Kas, Judith

Article — Accepted Manuscript (Postprint)

The effect of online reputation systems on intergroup inequality

Journal of Behavioral and Experimental Economics

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Kas, Judith (2022) : The effect of online reputation systems on intergroup inequality, Journal of Behavioral and Experimental Economics, ISSN 2214-8043, Elsevier, Amsterdam, Vol. 96, pp. --, https://doi.org/10.1016/j.socec.2021.101800

This Version is available at:
https://hdl.handle.net/10419/248471

The effect of online reputation systems on intergroup inequality*

Judith Kas

Department of Sociology / ICS, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands.

Currently works at the "Migration, Integration, Transnationalization" research unit, Social Science Center Berlin (WZB)

***Correspondence**: Judith Kas, judith.kas@wzb.eu. Reichpietschufer 50, 10785 Berlin, Germany

Funding: This work was supported by the Netherlands Organization for Scientific Research, grant 452-16-002.

Acknowledgements: I would like to thank Rense Corten, Arnout van de Rijt, Karen Cook and Robb Willer for their feedback that greatly improved the experimental design and the manuscript.

Ethics approval: The study is approved by the Ethics Committee of the Faculty of Social and Behavioral Sciences of Utrecht University (FETC19-047).

Highlights

- Reputation systems may negatively affect already-disadvantaged individuals.
- I used an experiment to test if reputation systems reduce intergroup inequality.
- The reputation system did not reduce demographic inequality.
- Emergent inequality in initial rounds affected inequality in later rounds.

Abstract

Recent studies state that online reputation systems may decrease or even eliminate intergroup inequality. These studies base their conclusion on the finding that the presence of reputation information may decrease platform users' reliance on demographic information. However, these findings tell only part of the story of the relation between reputation systems and intergroup inequality: they overlook the problem that reputation systems help those who obtained at least one review but may negatively affect those individuals who have not participated in interactions before. This study is the first to take the endogeneity of the reputation building process into account by providing a direct test of differences in inequality between individuals of different nationality between a situation with and without a reputation system. Using a preregistered online experiment, I show that inequality between American and Indian individuals is not lower when there is a reputation system than when there is no reputation system. Moreover, emergent inequality in initial rounds affects inequality in later rounds. This implies that platforms that wish to create equal opportunities for users with different backgrounds should not only try to make their reputation system more effective, but also to reduce initial differences between individuals with different backgrounds.

Keywords: Reputation systems, inequality, discrimination, trust, cumulative advantage

1. Introduction

Discrimination is a widespread and persistent problem in the peer-to-peer platform economy (Cui, Li, & Zhang, 2020; Edelman & Luca, 2014; Edelman, Luca, & Svirsky, 2017; Ert, Fleischer, & Magen, 2016; Ge, Knittel, MacKenzie, & Zoepf, 2016; Jaeger & van Beest, 2019; Laouénan et al., 2017; Mohammed, 2017; Pope & Sydnor, 2011; Tjaden, Schwemmer, & Khadjavi, 2018; Wu & Jin, 2018; Wu, Ma, & Xie, 2017). For instance, Black Airbnb hosts and eBay vendors earn less than White ones (Edelman & Luca, 2014; Nunley, Owens, & Howard, 2011). Discrimination is unconstitutional and has negative consequences both for individuals experiencing negative discrimination and for platforms that miss out on potentially fruitful transactions.

Recent studies suggest that reputation systems, which collect, aggregate and distribute feedback about past behavior (Resnick, Kuwabara, Zeckhauser, & Friedman, 2000) and that are widely used in online markets, may decrease or even eliminate discrimination (Abrahao, Parigi, Gupta, & Cook, 2017; Cui et al., 2020; Ert et al., 2016; Mohammed, 2017; Tjaden et al., 2018). This claim is based on the idea that digital discrimination is mostly statistical discrimination and that it can be solved by providing better information about individuals. Earlier research found that initial differences between individuals with different demographic backgrounds declined or even disappeared once these individuals had a positive reputation (Abrahao et al., 2017; Cui et al., 2020; Ert et al., 2016; Tjaden et al., 2018). For example, Tjaden et al. (2018) found that the difference in the number of clicks per ride between drivers with a typical German name and drivers with a typical Arab/Turkish/Persian name disappeared when the driver had a five-star rating. The mechanism underlying this finding is that reputational information is more relevant and specific to the choice of a seller or buyer, thereby decreasing the importance of more diffuse demographic information.

However, these findings about static effects of reputation only tell part of the story of the relation between reputation systems and intergroup inequality. Although statistical discrimination can be solved by providing more accurate information about disadvantaged individuals, reputation systems may do exactly the opposite when demographic information is available. They may provide more information about already-advantaged individuals than about disadvantaged ones, similar to the Matthew effect in science (Bol, de Vaan, & van de Rijt, 2018; Merton, 1968). Since reviews can generally only be written after a completed transaction,

individuals who have demographic characteristics that are preferred by others are more likely to be selected for transactions and thus more likely to obtain initial reviews.

This is relevant, because obtaining a first review is important for an individual's chances to participate in future interactions on the platform. As Frey and van de Rijt (2016) showed, reputation tends to cascade: Individuals who already have a positive review are more likely to be selected for new transactions and will as a consequence accumulate more reviews. Obtaining a first review is therefore of utmost importance for a platform user's future chances to participate in interactions. Random initial differences accumulate over time, resulting in inequality between equally trustworthy individuals.

If we accept that individuals with different demographic characteristics differ with respect to the probability of receiving a first review, and that a first review is critical for acquiring more reviews, then the difference between individuals with different backgrounds in the probability of being selected may grow over time. We know for example that requests from guests with distinctively African American names on Airbnb are less likely to be accepted than requests with distinctively White names (Edelman et al., 2017). As a consequence, guests with African American names may both participate in fewer transactions and accumulate fewer reviews, which may result in a prolonged disadvantage. Results from a study using data from a peer-to-peer rental platform support this claim (Kas, Corten, & van de Rijt, 2021). In the current paper, I study whether reputation systems decrease intergroup inequality by using an online experiment inspired by an experiment employed by Frey and van de Rijt (2016) to look at reputation cascades. Their study was designed to observe the emergence of reputation cascades that resulted from random intial differenes. In the current study I extended their design by studying structural differences between groups, rather than random differences between individuals.

