Tit for Tat?
The Difficulty of Designing
Two-Sided Reputation Systems

David Holtz and Andrey Fradkin

The importance of two-sided reputation systems: Perfect in theory ✕ Imagine you've just arrived in Barcelona for your family vacation, and rather than stay in a traditional hotel, you've decided to book an Airbnb. You might think to yourself, ''how can I be sure that this Airbnb property is as described on the site, and that the host will be responsive and professional?'' Along the same lines, suppose you're an Uber driver preparing to start your daily driving shift. You might wonder, ''how can I be sure that the passengers I pick up today will be clean and respectful?'' In a world without two-sided reputation systems, there would be no way to guarantee that ''sharing-economy'' transactions like those described above go smoothly. However, once a bilateral reputation system has been introduced, it is in both buyers' and sellers' best interest to be good transaction partners, since they do not want to receive a bad review that will negatively impact their ability to use the platform in the future. This data is used not just by the market participants but by the platforms as well. Platforms can use reputation data to identify struggling platform participants and help them improve, remove bad actors, and/or enable buyers and sellers to make informed decisions about who they'd like to transact with based on historical ratings. It's no coincidence that many of the most successful two-sided platforms, like Airbnb, Uber, and Upwork, feature two-sided reputation systems.

THE AUTHORS

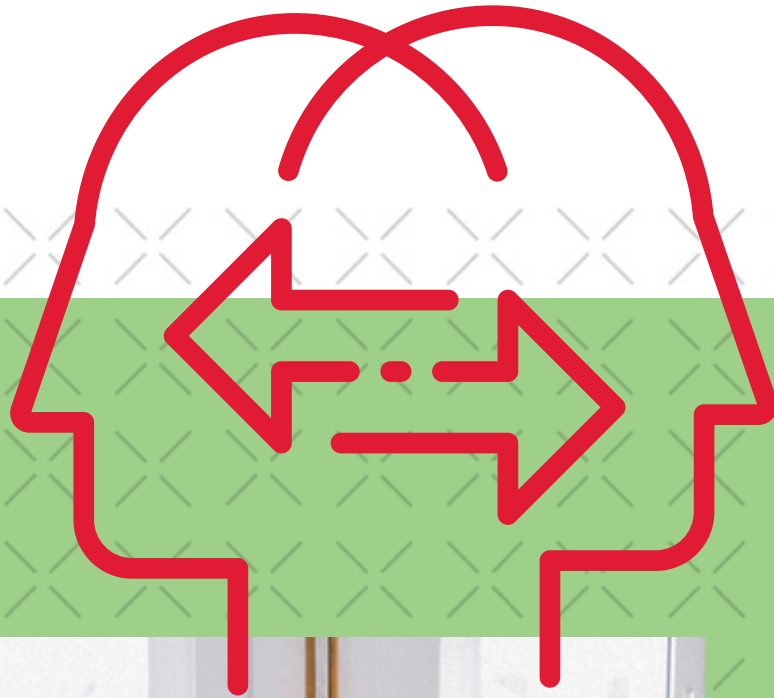**David Holtz**
PhD Candidate,
MIT Sloan School of Management,
Cambridge, MA, USA
dholtz@mit.edu

Andrey Fradkin
Assistant Professor of Marketing,
Boston University Questrom School of Business,
Boston, MA, USA
fradkin@bu.edu

Innovations in platform design increase the value of reputation systems while avoiding the common pitfalls.

←

BOX 1

# Experimenting with different review system designs

The effects and usefulness of different reputation system designs can be tested with randomized field experiments. In collaboration with Airbnb, we conducted such experiments to understand strategic reviewing behavior and the role of incentives to motivate more users to write reviews. Fairly simple system design modifications had substantial, and sometimes unexpected, effects.

Effects of simultaneously revealing reviews: design changes results
In the control group, hosts and guests could see each other's reviews before writing their own, whereas in the treatment group, hosts' and guests' reviews were not visible until both parties had written their review, or a fixed amount of time had passed.

