Fighting bias with bias: How same-race endorsements reduce racial discrimination on Airbnb

Keywords: Racial Bias, Racial Inequality, Impact of Recommendations, Platform Design, Sharing Economy

Extended Abstract

Over the past decade, the online sharing economy has flourished in several domains, including lodging, transportation, and even dating [1–3]. This success is surprising, given the risk to personal safety of sharing a home, accepting a ride, or going on a date with a complete stranger [4]. Sharing economy platforms have addressed this challenge by designing user-driven reputation systems (e.g., reviews, badges, and ratings) that discourage misbehavior and encourage mutual trust [1, 4]. In addition, most platforms also display users' names and photos, thereby reducing anonymity and conveying authenticity and accountability [1]. However, recent studies have documented an unintended consequence. Names and photos also reveal socio-demographic characteristics that can enable racial discrimination in users' choices of individuals with whom to interact [1–3, 5], thereby reinforcing racial barriers to participation in the sharing economy [5–8] and mirroring the historic bias observed offline.

Previous research shows that reputation systems can reduce discrimination in the sharing economy [5, 9, 10]. Other studies show that users pay more attention to and place greater trust in third-party recommendations from in-group members [11–13]. These findings are consistent with a half-century of social identity research on in-group bias that suggests greater responsiveness to same-race peer recommendations [14, 15].

Growing evidence of in-group bias in third-party recommendations poses the central question that motivates our study: Does racial bias in the response to peer endorsements promote or attenuate racial bias in the selection of exchange partners in the sharing economy? The preference of Whites for same-race recommendations, combined with racial bias in the strength of reviewers' endorsements, could reinforce racial discrimination against Black providers with few White customers and even fewer White endorsements. If so, previous studies may have overstated the effects of racial discrimination against Black providers by failing to control for the confounding effects of racial bias in third-party endorsements.

To find out, we analyzed data on host selection and peer recommendations in New York City using "instant bookings" collected from Airbnb, the leading online marketplace for shared accommodations. New York City has the booking density needed to control for neighborhood racial composition, and instant bookings are needed to isolate guests' racial preferences since they allow guests to select a host but do not permit hosts to choose a preferred guest. We used the Face++ facial recognition algorithm to classify users' race based on their profile photos (https://www.faceplusplus.com/). Visually revealed racial identities may differ from the user's self-identification, but Airbnb does not include race in user profiles and it is the racial identity perceived by guests that influences their behavior.

Given the powerful influence of peer recommendations on host selection, unmeasured effects of racial bias in peer recommendations may have confounded the preference for same-race hosts reported in previous studies. However, we found the opposite. The surprising result is that these two manifestations of racial bias are offsetting, not reinforcing. Guests largely overcame their racial bias in host selection when hosts were endorsed by previous same-race guests (**Figure 1**). This effect of same-race endorsement was far larger than the overall effect of an endorsement without regard to race. Moreover, we found no evidence of racial bias in the affective enthusiasm of endorsements, which suggests that the preference for same-race endorsements is motivated by the racial identity of the recommender, not the content of the recommendation (**Figure 2**).

These results have policy implications for reducing racial inequality in access to the sharing economy. Complaints about this inequality have led some companies to obscure the racial identity of users by removing profile photos [3]. However, this strategy can also backfire, by shifting discrimination to the point where racial identities become apparent, and by undermining trust when transactions are between anonymous users [3, 29]. Instead, our study suggests that it may be possible to algorithmically harness racial bias to combat racial inequality. Same-race booking preferences limit opportunities for Black hosts to be booked by, and thus reviewed by, guests of other races. Instead of hiding profile photos, platforms should selectively display the same racial composition of front-page endorsements for all providers, creating a level playing field in the positive effects of same-race endorsements on the willingness of White guests to book with hosts of a different race. Increasing the exposure of White guests to White-authored endorsements of Black hosts may lead to more White bookings, thereby making the algorithmic correction less necessary over time.

