

Wie du mir, so ich dir? Die Herausforderungen zweiseitiger Bewertungssysteme

David Holtz und Andrey Fradkin

Zweiseitige Bewertungssysteme: Theoretisch perfekt... ✕

Stellen Sie sich vor, Sie kommen mit Ihrer Familie in Barcelona an und suchen eine Urlaubsunterkunft über Airbnb statt in einem traditionellen Hotel. Vermutlich überlegen Sie, wie Sie sicher sein können, dass das angepriesene Apartment hält, was es auf der Webseite verspricht. Oder angenommen, Sie sind ein Uber-Fahrer, der sich darauf vorbereitet, seine tägliche Fahrerschicht zu beginnen. Sie fragen sich vielleicht: „Wie kann ich sicher sein, dass die Fahrgäste, die ich heute mitnehme, sauber und respektvoll sein werden?“ In einer Welt ohne wechselseitige Reputations- bzw. Bewertungssysteme gäbe es keine Möglichkeit, zu garantieren, dass „Sharing Economy“-Transaktionen wie die oben beschriebenen reibungslos ablaufen. Sobald jedoch ein bilaterales Bewertungssystem existiert, liegt es sowohl im Interesse von Käufern als auch von Verkäufern, ein guter Transaktionspartner zu sein. Niemand will eine schlechte Bewertung erhalten, die sich negativ auf die zukünftigen Nutzungsmöglichkeiten der Plattform auswirken könnte. Käufer und Verkäufer helfen die Informationen, auf der Grundlage vergangener Bewertungen fundierte Entscheidungen darüber zu treffen, mit wem sie ins Geschäft kommen wollen. Die Informationen aus den Bewertungen werden aber nicht nur von den Marktteilnehmern, sondern auch von den Plattformen selbst genutzt. Sie ermöglichen es den Plattformbetreibern, Teilnehmer, bei denen Probleme auftreten, zu identifizieren und sie zu unterstützen oder im schlimmsten Fall von der Plattform zu entfernen. Es ist kein Zufall, dass viele der erfolgreichsten Plattformen wie Airbnb, Uber oder Upwork zweiseitige Reputationssysteme entwickelt haben.

KEYWORDS

Zweiseitige Bewertungssysteme, Bilaterale Bewertung, Reziprozität

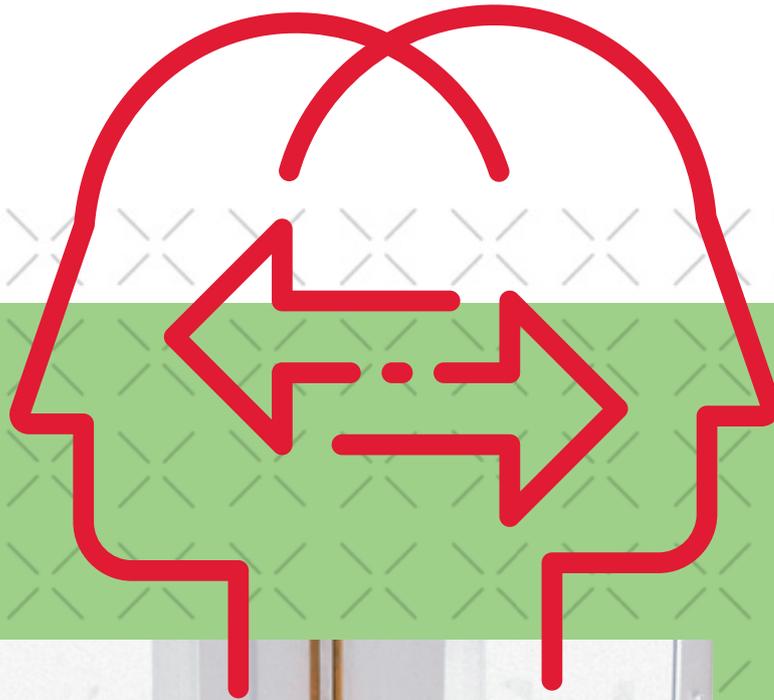
AUTOREN

David Holtz

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

Andrey Fradkin

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



Innovative Wege bei der Gestaltung steigern den Nutzen zweiseitiger Bewertungssysteme und helfen dabei, die üblichen Stolpersteine zu vermeiden.



