Reporting Bias and Reciprocity in Online Reviews: Evidence From Field Experiments on Airbnb

Andrey Fradkin*[1], Elena Grewal†[2], David Holtz‡[2], and Matthew Pearson§[2]

[1]The National Bureau of Economic Research and Airbnb, Inc.
[2]Airbnb, Inc.

September 15, 2014

*Corresponding Author: afradkin@gmail.com
†elena.grewal@airbnb.com
‡dave.holtz@airbnb.com
§matthew.pearson@airbnb.com

Extended Abstract

Reviews and reputation scores are increasingly important for decision-making, especially in the case of online marketplaces. Sixty-eight percent of respondents in a 2013 Nielsen survey said that they trusted consumer opinions posted online.[1] However, online reviews may not provide an accurate depiction of the characteristics of a product, either because many people do not leave reviews or because some reviewers omit salient information. We study the causes and magnitude of bias in online reviews by using large-scale field experiments that change the incentives of buyers and sellers to honestly review each other.

Our setting is Airbnb, a prominent online marketplace for accommodations where guests (buyers) stay in the properties of hosts (sellers). We show that the review system on Airbnb does not fully represent the experiences of Airbnb users. This bias is caused primarily by two mechanisms: the fact that buyers with worse experiences are less likely to leave reviews and the fact that buyers omit negative feedback from publicly displayed reviews, presumably due to intrinsic reciprocity induced by socially interacting with the seller. The fear of a retaliatory review plays a comparatively minor role for public reviews. We find that by changing incentives, we are able to decrease the bias in reviews, and to create a more informative review system.

Reputation is particularly important for transactions on Airbnb because guests and hosts interact in person,[2] often in the primary home of the host. Guests must trust that hosts have accurately represented their property on the website, while hosts must trust that guests will be clean, rule abiding, and respectful. Airbnbs reputation system works as follows. After the guest checks out, Airbnb asks both the guest and the host to evaluate each other using star ratings (1 to 5 stars) and textual comments.[3] In the case of the host, the text of all previous guest reviews is visible on the profile of the user and the listing. Furthermore, rounded averages of the overall and category star ratings appear on the listing page after the listing has received 3 or more reviews. The host can only deduce the ratings left by a particular guest if the displayed star ratings change. In the case of the guest, the text of all previous host reviews appears on their user profile. The ratings that the host leaves for the guest are never shown to users. Both host and guest can also leave private feedback text for each other while completing the review. Airbnb's reputation system displays two empirical regularities that are seen in many other online reputation systems: many participants do not leave a review, and most reviews are positive.[4] Over 30% of guests in our experimental sample do not leave a review, and over 70% of reviews by guests are five stars, the highest rating possible (See Table 1 for sample averages and treatment effects).[5]

We design two experiments intended to decrease the bias in the review system. Our first

[1] Global Trust in Advertising Report, Nielsen 2013.

[2] Marketplaces that enable these types of transactions are often referred to as part of the "sharing economy" by the media. In a 2013 New York Times column, Thomas Friedman claimed that "Airbnb's real innovation is not online rentals. Its 'trust'."

[3] Fradkin [2014] shows that both guests and hosts use reputation scores as part of their decision of whom to trade with.

[4] Nosko and Tadelis [2014] and Dellarocas and Wood [2007] both document that approximately 70% of Ebay users do not leave feedback and that over 90% of public feedback on Ebay is positive.

[5] Note: These numbers do not necessarily represent the historical and current rates of reviews on the site, which change over time due to seasonality and changes in Airbnb policy.

experiment is intended to understand the experiences of guests who do not typically leave a review. The experiment works as follows. Guests who have not left a review within 9 days are assigned to either a treatment group or a control group. Guests in the treatment group are offered a $25 Airbnb credit if they leave a review. Only trips to hosts without any reviews are included in the experiment. The treatment increases the guest review rate in the sample by 16 percentage points (pp) or an 8pp increase in overall review rates given that only 50% of guests do not review within 9 days. Furthermore, the rate of five star reviews is 7pp smaller in the treatment although the total number of five star reviews increases in the treatment (see Figure 1 for the distribution of ratings in the experiments). These results are consistent with a theory where individuals who have a mediocre experience avoid leaving negative reviews because it is unpleasant. A similar sorting effect is found in the context of giving in dictator games and door-to-door fundraising in Malmendier, te Velde and Weber [2013], Lazear, Malmendier and Weber [2012], and DellaVigna, List and Malmendier [2012]. Another reason why people might not leave honest negative reviews is that they are afraid that the counterparty will respond with a negative review in return. Bolton, Greiner and Ockenfels [2012] show that feedback between buyers and sellers after a transaction is highly correlated on Ebay and other online platforms. They run a laboratory experiment with a stylized reviewing game, where some games allow for retaliation and other do not. They show that in the games where retaliation is possible, review ratings between buyers and sellers exhibit more correlation.

