et al., 2019	Restaurant	Observational	Positive effect with an increasing effect of temporal distance

Table 2. List of variables

Label	Definition
Socio-demographic variables	
Female	Dummy = 1 if the customer is female
Age	Customer's age in years
Italian citizenship	Dummy = 1 if the customer is Italian
Years of schooling	Customer's years of schooling (from 5, primary education, to 16, university education)
Tourist variables	
Travel type	Customer is travelling for business/as a couple/with family/with friends/solo
Travel length	Number of days spent at the hotel: 2–3 days/4–7 days/more than 7 days
First-time destination	The customer has never been to the tourist destination before
Repeat customer	Dummy = 1 if the customer has been to the hotel before
New customer	Dummy = 1 if the customer has not been to the hotel before
Type of stay	Full board/Full board All Inclusive/Half Board/Half Board All Inclusive/B&B
Period of stay	Period of stay at the hotel: 20/07–09/08; 10/08–23/08; 24/08–13/09
Rating behaviour variables	
Not reviewer	Dummy = 1 if the customer has never written an online review before
Reviewer	Dummy = 1 if the customer has written an online review before
Hotel advice	The hotel was recommended to the customer by no one/family or friends/advertising/other
Review read	Dummy = 1 if the customer has read an online review of the hotel before the stay
Review source	The customer has read an online review of the hotel on TripAdvisor or Yelp/Facebook or social networks/forums/other
Review influence	Dummy = 1 if the customer has been influenced by the review read about the hotel
Other prices	Dummy = 1 if customer looked at other hotel prices before booking the stay
Rating variables	
Overall	Hotel overall rating (from 1, terrible, to 5, excellent)
Excellent rating	Dummy = 1 if overall hotel rating is 5 (excellent)
Sleep quality	Rating for hotel sleep quality (from 1, terrible, to 5, excellent)
Value	Rating for hotel value (from 1, terrible, to 5, excellent)
Service	Rating for hotel service (from 1, terrible, to 5, excellent)
Ambience	Rating for hotel ambience (from 1, terrible, to 5, excellent)

Table 3. Summary statistics of selected variables

Variable	**Mean** (Std. Dev.)	Min.	Max.	Mode	ISTAT (2005) / Brau et al. (2009), mean or mode values
Female (D)	0.48	0	1		0.50 / 0.50
Age	45.58 (14.07)	16	74		47 / 40
Italian citizenship (D)	0.96	0	1		0.92 / 0.80
Years of schooling	11.51 (2.84)	5	16		
High-school	0.58 (0.50)				0.42 / 0.65
Travel type				Family	Family / Family
Travel length				4–7 days	
First-time destination (D)	0.27	0	1		
Repeat customer (D)	0.73	0	1		
Type of stay				Half board All inclusive	Full board / N.A.
Hotel advice				None	
Not reviewer (D)	0.52	0	1		
Review read (D)	0.33	0	1		
Review source				TripAdvisor/Yelp	
Review influence (D)	0.75	0	1		
Other prices (D)	0.49	0	1		
Period of stay				10/08–23/08	

Note. D refers to a dummy variable, for which standard deviation is not reported. The mode is reported for categorical variables, which encompass more than two categories.

Table 4. Summary statistics of the rating variables

Rating	**Mean** (Std. Dev.)	Min.	Max.
Overall	4.61 (0.70)	2	5
Sleep quality	4.49 (0.70)	2	5
Value	4.72 (0.70)	2	5
Service	4.81 (0.63)	1	5
Ambience	4.78 (0.65)	1	5

Table 5. Summary statistics of the overall rating, by treatment

Treatment	**Mean** (Std. Dev.)	Observations
Total	4.61 (0.70)	67
3-point	4.38 (0.82)	24
Control	4.52 (0.75)	21
5-point	4.95 (0.21)	22

Table 6. The effect of treatments on rating

Dep. Variable	(1) Excellent rating (logit)	(2) Overall rating (ordered logit)	(3) Overall rating (ordered logit)	(4) Overall rating (ordered logit)
5-point	2.648***	2.913***		
	(0.114)	(0.466)		
3-point	0.038	0.121		
	(0.093)	(0.308)		
Repeat customer			-0.961	
			(0.655)	
5-point * New customer			15.980***	
			(0.987)	
5-point * Repeat customer			3.088***	
			(0.824)	
3-point * New customer			-1.838***	
			(0.334)	
3-point * Repeat customer			1.787*	
			(0.790)	
Reviewer				1.587
				(1.073)
5-point * Reviewer				0.707
				(0.628)
5-point * Not reviewer				17.95***
				(1.606)
3-point * Reviewer				-0.255***
				(0.110)
3-point * Not reviewer				0.524
				0.680
Controls	Yes	Yes	Yes	Yes
(Pseudo) R^2	0.243	0.238	0.302	0.275
N. obs.	67	67	67	67

Note. ***, **, * indicate the significance at 1%, 5%, and 10% respectively. All models include the following control variables: gender, age, the stay period, being a new customer at the hotel, having read a review of the hotel, and being an online reviewer. Model 1 includes a constant term. Coefficients are reported, with standard errors (clustered at the treatment level) in parentheses.

