

Online Appendix
Inflation in Reputation Systems?
Newcomers, Veterans, and Socialization within a Platform Community

Online Appendix A: Initial Semi-structured Questionnaire (Qualitative Airbnb Study)

1. **You and Airbnb:** Are you a guest (or also a host), how often have you used it, when did you start using it, what for, and why. Why not staying in a hotel?
2. **Positive and negative experience:** Please describe the last time you booked a place/your most recent stay (and how did you consume the ratings/leave a rating). Now tell me about a positive experience. Now tell me about a bad (negative) experience. What are the different stages of your booking process/the Airbnb overall experience? Tell me about the weaknesses about the system. What went wrong. Disappointments.
3. **Reputation system:** Do you think Airbnb and users can be trusted? Why? What are relevant trust cues/signals that help you identify whether you should trust someone on Airbnb? Tell me more of how you interpret and leave star ratings and written reviews. How do you review others yourself? Have you ever reflected whether your (star and written) reviews may be flawed? What do you think are the socially accepted “norms” of review behavior on Airbnb? Have you ever reflected whether reputation scores may be fake? What are the problems with reviews?
4. **Dynamic changes over time:** How did the platform change/evolve over time? Do you remember an incident where information was not up to date? How did your review behavior change over time? Please compare your review behavior as a newcomer on Airbnb compared to your review behavior at this point in time.

Online Appendix B: Duration of Each Qualitative Interview (Airbnb Study)

Table B. Duration of Each Qualitative Interview

Interview	Duration
1	39:55min
2	40:07min
3	36:37min
4	41:38min
5	40:12min
6	40:31min
7	40:44min
8	43:07min
9	39:15min
10	34:39min
11	31:48min
12	40:32min
13	36:10min
14	33:51min
15	37:41min
16	36:27min
17	35:59min
18	36:31min
19	35:27min
20	35:20min
21	31:09min
22	34:12min
23	34:47min
24	33:29min
25	36:11min
26	35:32min
27	36:14min
28	33:57min
29	35:08min
30	34:45min
31	32:12min
32	33:26min
33	33:30min
34	34:34min
35	33:08min
36	33:13min
37	33:40min
38	34:17min
39	33:17min
40	32:11min
41	33:06min

42	33:09min
43	33:50min
44	32:48min
45	31:38min
46	35:12min
47	30:49min
48	31:39min
49	33:24min
50	30:02min
51	29:43min
52	29:28min
53	29:44min
54	28:26min
55	30:51min
56	31:14min
57	27:31min
58	25:36min
59	30:23min
60	32:15min

Online Appendix C: Quantitative Corroboration of the Inverted U-Shaped Relationship Between User Review Experience and Rating on TripAdvisor

Here, we present additional quantitative analyses that corroborate some of our findings from the quantitative Yelp study, and by extension, insights from the qualitative Airbnb study. Specifically, the TripAdvisor dataset confirms the same inverted U-shaped relationship between reviewer experience and rating inflation observed in the other studies, highlighting the robustness of this pattern across platforms.

Research Design, Data Collection, and Data Analysis

We obtained a dataset from TripAdvisor, which is a well-established platform where users can leave reviews for not only vacation rentals, but also hotels, restaurants, airlines, and attractions. We chose TripAdvisor, given that there is some topical overlap to our qualitative Airbnb data (given that both platforms offer vacation rentals), while it also allows us to increase the generalizability of our findings to other contexts (similar to the quantitative Yelp data set). The variety of products also means that users are more likely to have an established review history with the platform (e.g., they have both stayed at and rated a vacation rental and eaten at and rated a restaurant) and allows us to have a more comprehensive understanding of a user’s review experience.

We took the following approach for data collection. First, we obtained data for all vacation rentals listed on TripAdvisor’s vacation rental homepage. Because multiple listings on the platform are pointers to partner websites, we focused on reviews that are in TripAdvisor’s platform ecosystem

and have been reviewed at least four times. This sampling process led to a pool of 1,613 vacation rental properties. For these properties, we identified those properties with an average review of 4.5 or below.¹ Once these properties were identified, we scraped the user IDs of all users who left reviews for these properties. Since our focal interest lies in the changes in rating behaviors of reviewers over time, we then collected all the user ID's review data on the platform. Specifically, the review data encompasses the ratings, the review text data, and the property (i.e., vacation rental, hotel, restaurant) in which the review was posted. Finally, for each property identified, we collected the average review score to establish the baseline rating of the property.