BOX 1

Experimente mit verschiedenen Gestaltungsvarianten von Bewertungssystemen

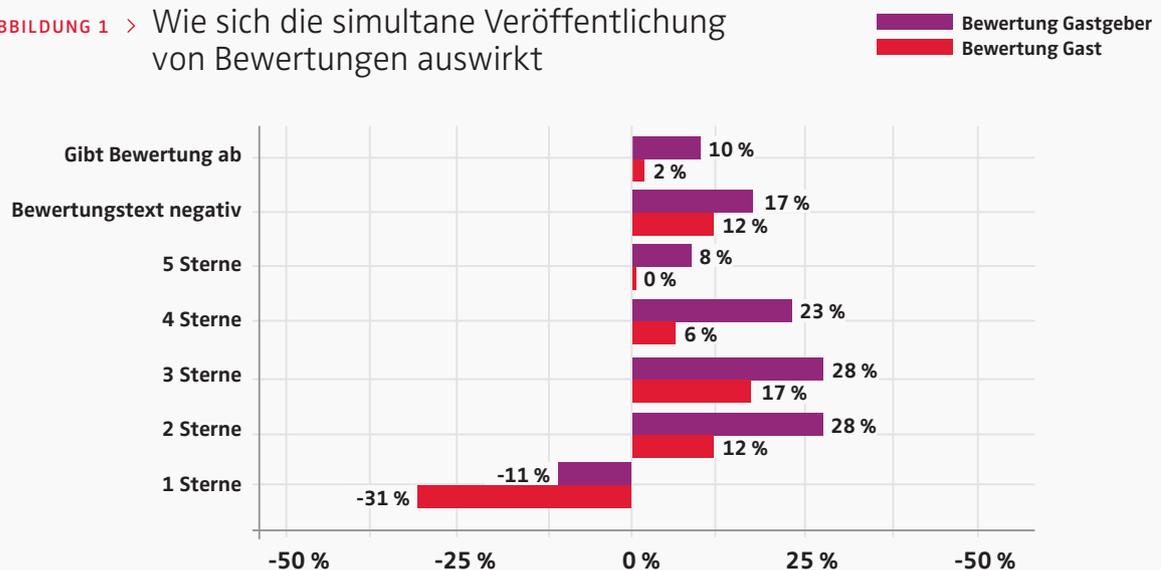
Auswirkungen und Nützlichkeit verschiedener Reputationssystem-Designs können mit randomisierten Feldexperimenten getestet werden. In Zusammenarbeit mit Airbnb führten wir solche Experimente durch, um strategisches Bewertungsverhalten zu erkunden und zu eruieren, ob und wie Anreize mehr Benutzer zum Schreiben von Rezensionen motivieren können. Es zeigte sich, dass kleine Änderungen in der Konzeption des Bewertungssystems erhebliche und manchmal unerwartete Auswirkungen hatten.

Die Auswirkungen von gleichzeitig veröffentlichten Bewertungen: Anderes Bewertungsergebnis

Im ersten Experiment konnten in der Kontrollgruppe Gastgeber und Gäste die Rezension des jeweils anderen sofort einsehen, noch bevor sie ihre eigene schrieben, während in der experimentellen Gruppe die Rezensionen von Gastgeber und Gästen erst sichtbar waren, nachdem beide Parteien ihre Rezension geschrieben hatten oder eine bestimmte Zeitspanne verstrichen war. Das Experiment zeigte die folgenden Ergebnisse:

- > Sowohl Gäste als auch Gastgeber fühlten sich wohler dabei, ehrliches Feedback zu hinterlassen, wenn die jeweiligen Bewertungen gleichzeitig veröffentlicht wurden.
- > Die durchschnittliche Bewertung durch die Gäste war bei gleichzeitiger Veröffentlichung um 0,25 % niedriger.
- > Gäste und Gastgeber waren bei einer gleichzeitigen Veröffentlichung öfter bereit, Bewertungen zu schreiben, die negative Anmerkungen enthielten. Die Anzahl der Bewertungen mit negativem Text stieg um 12 % bzw. um 17 % an.
- > Die Anzahl von Vergeltungsmaßnahmen von Gästen gegenüber Gastgebern, die ihnen wenig schmeichelhafte Kritiken hinterließen, sank deutlich: Die Rate, mit der Gäste Ein-Sterne-Bewertungen für Gastgeber abgaben, sank um 31 %.
- > Die Anzahl der Bewertungen insgesamt konnte gesteigert werden, da sowohl Gäste als auch Gastgeber neugierig waren, was ihr Gegenüber geschrieben hatte.