Table 7. Predicted probability of rating 5, repeat vs new customers

	5-point treatment	3-point treatment / Control
Repeat customer	0.938***	0.840***
	(0.017)	(0.034)
New customer	1***	0.476***
	(0.000)	(0.036)

Table 8. Predicted probability of rating 5, reviewers vs non-reviewers

	5-point treatment	3-point treatment / Control
Reviewer	0.774***	0.665***
	(0.083)	(0.008)
Not reviewer	1***	0.743***
	(0.000)	(0.059)

Appendix

Figure A1. The flyer handed out to the customers

Dear Customer, **please help us to improve the quality of our hotel, and to study rating websites** together with researchers at the [redacted]

Once back home, please fill in the questionnaire you can find at the following link: http:// , and write the following code in the last page:

____

By filling in the questionnaire, you will participate to the draw of a wonderful prize. Your contribution will be really useful for our work, and for the academic research. Thanks!

The questionnaire.

Rate your experience at Hotel X

Please fill in the following questionnaire about your experience at Hotel X. The questionnaire will not take you more than 10 minutes. The questionnaire is completely anonymous, and it will be useful in the study of rating systems on the Internet.

Your overall rating of this property *

	1	2	3	4	5	
terrible	○	○	○	○	○	excellent

Please indicate your rating for each of the following categories

Sleep quality *

	1	2	3	4	5	
terrible	○	○	○	○	○	excellent

Value (quality/price) *

	1	2	3	4	5	
terrible	○	○	○	○	○	excellent

Service *

	1	2	3	4	5	
terrible	○	○	○	○	○	excellent

Atmosphere *

	1	2	3	4	5	
terrible	○	○	○	○	○	excellent

Questionnaire

Gender: *

- ○ Female
- ○ Male

Age: *

Nationality *

- Austria
- Belgium
- Bulgaria
- Croatia
- Czech Republic
- Denmark
- Finland
- France
- Germany
- Greece
- Hungary
- Ireland
- Italy
- Liechtenstein
- Luxembourg
- Netherlands
- Norway
- Poland
- Portugal
- Romania
- Russia
- Spain
- Sweden
- Switzerland
- Ukrain
- United Kingdom
- United States
- Other (Europe)
- Other (Asia)
- Other (America)
- Other

Educational qualification: *

- Primary
- Secondary
- High school
- University
- No qualification

What sort of trip was? *

- [ ] Business
- [ ] Couple
- [ ] Family
- [ ] Friends
- [ ] Solo

How many people were with you during the stay at Hotel X? *

____

When did you stay at Hotel X? *

Choose the week(s) which include the days of your stay

- [ ] 20/7 - 26/7
- [ ] 27/7 - 02/8
- [ ] 03/8 - 09/8
- [ ] 10/8 - 16/8
- [ ] 17/8 - 23/8
- [ ] 24/8 - 30/8
- [ ] 31/8 - 6/9
- [ ] 7/9 - 13/9

How much did your trip last overall? *

- ( ) 1 day
- ( ) 2-3 days
- ( ) 4-7 days
- ( ) More than 7 days

Was this your first time in [location of the hotel]? *

- ( ) Yes
- ( ) No

If this was not your first time in [location of the hotel], how many times have you already been there?

____

Have you already been at Hotel X before this trip? *

- Yes
- No

If you have already been at Hotel X, how many times?

____

Which kind of stay did you have at Hotel Estense? *

- Full board
- Full board All Inclusive
- Half board
- Half board All Inclusive
- Bed & breakfast

Do you write restaurant / hotel reviews on web sites and / or social networks: *

- Never
- Rarely
- Once in a while
- Often

Was the hotel recommended to you by anyone? *

- [ ] No one
- [ ] Family / Friends
- [ ] Advertising

Have you read reviews for Hotel X on the Internet before booking? *

- Yes
- No

If yes, on which website/s?

- [ ] TripAdvisor / Yelp
- [ ] Facebook / Social networks
- [ ] Forums

If yes, did those reviews influence your choice?

- [ ] Yes
- [ ] No

Did you have a look at the prices and / or services of other hotels through online web-sites?

- [ ] Yes
- [ ] No

Thank you for filling out the questionnaire

If you wish you can leave a comment about your stay at Hotel Estense in the box below