Kas et al. (2021) also studied the endogenous nature of reputation systems on intergroup inequality. However, because they used data from a real-life platform, they could only study how the effect of the reputation system on inequality changed over time, rather than comparing inequality between the same situation with and without a reputation system. Other limitations of their study were that the number of subjects with a minority background was low, and that they could only observe a single sequence of reputation building. Finally, in their experiment, the decision to participate in future interactions was made by the subjects, and as a consequence depended on outcomes of earlier interaction, which distorted the interpretation of the result.

The current study is the first study that directly compares inequality between individuals of different nationalities in situations with and without a reputation system. Using an experiment allows me to draw causal conclusions about the effect of reputation on demographic inequality based on multiple sequences of reputation accumulation. Furthermore, it allows me to control the variation of background among the respondens, and the number of interactions they engage in. Finally, it allows me to study multiple sequences of reputation building. The goal of this study is to evaluate whether reputation systems may reinforce rather than reduce intergroup inequality when demographic information is available.

2. Literature review: trust and discrimination in online markets

Discrimination is an important and persistent problem in the peer-to-peer platform economy (Cui et al., 2020; Edelman & Luca, 2014; Edelman et al., 2017; Ge et al., 2016; Jaeger & Sleegers, 2020; Jaeger & van Beest, 2019; Laouénan et al., 2017; Mohammed, 2017; Nunley et al., 2011; Pope & Sydnor, 2011; Tjaden et al., 2018; Wu & Jin, 2018; Wu et al., 2017). These online platforms enable people who need particular goods and people who want to provide them to connect. Examples of platforms include the hospitality platform Airbnb and sales platform eBay. Trust and reputation building are particularly important on these platforms for a number of reasons. First, in these markets buyers interact with individuals rather than with businesses, which results in increased risk and uncertainty. Online exchanges are anonymous, and strangers often interact just once, with no prospect of future interactions (Kuwabara, 2015; Parigi, Santana, & Cook, 2017). Both buyers and sellers face a risk when trading with unknown others (Macy & Skvoretz, 1998; ter Huurne et al., 2018). Buyers may have difficulties assessing the value of the goods or services traded or in determining what the interests of the seller are (Akerlof, 1970). Likewise, sellers may face uncertainty regarding the intentions of the buyer with respect to the payment. Second, buyers and sellers interact through online platforms, rather than face-to-face. This limits the opportunity to get to know the other person before a transaction. Third, peer-to-peer interactions are less strictly regulated than traditional business-to-consumer trade, allowing for a broader interpretation of what acceptable behavior is (Katz, 2015; ter Huurne, Ronteltap, Corten, & Buskens, 2017).

Platforms have applied different methods to solve this trust problem. The majority of platforms try to reduce anonymity by allowing their users to create a user profile that generally contains their name, a photo and a short personal description. These profiles foster a sense of personal

contact (Dubois, Willinger, & Blayac, 2012; Guttentag, 2015) and to provide information that platform users use to assess the trustworthiness of their potential trading partners. However, an unintended consequence of these profiles is that they may lead to discrimination (Ahuja & Lyons, 2019; Carol, Eich, Keller, Steiner, & Storz, 2019; Cui et al., 2020; Edelman & Luca, 2014; Edelman et al., 2017; Ert et al., 2016; Laouénan et al., 2017; Tjaden et al., 2018). Names and photos convey information about users' demographic characteristics, of which ethnicity and nationality, gender, and age are the most obvious.

Platform users may discriminate on the basis of this information for different reasons. First, users may infer certain unobservable qualities, such as trustworthiness, from these demographic characteristics. Individuals may believe that demographic characteristics correlate with trustworthiness, and form expectations about the other's trustworthiness based on his or her demographic characteristics. It is referred to as statistical discrimination (Arrow, 1973). This type of discrimination may be reduced by providing better information about the unobserved characteristics of other users. This more accurate and more specific information decreases the informativeness of more diffuse demographic information and may therefore reduce the reliance on this information (Resnick et al., 2000; Robbins, 2017; Wozniak & MacNeill, 2020)

Users may also have other reasons to include demographic information when assessing others. They may have a preference for one group of people over the other, without having an underlying belief on the relation between demographic characteristics and other merits. This is referred to as "taste-based discrimination" (Becker, 1957). Generally, people tend to have a preference for others who are similar to them (McPherson, Smith-Lovin, & Cook, 2001). The result of this preference for similar others may be that individuals who are in the numerical majority (with respect to ethnicity, nationality, gender and age) have a relative advantage compared to those in the numerical minority. The goal of this paper is not to distinguish between statistical and taste-based discrimination, but to study the effect of the presence of a reputation system on intergroup inequality – regardless of whether inequality is caused by beliefs about unobserved characteristics, or by a mere preference for (similar) others.

3. Conceptual framework

3.1 Using reputation systems to reduce intergroup inequality

Reputation systems are proposed to be the most promising solution to statistical discrimination on online platforms, by researchers (Abrahao et al., 2017; Cui et al., 2020; Ert et al., 2016; Mohammed, 2017; Tjaden et al., 2018) and practitioners (Murphy, 2016) alike. Most platforms allow their users to provide a numeric rating and write a review about their trading partner after the transaction has been completed. These ratings and reviews are displayed on the user's profile and are visible to potential future trading partners. They may use these reviews to assess the trustworthiness of the vendors and consumers who are reviewed (Buskens & Raub, 2002; Cook, Hardin, & Levi, 2005; Weigelt & Camerer, 1988). A person whom trust is granted, is known in the literature as the 'trustee'.

Reputational information is believed to improve the evaluation of the trustee's trustworthiness. The reliance on reputational information when assessing if a trustee can be trusted or not assumes that people's past behavior is related to their future behavior. Reputation is a costly signal: For individuals who are sincerely trustworthy, it is easy to acquire a reputation, while it is difficult for less trustworthy individuals to obtain good reviews (Gambetta, 2009; Przepiorka & Berger, 2017; Przepiorka & Diekmann, 2013; Raub, 2004). Individuals who have been trustworthy in the past are believed to be more trustworthy in general, and thus to be more trustworthy in the future. Therefore, individuals who have behaved well in the past can be expected to also behave well in the future, making reputational information valuable for those who are assessing other people's qualities. Individuals who have a better reputation (i.e., more and more positive ratings) should be more likely to be selected as interaction partners than individuals with a worse reputation. Because of the high risk involved with online peer-to-peer transactions, reputation systems are crucial for interpersonal trust in the platform economy. The positive effect of a good reputation on trust is well established in the literature (Boero, Bravo, Castellani, & Squazzoni, 2009; Bolton, Katok, & Ockenfels, 2004; Charness, Du, & Yang, 2011; Duffy, Xie, & Lee, 2013; Fehrler & Przepiorka, 2013; Jiao, Przepiorka, & Buskens, 2021). Obtaining a first review has been found to be of utmost importance for future trust. For example, the probability that a guest will be accepted by a host on Airbnb increases from 8.4% for guests without positive reviews, to 29.5% for guests with one positive review (Cui et al., 2020). At the same time, reputation systems provide an incentive for users to behave well on the platform, since misbehavior in the present limits their opportunities for future transactions (Bolton et al., 2004; Fehrler & Przepiorka, 2013; Resnick & Zeckhauser, 2002).