› *Both guests and hosts felt more comfortable leaving honest feedback when their reviews were revealed simultaneously.*

› *The average rating left by guests was 0.25% lower for simultaneous revelations.*

› *The rate at which guests and hosts wrote reviews that included negative text increased by 12% and 17%, respectively, for simultaneously revealed reviews.*

› *Guests were less likely to retaliate against hosts who left them less-than-glowing reviews when the results were revealed at the same time: the rate at which guests left 1-star ratings for hosts decreased by 31%.*

› *Review rates increased with simultaneous reveals because both guests and hosts were curious what their counterpart had written.*

FIGURE 1 › Effects of simultaneously revealing reviews

Host review
Guest review
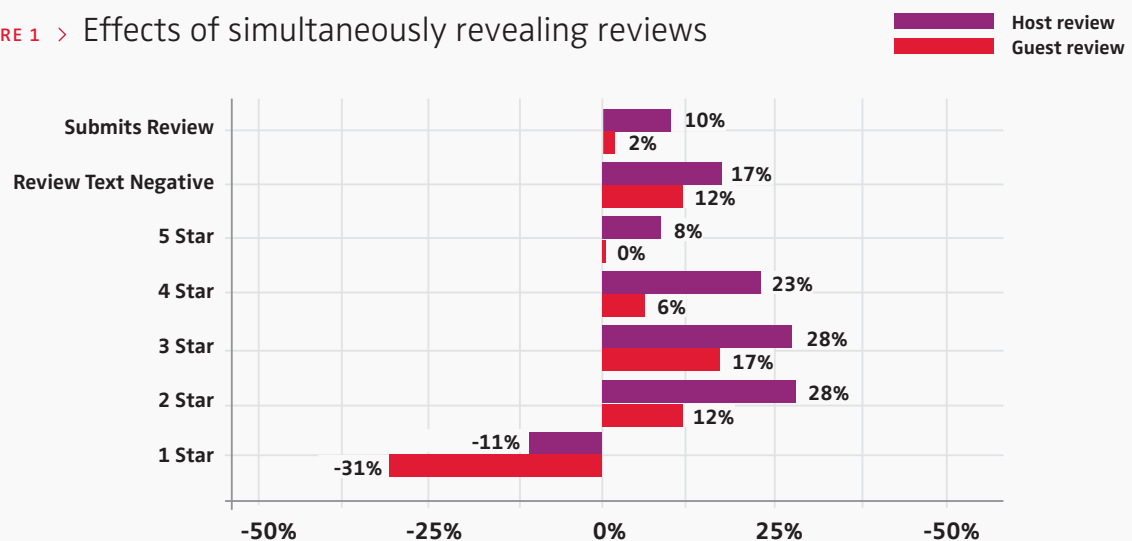
Effects of incentivizing reviews
The second experiment focused on Airbnb guests who had finished their trip but gone many days without yet writing a review of their host. Half of guests were offered a $25 Airbnb coupon in exchange for writing a review, whereas the other half did not receive any incentive to write a review.

› *Review rates increased by 70% when guests were offered a coupon.*

› *The rate of five-star reviews for hosts decreased by eight percentage points when offered a coupon.*

»
Reputation systems can have flaws due to factors such as reciprocity and retaliation, selective reviewing, and reputation inflation.
«

In practice it's quite complicated ✕ While two-sided reputation systems may seem like silver bullets that solve many of the problems that can keep an online marketplace designer up at night, the reality is unfortunately not so simple. Reputation systems can have flaws due to factors such as reciprocity and retaliation, selective reviewing, and reputation inflation. These flaws cause the ratings collected on the platform to diverge from the actual experiences that marketplace participants are having. When this occurs, two-sided reputation systems are less effective at mitigating moral hazard and adverse selection, which can lead to worse experiences for buyers and sellers alike.

Potential flaws and how to solve them ✕ Strategic reviewing behavior or reviewer bias can have a strong impact on the reviews that platform participants leave. But reputation design decisions, such as simultaneously revealing reviews or offering incentives to write reviews, can help to deliver a less biased picture of the average experiences.