References

1. K. Levy, S. Barocas, Designing Against Discrimination in Online Markets. *Berkeley Technol. Law J.* **32**, 1183 (2017).
2. B. Edelman, M. Luca, D. Svirsky, Racial discrimination in the sharing economy: Evidence from a field experiment. *Am. Econ. J. Appl. Econ.* **9**, 1–22 (2017).
3. J. Mejia, C. Parker, When transparency fails: Bias and financial incentives in ridesharing platforms. *Manage. Sci.* **67**, 166–184 (2021).
4. J. Santana, P. Parigi, Risk aversion and engagement in the sharing economy. *Games*. **6**, 560–573 (2015).
5. J. D. Tjaden, C. Schwemmer, M. Khadjavi, Ride with me - Ethnic discrimination, social markets, and the sharing economy. *Eur. Sociol. Rev.* **34**, 418–432 (2018).
6. J. Kleinberg, J. Ludwig, S. Mullainathan, C. R. Sunstein, Discrimination in the Age of Algorithms. *J. Leg. Anal.* **10**, 113–174 (2018).
7. Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, Dissecting racial bias in an algorithm used to manage the health of populations. *Science (80-.).* **366**, 447–453 (2019).
8. P. Törnberg, How sharing is the "sharing economy"? Evidence from 97 Airbnb markets. *PLoS One*. **17**, e0266998 (2022).
9. B. Abrahao, P. Parigi, A. Gupta, K. S. Cook, Reputation offsets trust judgments based on social biases among Airbnb users. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9848–9853 (2017).
10. R. Cui, J. Li, D. J. Zhang, Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Manage. Sci.* **66**, 1071–1094 (2020).
11. P. Racherla, M. Mandviwalla, D. J. Connolly, Factors affecting consumers' trust in online product reviews. *J. Consum. Behav.* **11**, 94–104 (2012).
12. M. Zhao, J. Xie, Effects of social and temporal distance on consumers' responses to peer recommendations. *J. Mark. Res.* **48**, 486–496 (2011).
13. K. Munger, Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment. *Polit. Behav.* **39**, 629–649 (2017).
14. J. C. Turner, in *Rediscovering social identity*, T. Postmes, N. R. Branscombe, Eds. (Psychology Press, 2010), pp. 243–272.
15. H. Tajfel, Social identity and intergroup behaviour. *Soc. Sci. Inf.* **13**, 65–93 (1974).