Our second experiment tests whether fear of retaliation is an important determinant of reviewing behavior by removing the ability of guest and hosts to retaliate against each other in reviews. The treatment changes the review system to hide any feedback until both buyer and seller have left a review (or the review time has expired). An important detail of the experiment is that once a guest or a host submits a review, an email is sent to the other party. In the case of the treatment, the first review is not revealed until the second review is submitted; therefore there could be an overall increase in the incentive to leave a review unrelated to retaliation. We first consider the decisions of the guest in response to this experiment. The treatment increases review rates by guests by 2.7pp while the share of treatment reviews that are five stars decreases by only 1.4pp. Another possibility is that guests respond to the experiment by including more negative or critical text in their reviews. We measure the presence of such text by identifying phrases that appear disproportionately in low-rated reviews.[6] We then test for changes in the frequency of reviews with these phrases and find that the rate of negative sentiment per review increases by 2pp. Therefore, the fear of retaliation plays only a minor role in biasing observed feedback compared to non-participation by guests who would leave lower ratings.

The fear of retaliation does affect reviewing behavior by guests in the context of private

[6]For the case of guest reviews of hosts, we take the sample of all reviews in the control group. We look at the set of n-grams which appear at least 4 times as much in reviews with a lower than five star rating and at least 50 times in the bad review data. For the case of host reviews of guests we look at phrase frequencies at a sample of low rated reviews which appeared before the experiment in comparison to phrase frequencies in the control group. We select phrases that appear at least 4 times as much and at least 1% of the time in the bad review data. This method identifies phrases such as "would not recommend", "complained", and "charge" for reviews of guests and phrases such as "was dirty", "the mattress", and "rude" for reviews of hosts.

feedback written from the guest to the host. Airbnb asks guests to suggest ways for hosts to improve their listings and hospitality. We find that removing the possibility of retaliation increases the share of reviews with this type of private feedback by 7pp. The increase in private feedback suggests that if guests do not fear retaliation, they are willing to provide more (constructive) criticism to hosts. However, this criticism is not necessarily reflected in public review ratings. For the set of reviews with private feedback about improvement, 66% of public feedback is a five star rating and the average overall rating is 4.6. Five star reviews are also common for trips where guests call customer service, state a low likelihood to recommend Airbnb,[7] and submit private feedback to Airbnb. In those cases, guests might be choosing to omit negative feedback from public reviews even when there is no possibility of retaliation because they do not want to hurt the hosts theyve socially interacted with. These results are consistent with experimental findings in which reductions in social distance induce reciprocal behavior.[8] This type of reciprocity can also explain why review rates on Airbnb are more than twice as high as review rates on Ebay, where buyers and sellers do not interact in person.

On the host side, the simultaneous reveal treatment increases review rates by 7pp. However, only 2pp of this increase is by hosts leaving reviews before a guest leaves a review. If fear of retaliation were the main reason for the increase, we would expect a larger share of additional reviews to be left by the host before the guests review in the treatment. Furthermore, the mean time between the review of a host and a guest decreases by 18 hours in the treatment. These facts suggest that hosts want the guests review to be publicly displayed on the website and try to leave a review soon after the guests review, if they have not done so already. The fact that hosts increase their review rates more than guests in response to the treatment likely reflects the fact that hosts are more affected by reviews than guests. Hosts earn money on Airbnb and the amount of money they earn is a function of the reviews, while guests can always use hotels as an alternative. The simultaneous reveal treatment increases negative sentiment in host reviews of guests by 1pp and the increase is driven entirely by reviews where the host does not recommend a guest. However, the rate at which hosts recommend guests is not affected by the treatment. Although this increase in negative text does make the site more informative, it affects less than 1% of stays. Therefore, the fear of retaliation is does not explain the high rate of positive reviews submitted by hosts.

We show that there is bias in online reviews and that interventions can reduce that bias. Buyers with more negative experiences are less likely to review and the fear of retaliation makes them less likely to criticize sellers in private feedback. Furthermore, even in the scenario without the possibility of retaliation, some buyers leave positive reviews when other indicators suggest a negative experience. Sellers respond to the simultaneous reveal experiment by greatly increasing review rates, although the ratings they submit are only slightly smaller in the treatment.

Based on our results, marketplaces should design reputation systems to induce a higher share of participants to leave reviews and to solicit feedback in a way that maximizes per-

[7] At the end of the review flow, guests are asked their likelihood to recommend Airbnb on a 0 10 Likert scale. This response is used to measure the Net Promoter Score.