Since our focal interest lies in the temporal aspect of rating behaviors for each reviewer, we sampled users who posted seven or more ratings during our data period (April 2005 – September 2022). The resulting dataset included 73,062 observations, with information on the ratings posted by 3,370 users for 51,172 unique properties.

Empirical Specification and Findings

Table C1 lists the variable definitions and summary statistics for our dataset.

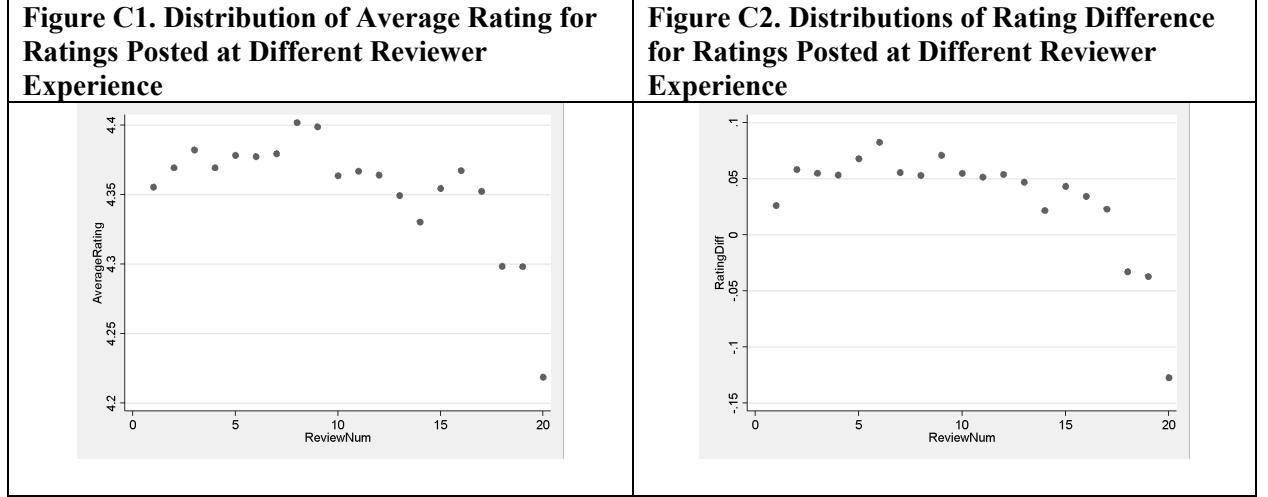
Table C1. Variable Definition and Summary Statistics

Variable	Definition	Mean	Min.	Max.	Std. Dev.
NumRating	Cumulative number of ratings posted by each user	15.19	7.00	20.00	4.80
Rating	Average rating posted by each user	4.38	1.86	5.00	0.43
NumProperties	Number of properties a user posted ratings on	14.93	5.00	20.00	4.78
PropertyRating	Average rating of properties	4.29	1.00	5.00	0.48
RatingDiff	The difference between the rating posted by a user and the average rating of a property (i.e., $Rating - PropertyRating$)	0.04	-4.00	3.22	0.88

Notes. The units of variables of *NumRatings*, *AvgRating*, and *NumProperties* are individual (*i*). The unit of variable *PropertyRating* is property.

To assess the inflation bias of reviewers, we operationalized two measures: rating (i.e., *Rating*) and the rating difference between *Rating* and the average rating of the property (i.e., *RatingDiff*). We first present model-free evidence illustrating the relationships between *Rating/RatingDiff* and reviewer experience (i.e., the *n*th review) (Figures C1 and C2). As shown in both figures, we observe similar inverted U-shaped patterns of individual rating behaviors in users' review cycle for both the *Rating* (Figure C1) and the *RatingDiff* (Figure C2).