〉 **Strategic reviewing behavior** ✕ Consider the seemingly minor detail of the timing of when reviews are displayed to platform users. Some two-sided reputation systems immediately post reviews online once they are written, whereas others do not. When one party's review is visible before the other has written their own review, the first reviewing party can use that first review to induce a positive review from their counterparty. Alternatively, one reviewer may wait to write their negative review (or never write it at all), out of fear that their counterparty will retaliate with a negative review of their own. Both of these factors can

lead to a two-sided reputation system that makes it seem like peoples' experiences are, on average, better than they are in reality. This would not be possible if reviews are hidden until both parties submit their reviews. An experiment on Airbnb (Box 1, Figure 1) confirmed that simultaneous reviews contain more negative feedback but reduce retaliation. We also discovered that simultaneously revealing reviews increased review rates and review speed, since guests and hosts alike were curious what their counterparty had written.

〉 **Not everyone writes reviews** ✕ In general, online marketplaces and platforms are not incentivizing buyers and sellers to write reviews. Instead, contributing to a reputation system is something people do for intrinsic reasons, e.g., to feel like an expert, or because they like the feeling of contributing to a public good. The intensity of this intrinsic motivation differs from person to person. On top of that, sometimes people just get busy and can't find time to write a review. As a result, not everyone writes reviews. For instance, on Airbnb, guests review their host 69% of the time, and hosts review their guest 79% of the time. If the subset of the population that chose to write reviews were representative, this wouldn't be a problem. Unfortunately, the population of reviewers can often be quite different from a platform's overall user population. Another randomized field experiment on Airbnb presented in Box 1 provides insight into this effect: without monetary incentive, the Airbnb reputation system missed out on a large number of guest reviews, and those missing reviews were, on average, less positive. While coupons or other monetary