Reputation systems increase trust and they are also considered a better and more accurate source of information when assessing another user's trustworthiness on a platform than demographic

information. While demographic information can be considered a diffuse type of information with limited informativeness for specific cases, ratings and reviews are very specific to the situation (Resnick et al., 2000; Robbins, 2017). When no reputational information is available, users will rely on the limited demographic information they have, but when they have both demographic information and reputational information at their disposal, they are expected to rely more on the relevant and specific information rather than the more diffuse information. Indeed, previous studies found evidence for this "compensation effect:" The presence of reputational information reduced the importance of demographic information (Abrahao et al., 2017; Cui et al., 2020; Ert et al., 2016; Mohammed, 2017; Tjaden et al., 2018). Based on these findings researchers concluded that reputation systems are a solution to discrimination in the platform economy.

3.2 How reputation systems may increase intergroup inequality

However, the difference between individuals with or without positive reviews only tells part of the story. The conclusion that reputation systems reduce inequality assumes that every individual is equally likely to receive their first review. I contest this assumption, because in general reviews can only be written and ratings assigned after a completed transaction, and not all users are equally likely to get a transaction. Due to statistical and taste-based discrimination, some individuals are more likely to be selected for an interaction than others. Kas et al. (2021) showed that rental requests from renters with a minority background on a peer-to-peer motorcycle sharing platform were less likely to be accepted than requests from renters who belong to the majority, and, as a consequence, were less likely to receive a first review.

We also know that getting a first review is essential for the opportunity to participate in future transactions and for gathering more reviews (Frey & Van De Rijt, 2016). Those who already have good ratings and reviews are more likely to be selected for future transactions, and may therefore accumulate even more reviews, while those individuals who do not have a reputation yet will have a hard time getting both transactions and reviews. In their experiment Frey and van de Rijt (2016) showed that reputation systems may lead to arbitrary inequality between equally trustworthy trustees. Initial random choices between trustees may give selected trustees the benefit of building a reputation, while others, who may not get the chance to prove that they are trustworthy, do not have this benefit.

Kas et al. (2021) used simulations to show that when these initial differences are not random but based on the demographic characteristics of the trustee, the initial disadvantages for

individuals with a minority background may accumulate when there is a reputation system. This may lead to a reinforcement of the initial disadvantage experienced by groups of individuals with certain characteristics. The reputation system may especially amplify initial differences between groups when the initial level of trust in disadvantaged individuals is low. In that case, individuals belonging to a disadvantaged group may not get the opportunity to receive a first review, which hampers the extent to which they can build a reputation and thus their future opportunities on the platform. By trying many times to participate in an interaction, disadvantaged individuals may in theory eventually manage to acquire that first necessary review. However, we know from earlier research that users of online platforms typically interact only a handful of times via those platforms (Lauterbach, Truong, Shah, & Adamic, 2009; Resnick & Zeckhauser, 2002; Teubner, 2017). It is especially likely that individuals who are repeatedly rejected by potential trading partners will give up after a few attempts.

Figure 1 summarizes the mechanism discussed in the conceptual framework (section 3). In order to draw conclusions about the effect of the presence of a reputation system on intergroup inequality, one should both look at static differences between trustees with and without a reputation and consider the system as a whole. By only comparing subjects with and without reviews, as most previous papers that look at the effect of reputation on inequality have done (Abrahao et al., 2017; Cui et al., 2020; Ert et al., 2016; Tjaden et al., 2018), researchers overlook the endogenous nature of the relation between reputation and inequality. This study is the first to compare the effect of the presence of a reputation system on the reinforcement of intergroup inequality with a control condition without a reputation system.

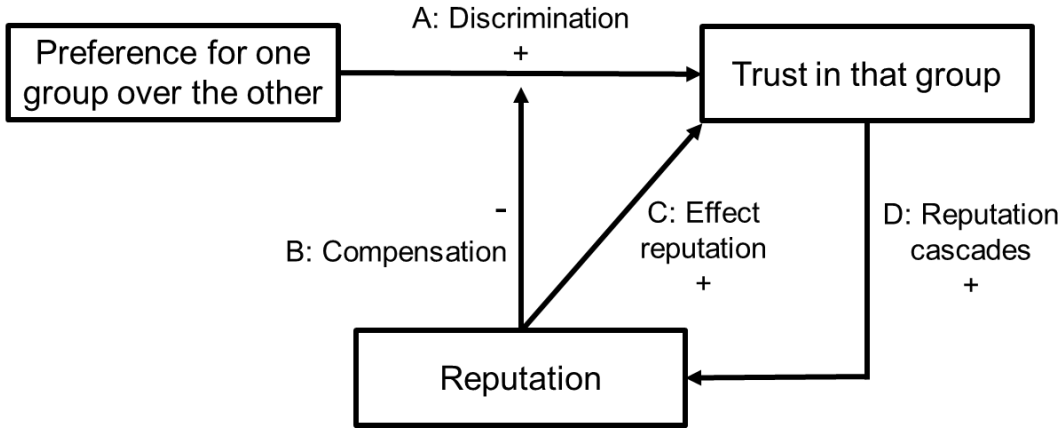

Figure 1: Conceptual model

There are thus two different mechanisms that determine the total effect of a reputation system on intergroup inequality. On the one hand, reputation systems may decrease the importance of demographic information, thereby decreasing overall discrimination. I hypothesize that:

> **Hypothesis 1**: The relative difference in trust received between individuals of different nationalities is *smaller* when there is a reputation system than when there is no reputation system.

On the other hand, initial differences between users with different nationalities may translate into differences in reputation, which in turn affect future possibilities for interaction on the platform. This leads to Hypothesis 2.

