This is the accepted manuscript of the article:

Ji, X., Nicolau, J. L., & Liu, X. (2023). Long rating scales trigger positive ratings: When referent thinking beats relative thinking. *Annals of Tourism Research*, *102, 103625*.

Long rating scales trigger positive ratings: When referent thinking beats relative thinking

1. Introduction

Online ratings play a critical role in tourism and hospitality (Ghimire et al., 2022). 1% increase in a hotel's ratings can boost room occupancy by up to 0.54% and revenue per available room by 1.42% (Anderson, 2012). Extant studies have explored various factors that may affect consumer ratings, for example, consumer characteristics, product/service attributes, and rating systems (Chen et al., 2018). However, whether and how rating scale lengths (e.g., 1-5 or 1-10) affect consumers' rating behavior remain unclear. In hospitality, Booking Holdings (i.e., Booking.com, Priceline.com, Agoda, and Kayak) uses a 10-point rating scale, while Expedia Group (i.e., Expedia.com, Hotels.com, Hotwire, Orbitz, and Travelocity) uses a 5-point rating scale.

Considering the relevance of online reviews in different facets of hospitality operations (e.g. setting prices or rewarding employees), if an artifact to measure a service—such as the use of a scale—makes its assessment vary, then, the analysis of that artifact is not a minor issue. Previous studies have verified the importance of the right-endpoint on respondents' judgement of rating scales. Specifically, the right-endpoint indicates the end value of a range and always serves as an anchoring role in the judgement scale (Sherif & Hovland, 1961). Enlightened by the range theory and the right digit effect in the context of price perception (Coulter & Coulter, 2007; Janiszewski & Lichtenstein, 1999), the right-endpoints should receive more attention from raters and serve as the reference point in the context of online ratings when the left-endpoints (i.e., 1-score) are the same. Therefore, in the context of online ratings of hospitality related products/services wherein consumer ratings are overwhelmingly positive (for example, Liang et al. (2019) observed that the average rating score of rooms listed on Airbnb in New York City is 4.6 (out of 5), and Schoenmueller et al. (2020) found that online travel agencies exhibit high degree of positive imbalance), we posit that unless encountering extremely negative experiences that forces consumers to give low ratings (i.e., left-endpoint), the majority of consumers should perceive the right-endpoint of the rating scale as the reference point, and move from it to the left side in steps (see Figure 1).

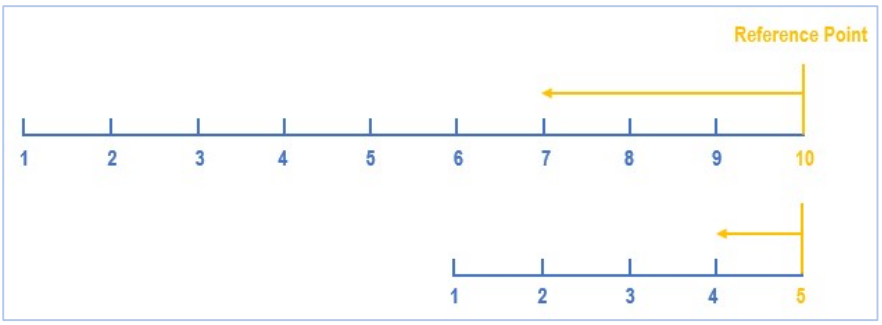

Figure 1. The reference point.

In fact, the existence of this reference point elicits a battle between two psychological effects: referent thinking and relative thinking. Relative thinking makes consumers think that moving from 5 to 4 is higher than moving from 10 to 9 because in relative terms 20% is greater than 10% (although the change is the same in absolute terms). Referent thinking sets a benchmark (5 or 10) against which the final outcome is compared.

Let us assume that consumers behaved rationally, and for a specific service, they consider that the satisfaction they obtained should correspond with a value of 4.6 (in a 1–5 scale). However, because platforms do not allow consumers to post values between integers, they would have to choose between 4 and 5. Selecting either 4 or 5 leads to a situation that is unfair: 4 is penalizing the firm too much and 5 reflects a "perfect" service that the consumer does not intend to post. More specifically, these opposing outcomes are explained as follows: In line with relative thinking, consumers would tend to avoid posting 4 because they would be aware that they would be penalizing the firm too much; rather, they would select 5 (as 4.6 is closer to their reference point, i.e. 5). However, as they perceived that the service was not "perfect", they would feel compelled to necessarily avoid 5 and would opt for 4. For the firm, this assessment would be unfair because it represents an evaluation (4) which is lower than the one the consumer actually wanted to post (4.6). However, the consumer does not have a mid-point between 4 and 5. In a 1-10 scale, this "injustice" does not occur—at least with the same degree—because the value 9 represents an assessment that is much closer to that the consumer intends to post, i.e. 9.2.