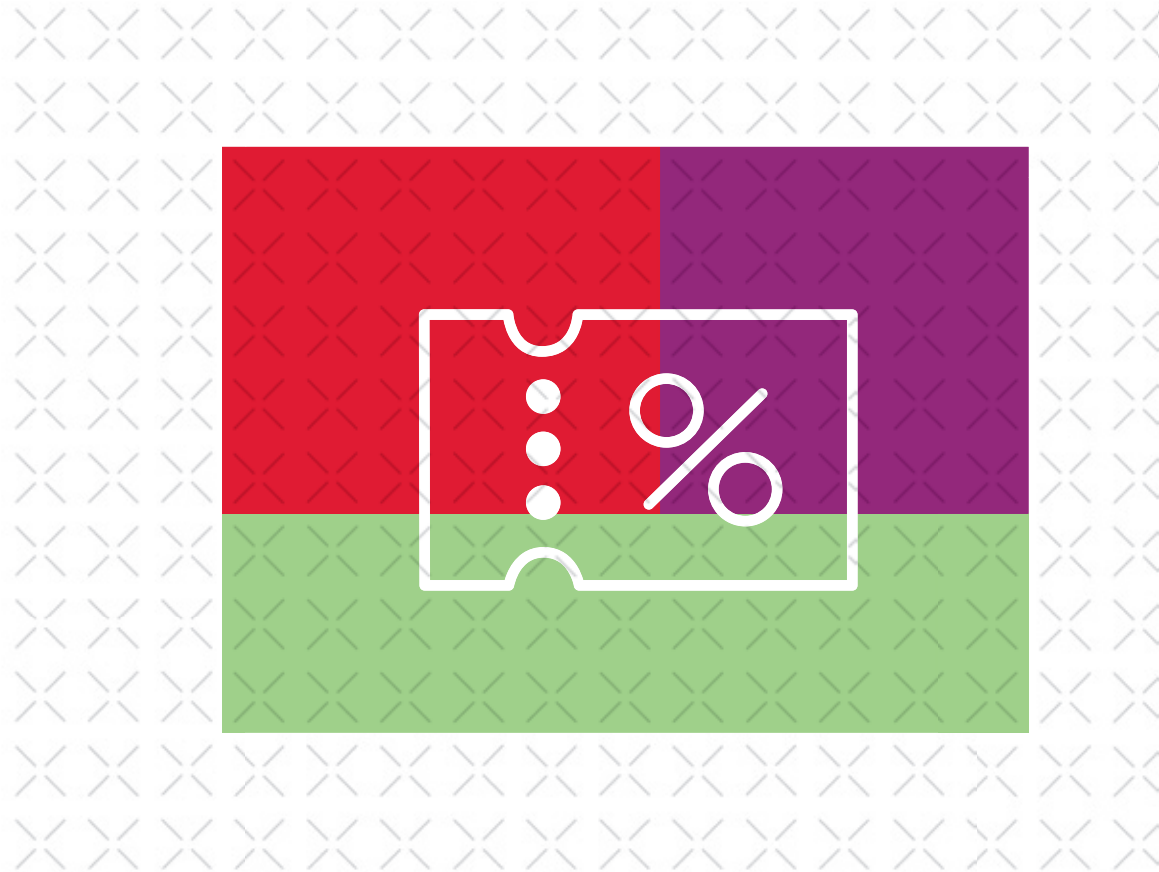FIGURE 2 > Design decisions for more efficient reviewing processes
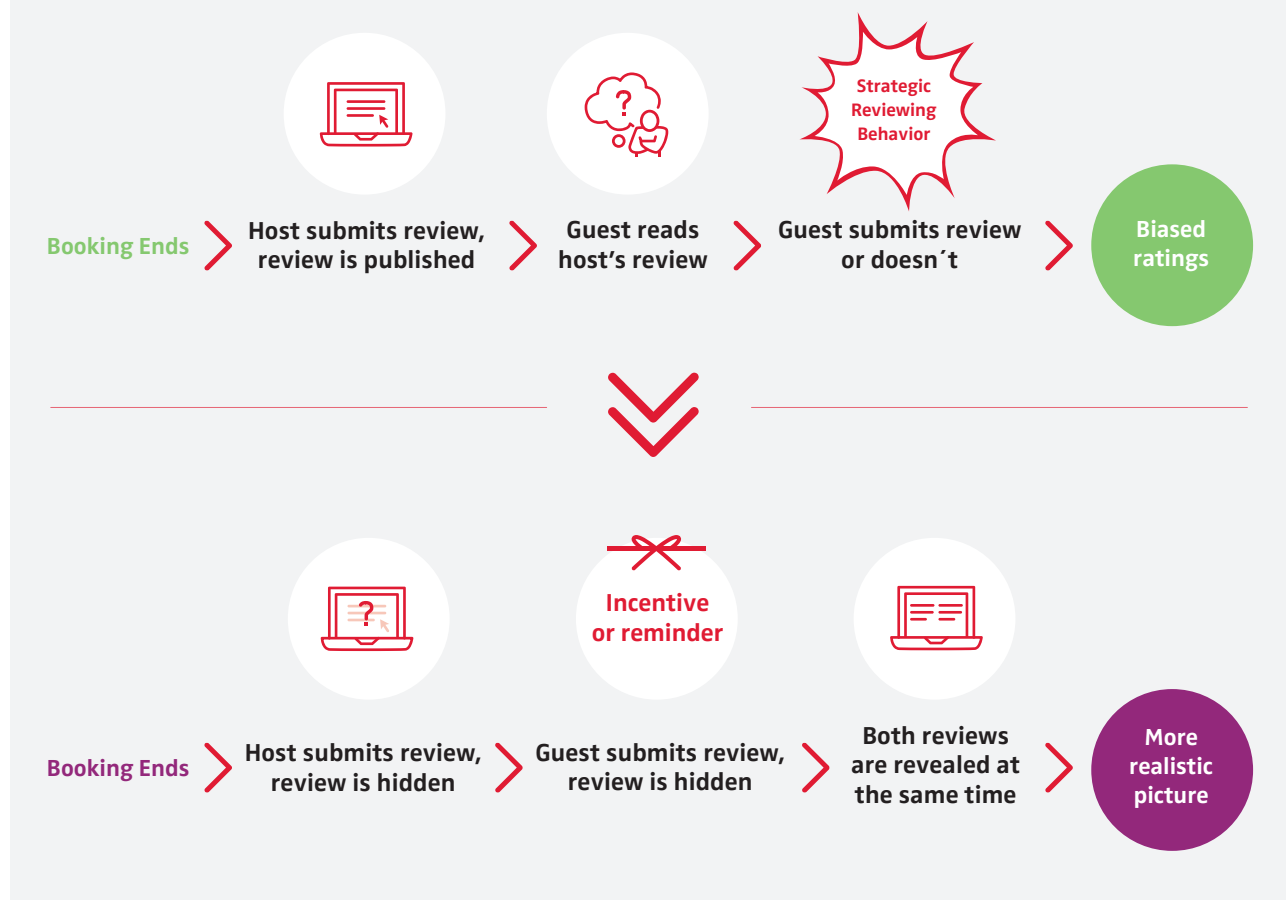
incentives provide one solution for collecting more representative reviews, it is unfortunately a costly one. Other policies that may increase review rates include reminder emails and changes to the text of those emails.