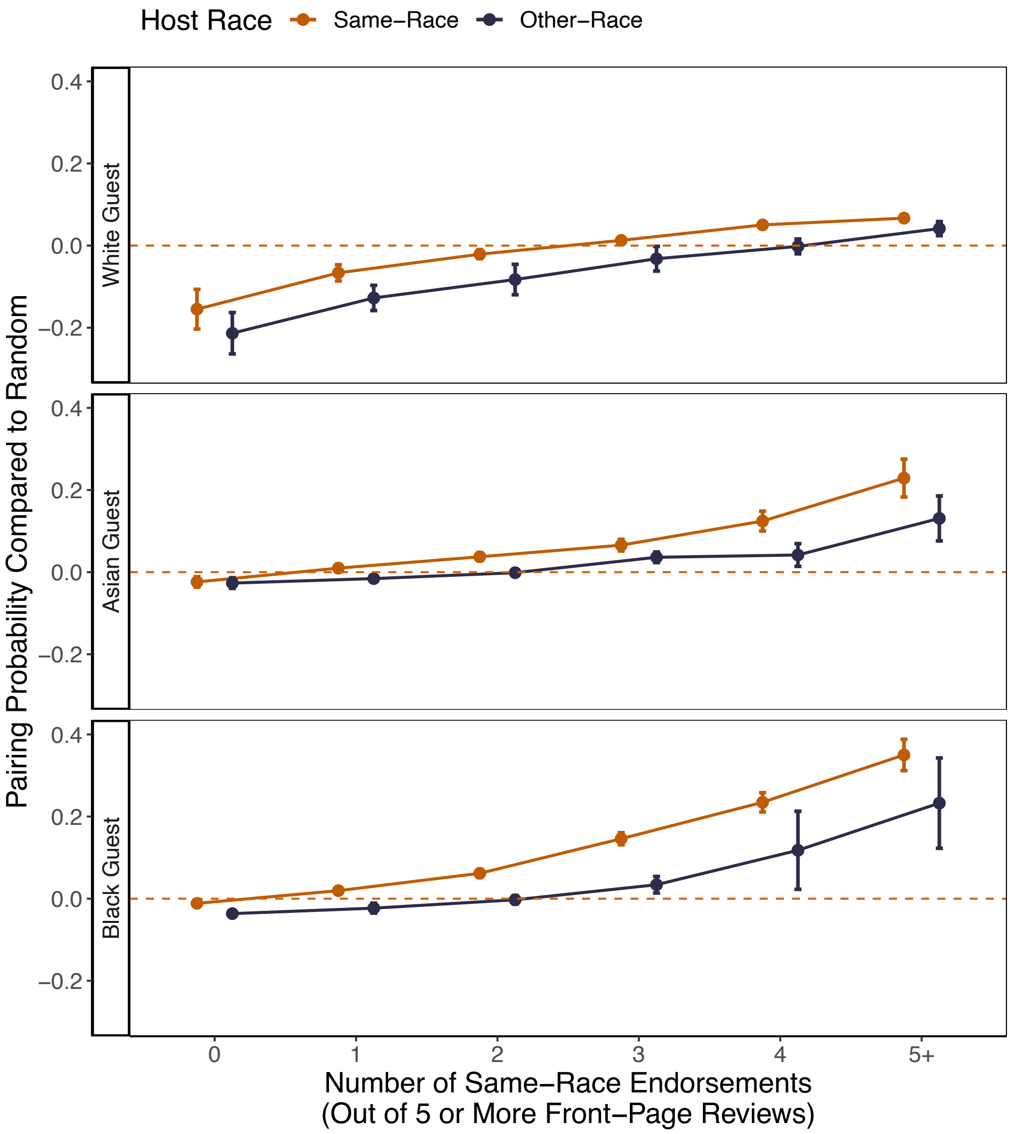

Figure 1. Response to same-race endorsements of same- and other-race hosts. The x-axis is the number of same-race endorsements of the host (out of five or more front-page reviews). The y-axis is the probability that a White (top panel), Asian (middle panel), and Black guest (bottom panel) chooses a host, compared to chance. Color indicates whether the race of the host is the same as the guest. As same-race endorsement increases from zero to five or more, the normed probability of booking with the host increases, regardless of the race of the guest or host. However, among White guests, the increase in booking probability is larger for other-race hosts (D = 0.254, compared to D = 0.222), while among Black and Asian guests, the increase is larger for same-race hosts (D = 0.361 and D = 0.253), compared to other-race hosts (D = 0.269 and D = 0.158), among Black and Asian guests respectively.

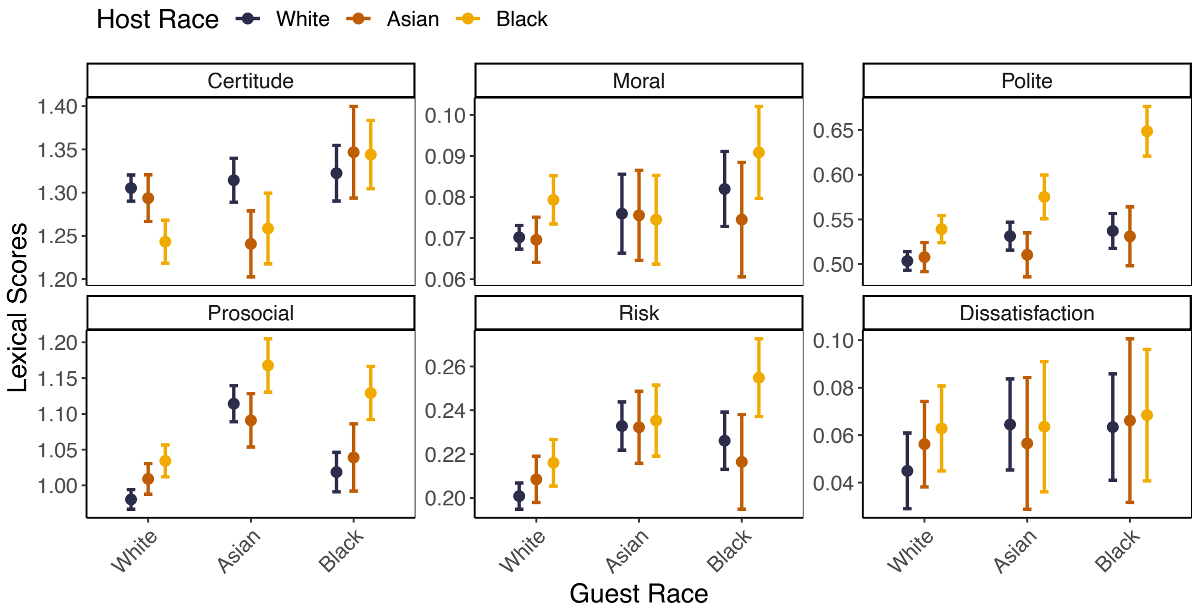

Figure 2. Lexical scores for six attributes of review content by race of guest and host. All reviews were combined for each of the nine combinations of race of guest and host. Scores for each attribute were calculated as the percent of words in the combined reviews that match words in the LIWC dictionary for that attribute. Additionally, using VADER, we compared the mean sentiment scores for endorsements written by same-race (0.968) and other-race guests (0.970) and found that sentiment scores were close to 1.0 (the upper limit) and nearly identical ($P = 0.305$), indicating that there were almost no negative reviews and other-race endorsements were no less positive than endorsements for hosts of the same race.