> **Hypothesis 2:** The relative difference in trust received between individuals of different nationalities is *larger* when there is a reputation system than when there is no reputation system.

The aim of this paper is to study how reputation systems affect intergroup inequality when the dynamics of reputation building processes are taken into account. Reputation systems are expected to affect different types of discrimination: Regardless of whether one group is preferred over the other by all individuals, or whether individuals prefer others who are similar to them, the theory predicts that reputation systems may sustain or reinforce these differences. Similarly, the theory applies to any situation with intergroup inequality, whether they are based on nationality, gender, age, or other characteristics.

In the current paper I focus on discrimination by national origin. I conducted a preregistered experiment in which American and Indian subjects interacted with each other. I investigated if individuals were more likely to select one group over the other, and if this tendency was different when there was a reputation system than when there was no reputation system.

4. Material and methods

4.1 Design

The current study builds on previous work on the dynamics of reputation and discrimination by using a pre-registered laboratory-style online experiment (see the supplementary materials for the preregistration report). The experiment was inspired by an experiment by Frey and van de

Rijt (2016). In the experiment, subjects played two trust games of eight rounds each in groups of eight[1]. The left panel of Figure 2 shows the structure of a one-shot trust game without a reputation system. In this figure, T_i, R_i, P_i and S_i represent the material payoffs of the actors. The right panel of Figure 2 contains the numbers used in the current experiment. Half of the group members played the role of trustor, the other half the role of trustee. Subjects played the same role throughout the game. The trustors took turns: Trustor 1 played in round 1, trustor 2 in round 2 and so on. In round 5, it was again the turn of trustor 1. When it was their turn, the trustor could choose to place trust in one of four trustees. When the trustor selected a trustee, that trustee could choose to honor the trust or abuse the trust. The trustor's payoff was highest (50 points) when they placed trust in a trustee and when the trustee honored their trust, but lowest when the trustee abused their trust (0 points). The trustee's payoff was highest when they abused trust (80 points). Trustors who were not active in a round and trustees who were not selected got a payoff of 30 points. When the trustor did not place trust in anyone, all players got a payoff of 30 points. At the end of the round all players were informed about the points they collected in that round. Under the assumption of rationality, trustors would anticipate that selfish trustees will always abuse trust. Trustors would therefore be motivated to select the most trustworthy trustee.

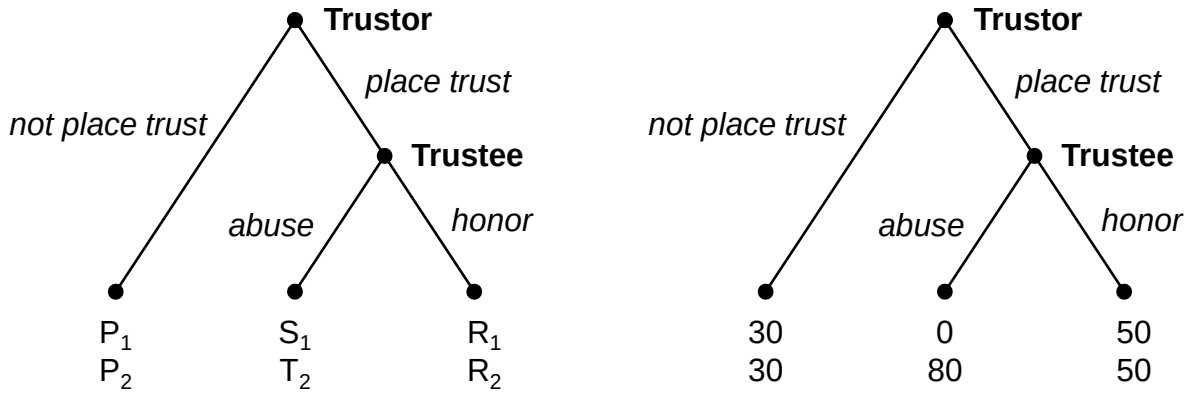

Figure 2: Left panel: one round of the trust game, where $R_1 > P_1 > S_1, T_2 > R_2 > P_2$. Right panel: numerical example used in the current experiment.

Frey and van de Rijt (2016, p. 1) observed the emergence of reputation cascades "that keep increasing the reputational advantage of one party while preventing others from building a reputation." In their study, trustors had no prior information about the trustees, so the inequality

[1] 19 subjects played the first game, but could not be matched to a group in the second game, and therefore played only one game.

that Frey and van de Rijt observed was based on the initial random choices of trustors in early rounds of the experiment. In the current study, I extended their design by providing the nationality of the trustees to all group members to evaluate the extent to which reputation systems may reinforce differences in outcomes between individuals with different nationalities.

Subjects participated in the control condition or in the reputation condition (between-subjects design). In the control condition, the trustors received no information about past decisions made by the other subjects. They only received information about the decision of the trustee they selected. In the reputation condition the trustors received full information about the decisions made by all other subjects in the group. This means that all players in a group could see which trustee was selected in a round, and what the decision of that trustee was. This information was visible in all subsequent rounds of the game. The supplementary materials contain an example of the interface of the trustors. In all conditions, trustees had full information about all decisions made by all players in their group.

In all groups, half of the trustees were United States citizens and half were Indian citizens. To test the theory, it is necessary to ensure intergroup inequality between the different groups of trustors. Trustors may have a range of possible preferences: They may all place more trust in the same group of trustees (e.g., both American and Indian trustors place more trust in Indian trustees than in American trustees). In that case the preferred group will be advantaged, regardless of the composition of the group of trustors. The decision of the trustors may also depend on the trustor's own nationality (i.e., homophily and heterophily). To also establish intergroup inequality in the case of homophily and heterophily, I created a minority-situation among the trustors in a group. Half of the groups had three American trustors and one Indian trustor, whereas the other groups had three Indian trustors and one American trustor. If individuals prefer others who are similar (different) to them, this minority-majority situation will result in an overall advantage of the majority (minority). Table 1 summarizes the treatment conditions.