¹ It would be challenging to identify rating inflation behaviors for properties and businesses with average review over 4.5.



Next, we formally investigated whether and how a user's rating inflation behavior changes over the reviewer lifecycle. Specifically, we examine both the linear and nonlinear relationships between a user's rating behaviors and their cumulative number of ratings. The regression specification takes the following form:

$$Outcome_{ij} = \beta_1 NumRating_{ij} + \beta_2 NumRating_{ij} * NumRating_{ij} + \varepsilon_{ij} \quad (A1)$$

where i denotes user i , j denotes property j . We employ two outcome variables ($Outcome_{ij}$):

$Rating_{ij}$ and $RatingDiff_{ij}$. $Rating_{ij}$ indicates the rating user i posted for property j ;

$RatingDiff_{ij}$ is the difference between user i 's rating for property j and property j 's average rating.

$NumRating_{ij}$ represents the cumulative number of ratings that user i posted when the user posted

the rating for property j . Robust standard errors were clustered at the reviewer level to reduce

potential heteroscedasticity concerns and to allow for correlation over time.

Table C2 reports the estimation results of our empirical model with $Rating$ as the dependent variable. As presented in column (1), the estimated coefficient of $NumRating$ is positive and significant ($p < 0.01$), and that of its squared term ($NumRating * NumRating$) is negative and significant ($p < 0.01$). These findings provide evidence of a non-linear relationship between $NumRating$ and $Rating$, indicating an inverted U-shaped pattern – thereby confirming some of the findings from the other two studies. Differently put, users tend to post higher ratings (i.e., rating inflation) as they post more reviews; however, such effects gradually diminish, and users post lower ratings (i.e., rating deflation) after a certain threshold. To assess whether these results are consistent

across different pools of users, we replicate the analyses using different samples (columns (2) – (5)) by varying the cumulative number of reviews left by users, and the results remain qualitatively consistent. We also replicate the series of analyses with *RatingDiff* as the dependent variable (Table C3) and find that the signs and significance of the coefficient of interest are consistent across all samples.

Table C2. Relationship Between the Number of Reviews and User Rating

	(1)	(2)	(3)	(4)	(5)
Variables	Users with More than 6 Reviews Rating	Users with More than 7 Reviews Rating	Users with More than 8 Reviews Rating	Users with More than 9 Reviews Rating	Users with More than 10 Reviews Rating
NumRating	0.0138** (0.00647)	0.0175*** (0.00659)	0.0184*** (0.00669)	0.0184*** (0.00681)	0.0195*** (0.00693)
NumRating * NumRating	-0.000786** (0.000325)	-0.000885*** (0.000329)	-0.000873*** (0.000334)	-0.000823** (0.000338)	-0.000838** (0.000344)
Observations	51,181	49,256	47,600	45,908	44,278
Log-likelihood	-55,157	-53,224	-51,556	-49,862	-48,135

Notes. Robust standard errors clustered on reviewer are in parentheses. Ordinal logistic regression is employed for all analyses. * p<0.1, ** p<0.05, *** p<0.01.

Table C3. Relationship Between the Number of Reviews and Rating Difference

	(1)	(2)	(3)	(4)	(5)
Variables	Users with More than 6 Reviews RatingDiff	Users with More than 7 Reviews RatingDiff	Users with More than 8 Reviews RatingDiff	Users with More than 9 Reviews RatingDiff	Users with More than 10 Reviews RatingDiff
NumRating	0.0131*** (0.00307)	0.0136*** (0.00312)	0.0143*** (0.00316)	0.0142*** (0.00321)	0.0150*** (0.00326)
NumRating * NumRating	-0.000871*** (0.000155)	-0.000893*** (0.000156)	-0.000921*** (0.000158)	-0.000914*** (0.000160)	-0.000961*** (0.000162)
Observations	48,617	47,301	46,186	45,056	43,459
Adjusted R-Squared	0.00122	0.00126	0.00128	0.00129	0.00143

Notes. Robust standard errors clustered on reviewer are in parentheses. OLS regression is employed for all analyses. * p<0.1, ** p<0.05, *** p<0.01.