ABBILDUNG 1 > Wie sich die simultane Veröffentlichung von Bewertungen auswirkt



Wie Incentives wirken

Das zweite Experiment konzentrierte sich auf Airbnb-Gäste, die auch mehrere Tage nach einer Reise noch keine Bewertung ihres Gastgebers abgegeben hatten. 50 % dieser Gäste erhielten einen Airbnb-Gutschein im Wert von 25 Dollar als Gegenleistung für das Verfassen einer Bewertung, während den anderen 50 % der Gäste kein Anreiz präsentiert wurde.

- > Die Bewertungsraten stiegen um 70 %, wenn den Gästen ein Gutschein angeboten wurde.
- > Die Rate der Fünf-Sterne-Rezensionen für Gastgeber ging um acht Prozentpunkte zurück.



Zweiseitige Bewertungssysteme können aufgrund von Faktoren wie Reziprozität und Vergeltung, strategischem Bewerten und „Reputation Inflation“ Mängel aufweisen.



... aber in der Praxis ziemlich kompliziert ✗ Was wie ein Wundermittel erscheint, das viele der Probleme löst, die den Managern von Online-Marktplätzen schlaflose Nächte bescheren, ist in der Realität leider nicht so einfach implementierbar. Zweiseitige Bewertungssysteme können aufgrund von Faktoren wie Reziprozität und Vergeltung, strategischem Bewerten und „Reputation Inflation“ (einer Tendenz zur Bestbewertung) Mängel aufweisen. Diese Mängel führen dazu, dass die auf der Plattform gesammelten Bewertungen von den tatsächlichen Erfahrungen der Marktplatzteilnehmer abweichen. Wenn dies der Fall ist, sind zweiseitige Reputationssysteme weniger wirksam im Kampf gegen Fehlverhalten oder schwarze Schafe, was sowohl für Käufer als auch Verkäufer Nachteile bringt.

Problembereiche und deren Lösung ✗ Strategische Bewertungen und eine vergleichsweise geringe Anzahl an Bewertungen haben einen starken Einfluss auf das Gesamtbild, das die hinterlassenen Bewertungen ergeben. Entscheidungen zur Gestaltung des Reputationssystems, wie z. B. die gleichzeitige Veröffentlichung von Bewertungen oder Anreize zum Verfassen von Rezensionen, können jedoch dazu beitragen, ein weniger verfälschtes Bild der durchschnittlichen Erfahrungen zu vermitteln.

> **Strategische Bewertungen** ✗ Betrachten wir zunächst das scheinbar unbedeutende Detail des Zeitpunkts, zu dem die Bewertungen den Benutzern einer Plattform angezeigt werden. Einige zweiseitige Bewertungssysteme stellen Beurteilungen sofort online, sobald sie geschrieben wurden, während andere dies nicht tun. Wenn die Bewertung einer Partei sichtbar ist, bevor die andere Partei ihre eigene Bewertung geschrieben hat, kann der erste Verfasser diese erste Rezension dazu nutzen, eine positive Bewertung der Gegenpartei zu fördern. Umgekehrt ist es genauso möglich, dass ein Nutzer mit einer negativen Bewertung lange zuwartet (oder sie überhaupt nicht schreibt), weil er oder sie fürchtet, dass die Gegenpartei die Bewertung mit einer eigenen negativen Bewertung vergelten könnte. Beide Faktoren können in zweiseitigen Reputationssystemen bewirken, dass die beschriebenen Erfahrungen im Durchschnitt besser sind als das tatsächliche Erleben. Das könnte man ausschließen, wenn Beurteilungen erst veröffentlicht würden, sobald sie

von beiden Seiten vorliegen. Unsere Experimente (Box 1, Abbildung 1) bestätigen, dass simultane Bewertungen mehr negatives Feedback enthalten und weniger Vergeltungsmaßnahmen nach sich ziehen. Zusätzlich konnten wir feststellen, dass die gleichzeitige Veröffentlichung von Bewertungen die Revisionsraten und die Revisionsgeschwindigkeit erhöhte, da beide Seiten gleichermaßen neugierig auf die Bewertung der Gegenseite waren.