Table 1: Summary of treatment conditions

Condition	**$N_{Subjects}$**	**N_{Groups}**	**Reputation information**	**Group composition**
Control	120	28	No information about past	Trustees: 2 US, 2 India
Majority US			decisions of other players	Trustors: 3 US, 1 India

Control Majority India	121	29	No information about past decisions of other players	Trustees: 2 US, 2 India Trustors: 1 US, 3 India
Reputation Majority US	135	28	Full information about past decisions made by other players	Trustees: 2 US, 2 India Trustors: 3 US, 1 India
Reputation Majority India	117	32	Full information about past decisions made by other players	Trustees: 2 US, 2 India Trustors: 1 US, 3 India

4.2 Procedure

In September and October 2019, 507 subjects completed the experiment; these subjects were divided into fourteen sessions (seven per condition) that each lasted for about 45 minutes. I recruited the subjects via Amazon Mechanical Turk (MTurk, 2019).[2] Compared to a student sample, there is more demographic variation among workers on MTurk. I used the fact that the large majority of MTurk workers were from the United States, followed by India (Difallah, Filatova, & Ipeirotis, 2018). I only allowed MTurk workers who were aged 18 or above, had one of these nationalities, had completed 1000 tasks or more on MTurk to participate in the experiment. In addition, I followed the procedure of Arechar, Gächter and Molleman (2018) and only allowed workers who had a task acceptance rate of 90%, to ensure that they would

[2] Following the process as described in Arechar, Gächter, and Molleman, (2018), each session was scheduled at a fixed date and time. Subjects could sign up for a specific session through a link that was posted on MTurk. A total of 780 respondents signed up and showed up for the experiment. 38 (5%) of them did not complete the instructions. 206 (26%) subjects who finished reading the instructions, but who could not be assigned to a group received a show-up fee of $2.00. Reasons for why respondents could not be assigned to a group were that the number or subjects who showed up was not a multiple of eight, or because there was a disbalance in the number of American and Indian respondent who showed up. Moreover, subjects received $2.00 when they waited more than 10 minutes for other respondents. In case a subject quit the experiment in the middle of a game, they did not receive a payment and the remaining group members received the points that they earned so far, plus the points they would have earned in the remaining rounds, based on their average number of points earned per round. This happened in six groups. In order to pay the subjects according to their earnings in the game, I had to collect their MTurk IDs. After completing the cleaning of the data, these IDs are stored separately from the experimental data.

complete the task with care. Subjects could only participate once.[3] 41.8% of the subjects were female. 49.5 % had Indian nationality and the remaining 50.5% had US nationality. The average age of the subjects was 36.5 years (*SD* = 10.6). Most subjects held a high school diploma (17.8%), an undergraduate degree (28.8%) or a graduate degree (43.2%). I conducted two pilot sessions.

All subjects received a show-up fee of $2. In addition, they received $1 for every 100 points they earned in the games, which resulted in a possible payment that ranged from $5.60 to $14.80. On average, subjects earned 6.96 dollars (256 points per game). I used oTree to program the experiment (Chen, Schonger, & Wickens, 2016).

At the beginning of the study the participants filled out a short survey about their age, gender and nationality. They read about the duration, payment, and general procedures of the experiment and they gave their consent to display their age, nationality and gender to other participants (see the supplementary materials for the full instructions). Their identity was never revealed, and the experiment used no deception. Subjects received on-screen instructions and could only proceed to the games if they correctly answered five comprehension questions to ensure that they had carefully read the instructions. I excluded six subjects who failed to reach a decision within the time limit (two minutes) from the remainder of the experiment, along with the groups of which they were a member.

4.3 Dependent variables

The main outcome variable is the inequality between trust placed in American and Indian trustees. Per group I calculated the proportion of trust placed in Indian trustees, which is the total number of times trust was placed in Indian trustees, divided by the total number of times trust was placed in a group. If there is no discrimination, the expected proportion of trust placed in Indian trustees would be the number of Indian trustees in a group divided by the total number of trustees in the group. Considering that in each group two of the trustees were American and the other two were Indian, trustors who randomly selected one of the trustees without taking their nationality into account would choose an Indian trustee half of the time. The more the proportion of trust placed in Indian trustees deviates from 0.5, the higher the inequality.

[3] Four subjects completed the study more than once. The data from the groups they were a part of in their second sessions have been excluded from the analyses.

.

Inequality had thus been calculated as the absolute difference between the proportion of trust placed in Indian trustees and 0.5. This difference is equal to the absolute value of half the value of the two-person Gini-coefficient.[4] The larger this difference, the more inequality there is in a group. This measure is robust to differences in discriminatory behaviors. It can be applied both to cases in which trustors select trustees who have the same or different nationality as they have, and to cases in which all trustors place more trust in American or Indian trustees, regardless of their own nationality. There were ten groups in which trust was either never placed or was placed only once — I excluded these from the analysis.

5. Results

5.1 Descriptive statistics

Figure 3 shows how inequality was distributed across groups in the different conditions. The bar in the middle indicates that the trustors in a group selected American and Indian trustees with the same frequency. The bars more to the left represent groups in which the trustors placed more often trust in Indian trustees than in American trustees. The bars more to the right represent groups in which the trustors placed more often trust in American trustees than in Indian trustees. The presence of a reputation system seemed to reduce the proportion of groups in which trust was equally divided between Indian and American trustees in groups where the majority of the trustors was American, but not in groups with an Indian trustor majority.

A generalized linear model with a binomial distribution and a logit link function and robust standard errors with the average fraction of trust placed in a the group as the dependent variable showed that there was no difference in the likelihood that American and Indian trustors placed trust (M_{US}= 54.3, M_{India}= 55.1, $z(115) = 1.08$, $p = 0.278$). The probability that trustors selected an American trustee does not significantly deviate from the probability that trustors selected an Indian trustee (M_{US}= 29.3, M_{India}= 23.9, $z(232) = -1.91$, $p = 0.057$). However, American trustees

[4] The inequality measure used here is the absolute value of 0.5 – S2 / (S1 + S2), where S1 is the frequency with which trust is placed in American trustees and S2 is the frequency with which trust is placed in Indian trustees. When this measure is multiplied by two, the result is |1 – 2S2 / (S1 + S2)|, which can also be written as |(S1 + S2)/(S1 + S2) – 2S2 / (S1 + S2)|, which is equal to the two-person Gini coefficient: |(S1 – S2) / (S1 + S2)|.

were on average more likely to honor trust (M_{US}= 61.4, M_{India}= 32.6, $z(193) = -4.68$, $p < 0.001$). Appendix A contains additional descriptive statistics.