Accordingly, based on this argument, as there are more available options on the positive side of a 10-point rating scale (in Booking.com's rating system) than in a 5-point rating scale (in Expedia.com's rating system), this study asks the research question: Do consumers have a higher tendency to give positive ratings in a longer scale than in a shorter scale? Or in terms of the aforementioned battle between psychological effects, do longer rating scales result in positively skewed distributions of consumer ratings wherein referent thinking beats relative thinking?

2. Data and research design

Booking.com and Expedia.com (5-point rating scale) are two giants in the market of hotel online booking. Booking.com completely adopted the 10-point rating scale as of April

2020. We thus collect the information of consumer reviews for all star-rated hotels in California listed on Booking.com and Expedia.com from April 2020 to June 2021, respectively. We then select the same hotels listed on both Booking.com and Expedia.com. As a result, a dataset of 732,961 consumer reviews (Booking.com: 364,935 reviews; Expedia.com: 368,026 reviews) of 2,354 hotels is constructed.

We estimate a logit model to determine the probability of consumers posting positive ratings. The dependent variable is a dummy variable (*Positive*), which in order to give response to the research question posited, is defined as: *Positive* equals to 1 if consumer ratings on Expedia.com (5-point rating scale) > 3 or on Booking.com (10-point rating scale) > 6.

The independent variable of interest is *Treat*, a dummy variable set to 1 (0) for ratings posted on Booking.com (Expedia.com). As discussed in prior studies, other factors also affect consumers' rating behavior, such as length of stay, review length, and reviewing date (Liu et al., 2019). We thus controlled the number of days consumers stayed (*Stay*), the length of a review (*Length*), and the reviewing date (a dummy variable to denote whether a review is posted on *Weekend*).

We use equation (1) to test whether a long rating scale triggers positive ratings through comparing consumer rating on Booking.com versus Expedia.com. In equation (1), β_1 denotes whether consumers on Booking.com (10-point rating scale) are more likely to give positive ratings than on Expedia.com (5-point rating scale), β_2 captures the effect of control variables, ε_i signifies the error term that follows an extreme value distribution.

$$Positive_i = \beta_0 + \beta_1 Treat_i + \beta_2 Controls_i + \varepsilon_i \qquad (1)$$

3. Results

Table 1 reports the logit regression results based on equation (1) with the positive values defined as Expedia.com (5-point rating scale) > 3 or on Booking.com (10-point rating scale) > 6. The coefficient of *Treat* (0.290***) in column 1 is statistically significant and positive, suggesting that the probability of generating positive ratings on Booking.com is 34% (exp(0.290)=1.34) higher than that on Expedia.com, thereby answering our research question in that consumers have a higher tendency to give positive ratings in the longer scale than in the shorter scale. We further conduct sub-sample analysis using ratings posted on 1-, 2-, 3-, 4-, and 5-star hotels in column 2-6, the coefficients of which are still significantly positive, thus confirming the previous results.

We also define *Positive* as ratings on Expedia.com (5-point rating scale) > 4 or on Booking.com (10-point rating scale) > 8, a stricter measure, to check the robustness of the above results. We observe that the results in Table 2 are essentially the same as those in Table 1, suggesting the robustness of our empirical results that long rating scales trigger positive ratings. We further drop the periods that were seriously affected by COVID-19 and find that our results in Tables 1-2 remain robust.

Table 1. Estimation results.

	(1)	(2)	(3)	(4)	(5)	(6)
	Total	1-Star	2-Star	3-Star	4-Star	5-Star
Variables	*Positive*	*Positive*	*Positive*	*Positive*	*Positive*	*Positive*
Treat	0.290***	0.595***	0.338***	0.304***	0.358***	0.325***
	(0.006)	(0.104)	(0.013)	(0.008)	(0.012)	(0.050)
Stay	-0.007***	-0.013	-0.012***	-0.008***	-0.019***	0.007
	(0.002)	(0.027)	(0.004)	(0.003)	(0.005)	(0.019)
lnLength	-0.353***	-0.350***	-0.341***	-0.365***	-0.349***	-0.340***
	(0.002)	(0.031)	(0.004)	(0.003)	(0.004)	(0.016)
Weekend	0.014**	0.053	0.012	0.013	0.020	0.087
	(0.006)	(0.106)	(0.015)	(0.008)	(0.013)	(0.057)
Constant	1.806***	0.248**	1.274***	1.891***	1.967***	2.235***
	(0.007)	(0.115)	(0.016)	(0.010)	(0.015)	(0.064)
Prob > chi^2	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo R^2	0.053	0.059	0.051	0.056	0.054	0.051
Obs.	732,961	1,857	110,830	419,519	188,125	12,630

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2. Robustness check.