> **Nicht jeder schreibt eine Bewertung** ✗ Im Allgemeinen bieten Online-Marktplätze und -Plattformen Käufern und Verkäufern keine Anreize für das Verfassen von Bewertungen. Die meisten Reputationssysteme gehen davon aus, dass Menschen aus intrinsischen Motiven Bewertungen abgeben, z. B., um sich als Experten zu fühlen, oder weil sie das Gefühl haben, einen öffentlichen Nutzen zu stiften. Die Intensität dieser intrinsischen Motivation ist jedoch von Person zu Person unterschiedlich. Hinzu kommt, dass sich vielbeschäftigte Menschen oft einfach nicht die Zeit nehmen, Bewertungen zu verfassen. Dies hat zur Folge, dass viele Erfahrungen unbewertet bleiben. Bei Airbnb geben beispielsweise Gäste in 69 % der Fälle eine Bewertung ihres Gastgebers ab und 79 % der Gastgeber bewerten ihre Gäste. Wenn die Gruppe der Bewertungsverfasser für die Gesamtbevölkerung repräsentativ wäre, wäre dies kein Problem. Leider unterscheidet sich jedoch die Gruppe der Verfasser von Rezensionen oft recht stark von der Gesamtpopulation der Nutzer einer Plattform. Ein weiteres randomisiertes Feldexperiment zu Airbnb, dessen Ergebnisse in Box 1 zusammengefasst sind, gibt Einblick in diesen Effekt: Ohne monetären Anreiz verpasst das Airbnb-Reputationssystem eine große Anzahl an Bewertungen von Gästen, und die fehlenden Bewertungen sind im Durchschnitt weniger positiv. Gutscheine oder andere monetäre Anreize sind zwar eine Möglichkeit, das Gesamtbild repräsentativer zu machen, aber leider eine kostspielige. Andere, kostengünstigere Maßnahmen, die Bewertungsraten zu erhöhen, sind Erinnerungs-E-Mails oder motivierender formulierte E-Mails.

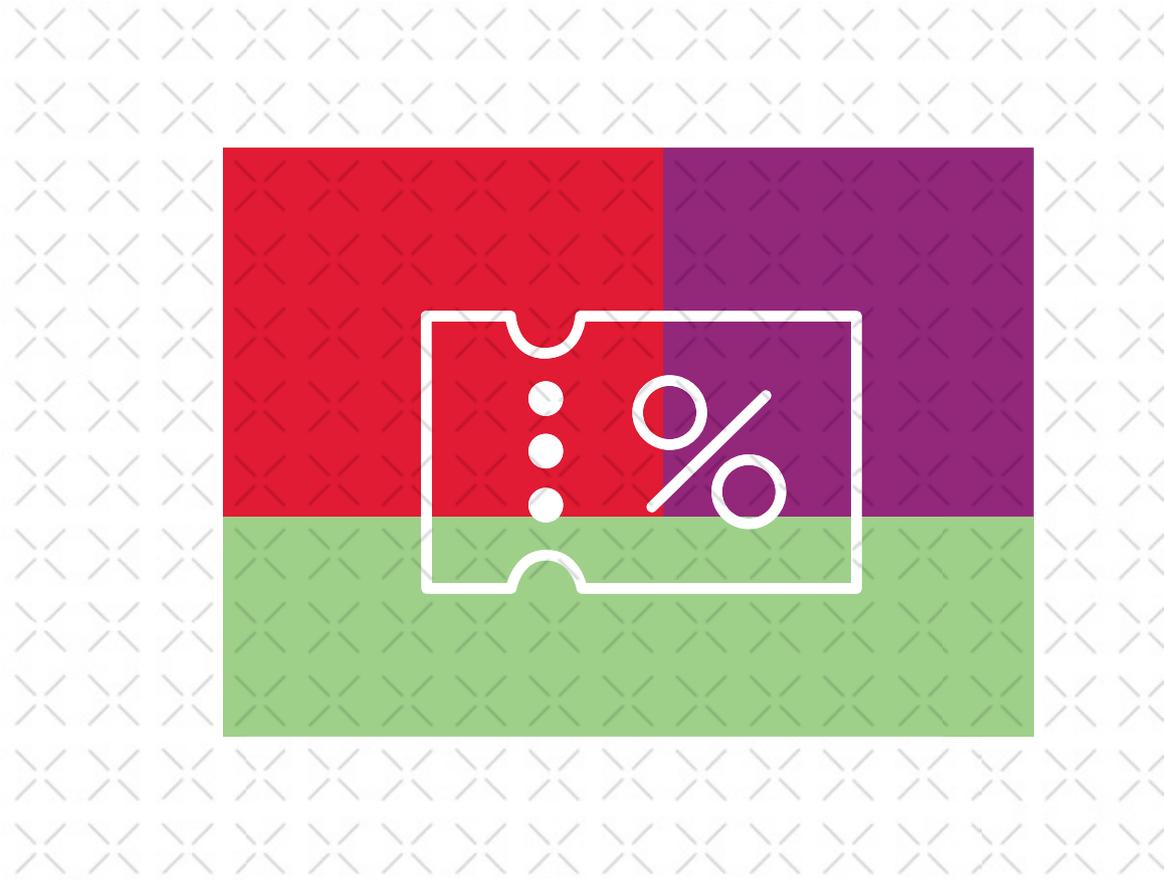
> **Reputation Inflation** ✗ Ein weiterer Faktor, der die Wirksamkeit zweiseitiger Reputationssysteme einschränken kann, ist das, was Forscher als „Reputation Inflation“