[8] For example, in dictator games, giving from the dictator to the recipient increases when the dictator and recipient face each other (Bohnet and Frey [1999*a*], Bohnet and Frey [1999*b*]) and decreases with more anonymity between participants (Hoffman et al. [1994], Hoffman, McCabe and Smith [1996]).

ceived social distance between the reviewer and the reviewed. One way in which marketplaces can induce more reviews is by offering reviewers rewards such as status, badges, or actual money.[9] The problem of bias due to reciprocity is more difficult to address. Future research should test whether review prompts that maximize the perceived social distance between the reviewer and the person being reviewed lead to more informative reviews. Alternatively, marketplaces can identify and feature more informative reviews (as on Amazon.com and Yelp.com).

References

Bohnet, Iris, and Bruno S Frey. 1999*a*. "Social Distance and Other-Regarding Behavior in Dictator Games: Comment." *American Economic Review*, 89(1): 335–339.

Bohnet, Iris, and Bruno S Frey. 1999*b*. "The sound of silence in prisoner's dilemma and dictator games." *Journal of Economic Behavior & Organization*, 38(1): 43–57.

Bolton, Gary, Ben Greiner, and Axel Ockenfels. 2012. "Engineering Trust: Reciprocity in the Production of Reputation Information." *Management Science*, 59(2): 265–285.

Dellarocas, Chrysanthos, and Charles A. Wood. 2007. "The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias." *Management Science*, 54(3): 460–476.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *The Quarterly Journal of Economics*, 127(1): 1–56.

Fradkin, Andrey. 2014. "Seach Frictions and the Design of Online Marketplaces."

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith. 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review*, 86(3): 653–60.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith. 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior*, 7(3): 346–380.

Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber. 2012. "Sorting in Experiments with Application to Social Preferences." *American Economic Journal: Applied Economics*, 4(1): 136–163.

Li, Lingfang (Ivy), and Erte Xiao. 2014. "Money Talks: Rebate Mechanisms in Reputation System Design." *Management Science*, 60(8): 2054–2072.

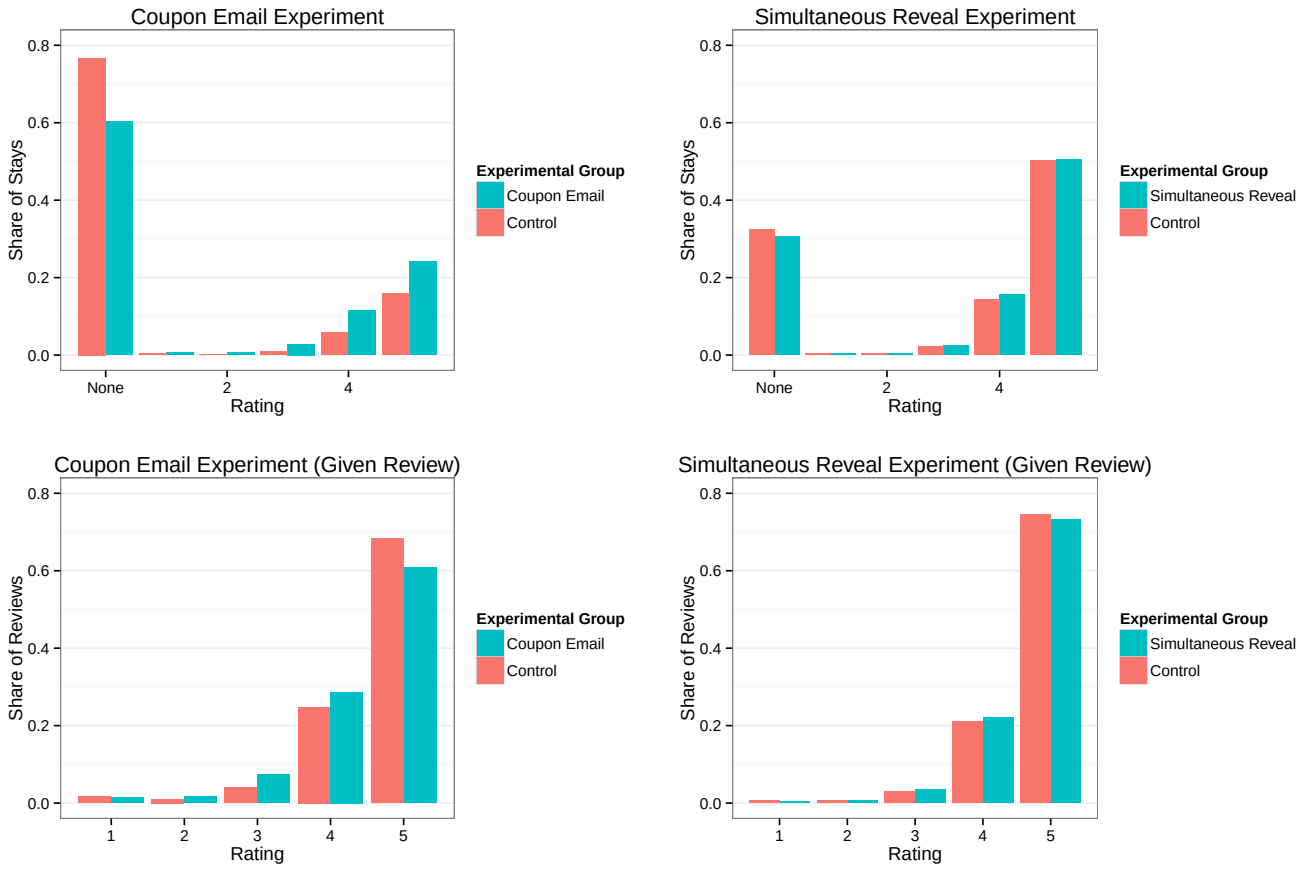
Malmendier, Ulrike, Vera L. te Velde, and Roberto Weber. 2013. "Rethinking Reciprocity."