	(1)	(2)	(3)	(4)	(5)	(6)
	Total	1-Star	2-Star	3-Star	4-Star	5-Star
Variables	*Positive*	*Positive*	*Positive*	*Positive*	*Positive*	*Positive*
Treat	0.109***	0.691***	0.183***	0.131***	0.133***	0.133***
	(0.005)	(0.130)	(0.013)	(0.006)	(0.010)	(0.039)
Stay	-0.036***	-0.053	-0.027***	-0.038***	-0.064***	-0.044***
	(0.002)	(0.033)	(0.004)	(0.002)	(0.004)	(0.015)
lnLength	-0.241***	-0.344***	-0.247***	-0.242***	-0.241***	-0.245***
	(0.001)	(0.038)	(0.004)	(0.002)	(0.003)	(0.012)
Weekend	0.032***	0.022	0.048***	0.029***	0.034***	0.110**
	(0.005)	(0.125)	(0.014)	(0.007)	(0.011)	(0.044)
Constant	0.494***	-0.943***	0.011	0.495***	0.741***	1.169***
	(0.006)	(0.129)	(0.014)	(0.007)	(0.011)	(0.046)
Prob > chi^2	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo R^2	0.027	0.052	0.027	0.027	0.029	0.030
Obs.	732,961	1,857	110,830	419,519	188,125	12,630

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

4. Conclusion

By virtue of the empirical analysis of 732,961 reviews collected from Booking.com and Expedia.com, we find that consumers are more likely to give positive ratings when encountering a longer rating scale. This study fills this gap in exploring whether and how rating scale lengths affect consumers' rating behavior with their actual evaluations, hence extending the limited understanding of the effect of rating systems on consumers' rating behavior. Moreover, to the best of our knowledge, this study is also an early attempt to provide empirical evidence and explanations on the positively skewed distribution of consumer ratings in hospitality.

In fact, this study contributes to theory as, in the battle waged between the two psychological effects, referent thinking beats relative thinking. Our findings indicate that the right-endpoint of a rating scale serves as the reference point in the context of overwhelmingly positive ratings in hospitality. Unless a few consumers with extremely unsatisfying experiences post negative ratings, most consumers perceive the right-endpoint as the

reference point and move from it towards the left side of the rating scale in discrete steps. The 10-point rating scale endows consumers more available score points (i.e., 7-, 8-, 9-, and 10-point) on its positive side than the 5-point rating scale (i.e., 4- and 5-point).

As the 10-point rating scale used by Booking Holdings is more likely to systematically trigger positive ratings than the 5-point rating scale used by Expedia Group, particularly for budget hotels, it could cause unfair competitions across hotels listed on different platforms. Therefore, online booking platforms using short scales should consider shifting to long scales. This way those who monitor online reviews either manually or through specialized software such as Reviewpro or Revinate would face homogenous information across platforms.

Reference

Anderson, C. K. (2012). The impact of social media on lodging performance. *Cornell Hospitality Report*, 12 (15), 6-11.

Chen, P. Y., Hong, Y., & Liu, Y. (2018). The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments. *Management Science*, 64(10), 4629-4647.

Coulter, K.S., & Coulter, R.A. (2007). Distortion of price discount perceptions: The right digit effect. *Journal of Consumer Research*, 34(2), 162-173.

Ghimire, B., Shanaev, S., & Lin, Z. (2022). Effects of official versus online review ratings. *Annals of Tourism Research*, 92, 103247.

Janiszewski, C., & Lichtenstein, D. R. (1999). A range theory account of price perception. *Journal of Consumer Research*, 25(4), 353-368.

Liang, S., Li, H., Liu, X., & Schuckert, M. (2019). Motivators behind information disclosure: Evidence from Airbnb hosts. *Annals of Tourism Research*, 76, 305-319.

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

Sherif, M., & Hovland, C. I. (1961). Social judgment: Assimilation and contrast effects in communication and attitude change. New Haven, CT: Yale University Press.