ABBILDUNG 2 > Gestaltungsoptionen für effizientere Beurteilungsprozesse


bezeichnen. Damit meinen sie das Phänomen, dass die durchschnittlichen Bewertungen auf einer Plattform mit der Zeit immer höher werden. Eine kürzlich durchgeführte Studie über das Feedback zu Upwork, einem Online-Marktplatz für Freiberufler, zeigt wie aus dem Lehrbuch, was damit gemeint ist. Von 2005 bis 2014 wurden die Bewertungen, die die Freiberufler auf Upwork erhielten, immer positiver, so dass bis 2014 bereits 80,7 % aller Bewertungen zwischen 4,75 und 5 Sternen (von maximal 5 Sternen) lagen. Dieses Phänomen entsteht, weil die Folgen einer negativen Bewertung für den Einzelnen gravierend und teuer sind: Niemand will einen Freiberufler beauftragen, der schlechte Bewertungen hat. Aus diesem Grund fühlt man sich schlecht, wenn man eine niedrige Bewertung abgibt, und lässt es lieber bleiben. Weil diese negativen Bewertungen fehlen, steigen die Durchschnittswerte immer weiter an und schlechte Bewertungen unterscheiden sich in absoluten Zahlen kaum mehr von den positiven. Bei vertraulichen Rückmeldungen, die nicht von allen eingesehen werden können, ist dieses Muster nicht

zu beobachten, da diese sich nicht gleich auf die langfristigen Geschäftsmöglichkeiten einer Person auswirken. Diese Art von Dynamik macht es schwierig, zwischen „qualitativ hochwertigen“ und „qualitativ minderwertigen“ Teilnehmern einer Plattform zu unterscheiden. Insbesondere neue Benutzer sind sich vielleicht nicht immer sofort dessen bewusst, dass eine scheinbar hohe Bewertung relativ gesehen tatsächlich eher schlecht ist. Plattformen könnten dieses Problem angehen, indem sie die Formulierungen der Bewertungsformulare ändern oder relative Werte im Vergleich zu anderen angeben anstatt der absoluten Werte. Eine andere Alternative bestünde darin, stärker auf vertrauliches Feedback zwischen zwei Partnern zu setzen, wobei das geschilderte Problem dann weniger auftritt.

Mit neuen Bewertungsprozessen die typischen Stolpersteine vermeiden ✗ Zweiseitige Bewertungssysteme leisten einen wichtigen Beitrag zum Funktionieren von Peer-to-Peer-Geschäften auf Plattformen wie Airbnb, Uber und Upwork,



»

Gutscheine oder andere monetäre Anreize sind zwar eine Möglichkeit, das Gesamtbild repräsentativer zu machen, aber leider eine kostspielige.

«

aber diese Systeme zu konzipieren, ist gar nicht so einfach. Wenn Reputationssysteme schlecht durchdacht sind, wird es schwierig, zwischen „guten“ und „schlechten“ Interaktionen zu unterscheiden, und für alle Teilnehmer steigt die Gefahr, einen unpassenden Austauschpartner zu wählen. Zusätzlich wird es auch für Plattformbetreiber herausfordernder, schwarze Schafe zu identifizieren und zu entfernen. Abbildung 2 fasst die Probleme wenig ausgereifter zweiseitiger Bewertungssysteme zusammen und zeigt mögliche Lösungen auf. Innovative Wege bei der Gestaltung von Reputationssystemen, wie z.B. die gleichzeitige Offenlegung der Bewertungen, Anreize für die Abgabe von Bewertungen und ein stärkerer Fokus auf vertrauliches Feedback, steigern den Nutzen zweiseitiger Systeme und helfen dabei, die üblichen Stolpersteine zu vermeiden. ✕

↓

LITERATURHINWEISE

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.