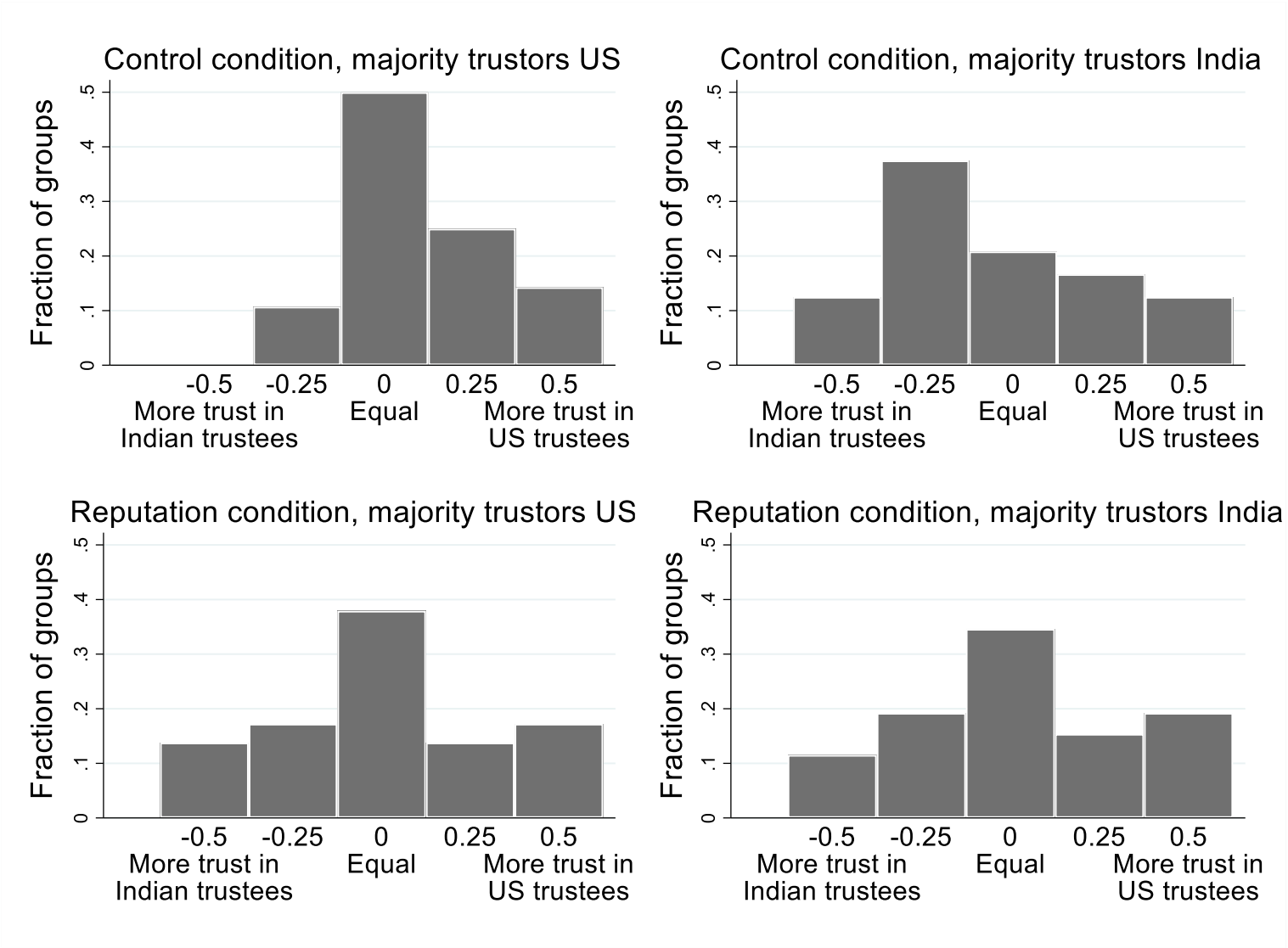

Figure 3. Distribution of inequality (defined as 0.5 – fraction of trust placed in Indian trustees) across groups in the different conditions.

5.2 Hypothesis test

A generalized linear model with a binomial distribution and a logit link function and robust standard errors at the level of the group revealed that the difference in inequality in the presence of a reputation system compared to the control condition was not significant ($M_{\text{Control}} = 0.21$, $M_{\text{Reputation}} = 0.26$, $z(105) = 1.50$, $p = 0.133$). This result does not change when separately analyzing the groups in which the majority of trustors was American ($M_{\text{Control}} = 0.18$, $M_{\text{Reputation}} = 0.25$, $z(55) = 1.61$, $p = 0.106$), and groups in which the majority of trustors was Indian ($M_{\text{Control}} = 0.24$, $M_{\text{Reputation}} = 0.27$, $z(48) = 0.48$, $p = 0.635$). I cannot reject the null hypothesis that there is no difference between the control condition and the reputation condition.[5, 6] Neither hypothesis 1 nor hypothesis 2 is accepted. This result is robust to using different measures of inequality (the results of the robustness checks are in Appendix B).

5.3 Exploratory analyses of the mechanisms

To get a better understanding of the main finding, I explored the extent to which the mechanisms outlined in the theory section (section 2) were at play in the experimental groups. I started by testing the compensation hypothesis for which earlier studies found evidence: In those studies the presence of reputational information reduced the importance of demographic information. After that I explored how choices of individual trustors and trustees led to the outcomes at the group level. The bottom part of Figure 3 shows that in the reputation condition in some groups trustors placed more trust in American trustees, while in other groups trustors placed more trust in Indian trustees. I conducted a number of analyses to explain what factors determine which group of trustees was advantaged over the other. I started with analyzing the prevalence of reputation cascades. After that I analyzed the extent to which the prevalence of reputation cascades predicted inequality. Finally, I studied whether the nationality of the first

[5] Because most subjects played two games, the groups were not completely independent. Moreover, the composition of some groups changed between the two games, while other groups remained the same. A regression with inequality as the dependent variable and with random intercepts for unique groups revealed that adding these intercepts did not significantly improve the model fit ($\chi 2$ (1) = 0.37, p = 0.270). Adding a random intercept for each session also did not improve the model fit ($\chi 2(1) = 0.12$, p = 0.366). This suggests that no significant proportion of variance was explained at the level of unique groups and sessions, so in the main text the results from the regressions without these random intercepts were reported.

[6] A power analysis shows that the experimental design would have a power of 0.8 to detect medium-size differences (Cohen's D of 0.55) between the conditions. If the reputation system has an effect on inequality, the effect is thus likely small.

selected trustee and the decision of that trustee affected the advantage of one group over the other.