Nosko, Chris, and Steven Tadelis. 2014. "The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment."

[9] One intriguing mechanism is to allow sellers to give rebates to buyers who leave reviews (Li and Xiao [2014]).

1 Tables and Figures

Figure 1: Distribution of Ratings



Notes: The above figure displayed the distribution of ratings in the control and treatment groups for the Simultaneous Reveal Experiment and for the Incentivized Review Experiment.

Table 1: Baseline Means and Experimental Treatment Effects

	Simultaneous Reveal			Incentivized Review		
Dependent variable	**Control Mean**	**Treatment Effect**		**Control Mean**	**Treatment Effect**	
Guest Reviews of Host						
Review Submitted	0.65	0.027 ***	0.028 ***	0.23	0.161 ***	0.161 ***
		(0.005)	(0.005)		(0.005)	(0.005)
Five Star | Review	0.75	-0.014 ***	-0.014 ***	0.68	-0.074 ***	-0.066 ***
		(0.003)	(0.003)		(0.009)	(0.009)
Rating | Review	4.68	-0.014 ***	-0.014 ***	4.57	-0.111 ***	-0.097 ***
		(0.005)	(0.004)		(0.016)	(0.015)
Improve Comment | Review	0.16	0.063 ***	0.063 ***	0.05	0.007	0.010
		(0.003)	(0.003)		(0.007)	(0.007)
Neg. Sentiment | Review	0.16	0.020 ***	0.020 ***	0.22	0.018 *	0.011
		(0.003)	(0.003)		(0.009)	(0.009)
Host Reviews of Guest						
Review Submitted	0.73	0.067 ***	0.068 ***	0.60	0.011 **	0.010 *
		(0.002)	(0.002)		(0.005)	(0.005)
Lowest Category Rating | Review	4.79	-0.015 ***	-0.014 ***	4.72	-0.014	-0.014
		(0.004)	(0.004)		(0.009)	(0.009)
Private Feedback | Review	0.24	-0.001	-0.002	0.17	-0.008	-0.009
		(0.003)	(0.003)		(0.006)	(0.006)
Neg. Sentiment | Review	0.16	0.010 ***	0.009 ***	0.16	0.001	-0.0001
		(0.003)	(0.003)		(0.003)	(0.003)
Controls		*No*	*Yes*		*No*	*Yes*
Potential Reviews		115797			37021	
Submitted Reviews by Guest		76799			11518	
Submitted Reviews by Host		88102			22290	

Notes: The averages in the Simultaneous Reveal section are taken for a sample of trips between 5-11-2014 and 6-11-2014. They do not necessarily represent the historical and current rates of reviews on the site, which differ over time due to seasonality and changes to Airbnb policy. Only guests staying with a non-reviewed host, who had not submitted a review within 10 days of check-out are eligible for the incentivized review experiment. "Neg. Sentiment" is a binary variable that takes the value 1 when a review includes at least some word or phrases identified as negative. "Lowest Category Rating" is the minimum of the cleanliness, communication, and houe rules ratings submitted by the host. Controls include guest and host gender, location and experience as well as the listing's room type and number of good pictures. Heteroskedasticity robust standard errors are displayed.
*p<0.1; **p<0.05; ***p<0.01