> **Reputation inflation** ✕ Another factor that can limit the effectiveness of two-sided reputation systems is what researchers call ''reputation inflation''. A recent study of feedback on Upwork, an online marketplace for freelance work, provides a textbook example of reputation inflation. From 2005 to 2014, the ratings provided to freelance workers on Upwork steadily rose, such that by 2014, 80.7% of all ratings were between 4.75 and 5 stars (out of a maximum of 5 stars). This phenomenon emerges because receiving a negative review is costly for freelancers: No one wants to hire a freelancer with a bad rating. Because of this, people feel bad leaving low ratings and are less likely to do so. Subsequently, what counts as a

''bad rating'' continually, in absolute terms, increases, until ratings are almost uniformly positive. This ratcheting pattern is not observed in private feedback because it does not have the same impact on a person's long-term business outcomes. This type of dynamic makes it difficult to distinguish between ''high-quality'' and ''low-quality'' participants on a platform, especially for new users who may not realize that what seems like a high rating is actually quite low in relative terms. The platform could influence the rate of inflation by changing the wording of the review form, and by displaying ratings computed relative to the other users on the platform. As an alternative, platforms can rely more heavily on private feedback, which is less subject to reputation inflation.

New reviewing schemes help avoid common pitfalls ✕
Two-sided reputation systems enable much of the peer-to-peer commerce occurring on platforms such as Airbnb, Uber,

»

While coupons or other monetary incentives provide one solution for collecting more representative reviews, it is unfortunately a costly one.

«

and Upwork, but designing these systems can be difficult! When reputation systems are not thoughtfully designed, it can be hard to distinguish between the "high-quality" and "low-quality" interactions. This makes it difficult to identify and remove bad actors and increases the chances of a "bad match". Figure 2 summarizes the problems of poorly designed two-sided reviewing systems and possible solutions. Innovations in reputation system design, such as simultaneous reveal of information, review incentives, and greater reliance on private feedback, are making it easier to implement two-sided systems while avoiding the common pitfalls. ✕

↓

FURTHER READING

Filippas, A.; Horton J. J.; & Golden, J. M. (2020): "Reputation Inflation." Working Paper.

Fradkin, A.; Grewal E.; & Holtz, D. (2020): "Reciprocity in Two-sided Reputation Systems: Evidence from an Experiment on Airbnb", Working Paper.

Fradkin, A.; Grewal E.; & Holtz, D. (2018): "The determinants of online review informativeness: Evidence from field experiments on Airbnb." Working Paper.

Garg, N.; & Johari, R. (2018): Designing Informative Rating Systems for Online Platforms: Evidence from Two Experiments.