5.3.1 The compensation effect

First, I explored the extent to which disadvantaged trustees benefitted more from reputation than already-advantaged trustees, because the presence of more relevant and specific reputational information reduced the importance of more diffuse demographic information (Figure 1, arrow A). This is the main argument of studies that propose that reputation systems may eliminate discrimination (e.g., Ert et al., 2016). To test if the experimental data supported this argument, I fitted McFadden's choice model for the reputation condition to analyze how the reputation of the trustee in the reputation condition affected the decision of the trustor. This model allowed me to study how the choice of the trustor depended on characteristics of the trustees and on variables that did not vary across trustees, such as the round number and the majority condition. Table 2 contains the results of these analyses. The models were constructed separately for American trustors (Models 1 and 2) and Indian trustors (Models 3 and 4). Table 2 only contains the outcomes for the variables of interest; the control variables are omitted. The full table including the control variables and the results for the same models of the control condition can be found in the Appendix C.

The choice of a trustor for one of four trustees in a round was the outcome variable. I excluded rounds in which no trustee was selected. The nationality of each trustee was included as an independent variable. I also included the number of times each trustee honored and abused trust in the past and I interacted these variables with the trustee's nationality to test if the presence of reputational information reduced the reliance on the nationality of the trustees. I also included the game and round number as control variables, and I specified that the residuals are clustered into trustors, because trustors made multiple decisions across the game(s) and these decisions were not independent.[7]

[7] Clustering for groups or sessions instead of individuals led to one slight change in the results of the models for Indian trustors: the coefficients of the frequency with which a trustee has abused trust in the past turned positive and significant. This does not alter the conclusions.

Table 2: Results of conditional logit models with the trustee choice of the trustor as dependent variable. Estimated logged odds (logit). Residuals clustered into trustors. Estimates for the control variables are omitted from this table. A table with full results and the results for the control condition can be found in the Appendix C.

	American trustor		**Indian trustor**	
	Model 1	Model 2	Model 3	Model 4
Main effects				
Trustee India	-0.87**	-0.84***	0.25	0.11
	(0.30)	(0.23)	(0.29)	(0.25)
Number of times trustee	0.76**	0.81***	0.29	0.28
honored trust	(0.25)	(0.17)	(0.21)	(0.15)
Number of times trustee	-2.72***	-2.83**	0.42	-0.30
abused trust	(0.72)	(0.89)	(0.22)	(0.54)
Interactions				
Trustee India * Number of	0.10	-	0.02	-
times trustee honored trust	(0.32)		(0.30)	
Trustee India * Number of	-	0.47	-	0.97
times trustee abused trust		(1.32)		(0.54)
Number of decisions	548	548	524	524
Number of trustors	137	137	131	131

Note: Standard error in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Supporting the findings from the Figure 3, the models show that American trustors placed more trust in American trustees, while Indian trustors did not condition their decision on the nationality of the trustee. American trustors placed more trust in trustees who had honored more trust in the past, and less trust in trustees who had abused more trust in the past. Indian trustors placed more trust in trustees who had honored more trust in the past, but not less trust in trustees who had abused more trust in the past, compared to trustees who had neither honored nor abused trust in the past. None of the interactions between the trustee's nationality and their reputation is significant. This means that there is no evidence for the compensation effect: The disadvantage for an Indian trustee was not smaller when that trustee had a more positive reputation. These findings do not support the argument from earlier studies that the presence of more relevant and specific information reduced the reliance on demographic information.

5.3.2 The effect of reputation cascading on inequality

I now move to a discussion of the mechanisms that may explain why reputation systems may not be an effective instrument to decrease inequality. The first mechanism is reputation cascading (Figure 1, arrow D). I analyzed the prevalence of reputation cascades in the reputation condition by estimating the probability that a cascade was continued. Consistent with the analysis by Frey and van de Rijt (2016), I used a logistic regression with clustering for trustors and a variable indicating whether the trustor selected the trustee that was selected in the previous round. I excluded the first round of each game, rounds in which the trustor did not place trust in the previous round, or if the trustee abused trust in the previous round. If trustors would randomly select a trustee, the probability that a trustor selected the same trustee as was chosen on the last turn the trustor observed should be 25%, since there were four trustees in every group. In the control conditions this probability was indeed not significantly different from 25% (27.6%, $z = 0.430$, $p = 0.667$). The observed probability was 52.5% in groups with an American trustor majority in the reputation condition, which is significantly higher than 25% ($z = 3.554$, $p < 0.001$). This means that in these groups the presence of a reputation systems led to reputation cascades. In groups with an Indian majority the observed probability was not significantly higher than 25% (31.3%, $z = 0.828$, $p = 0.408$) in the reputation condition, so I did not find evidence of reputation cascades in those groups. The probability that a cascade was continued was about 11.8% higher when the cascade length increases with one in the groups in which the majority of the trustors was American ($z = 2.08$; $p = 0.037$). The nationality of the reputable trustee did not affect the cascade continuation probability in those groups (mean difference = 15.8 percent point; $z = 1.38$; $p = 0.168$).

Next, I used a regression analysis at the group level to test if a higher likelihood of cascade continuation was related to more inequality (Figure 1, arrow C and D). I used inequality in a group as the dependent variable and the probability that a cascade was continued as an independent variable. I controlled for treatment (control or reputation), trustor majority (US or India) and the game number (first or second). I found evidence that there was more inequality in groups with a higher likelihood of cascade continuation ($b = 0.136$, $t = 2.75$, $p = 0.008$).

The presence of reputation cascades led to the prediction that the nationality of the first trustee who was trusted and honored trust in a group had a large influence on the choices of future trustors, because initial differences may be reinforced through accumulation of reputation (Figure 1, arrow A). A proportions test showed that when trust was placed for the first time in

the reputation condition, American trustors were more likely to select an American trustee (American trustees: 79%, $z = 3.024$, $p = 0.003$), while Indian trustors did not differentiate between American and Indian trustees (Indian trustees: 52%, $z = 0.186$, $p = 0.853$). Figure 4 shows the relation between the nationality of the first trustee who honored trust and the nationality of the trustee who was chosen in subsequent rounds. A generalized linear model with a binomial distribution and a logit link function and robust standard errors at the group level showed that the difference in inequality between American and Indian trustees in the remaining rounds was significantly affected by the nationality of the first trustee who honored trust (mean difference = 0.336, $z(44) = 3.09$, $p = 0.002$). When the first trustee who honored trust was Indian, the probability that future trustors also select an Indian trustee was higher than when the first trustee who honored trust was American (and vice versa).

Together these findings suggest that in some cases trustees belonging to the majority benefitted more from the reputation system, while in other cases trustees belonging to the minority benefitted more. Which group benefitted more depended on the nationality of the trustee who first honored trust in a group. The selection of this first trustee depended on the nationality of the first trustor who placed trust in a group, which in turn depended on the composition of the group: in 69% of the rounds the first trustor who placed trust belonged to the majority (69.3% ($z(113) = 4.121$, $p < 0.001$). These results supported the theory that differences in group composition, discrimination, and cascading of reputation may lead to reinforcement of intergroup inequality.

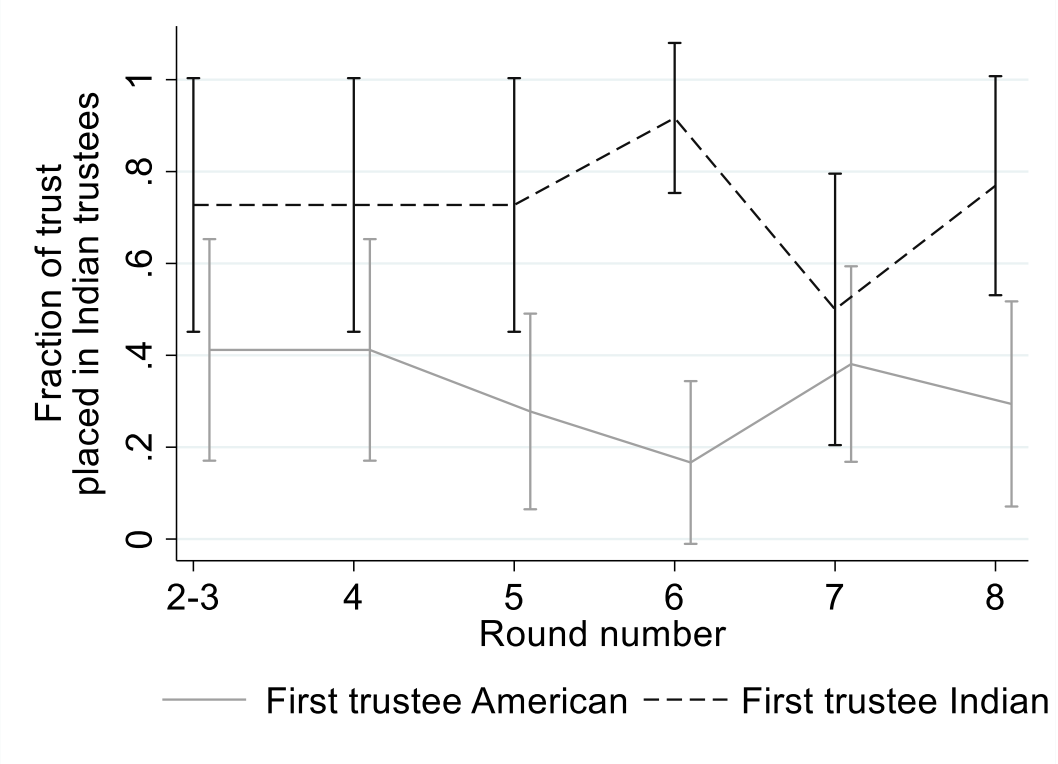

Figure 4. Relation between the nationality of the first trustee who honored trust in a group and proportion of trust placed in Indian trustees in the remaining rounds. Groups in which trust was placed never or only once are excluded.

6. Discussion

Reputation systems are often argued to be the most promising solution to discrimination in the platform economy, because the presence of ratings and reviews may decrease the importance of demographic information (Abrahao et al., 2017; Cui et al., 2020; Ert et al., 2016; Tjaden et al., 2018). However, this argument implicitly assumes that reputation scores are exogenously generated. In this paper I showed that reputation systems are not free from bias. Reputation building is a sequential process: Decisions made early on affect decisions made later. In the presence of a reputation system some individuals will accumulate reviews, while others will not. If individuals with a certain nationality structurally get more opportunities for reputation building than individuals with a different nationality, their initial advantage may accumulate and result in a sustained or even larger advantage. Using an online experiment, I found that the presence of a reputation system failed to decrease intergroup inequality when demographic information is available. This finding is robust to using different measures of inequality.

Moreover, decisions made in initial rounds of the experiment determine to a large extent which individuals will be trusted more later on.

The presence of a reputation system did not decrease inequality and the exploratory analyses revealed that the mechanisms pertaining to reputation systems worked as predicted. The difference in the probability of being trusted between Indian and American trustees in the remaining rounds strongly depended on the nationality of the first trustee who honored trust in a group, because reputation tends to cascade. The higher the likelihood that a cascade is continued in a group, the larger the inequality in that group.

These results support the theory that reputation systems may sustain or even increase intergroup inequality. While earlier studies concluded that reputation systems may decrease inequality (Abrahao et al., 2017; Cui et al., 2020; Ert et al., 2016; Tjaden et al., 2018), they solely looked at the effect of having a positive reputation, and they overlooked the endogenous nature of the reputation building process. I argue that the current experiment is a better test of the effect of reputation systems on inequality, because it took the static effect of the presence of reputation information on the reliance on demographic information into account, as well as differences in the speed with which different groups collect reputation scores.

A study by Kas et al. (2021) in which the endogenous nature of reputation was accounted for did not have data on a control group without a reputation system, and could therefore only study the change in inequality over time when there is a reputation system. According to the theory, inequality is likely to be increased by the presence of a reputation system if the initial level of trust in disadvantaged individuals is low (Kas et al., 2021). In the current experiment the observed trust levels were relatively high, which may explain why the presence of a reputation system was not linked to less intergroup inequality, but also not more. Other reasons for why the reputation system did not lead to an increase in inequality may be that Indian trustees were less likely to honor trust than American trustees, while the theory assumes that all trustees are equally likely to honor trust. One explanation for this latter finding may be that Indian trustees had a smaller incentive to honor trust because they were less likely to be selected, and therefore had a shorter shadow of the future. To avoid these differences in behavior between different users, I could have simulated trustee responses. However, using simulations would imply deceiving the subjects and would lower the external validity of the study, because in real markets there may also be differences between trustees. A final explanation for why intergroup inequality was similar and not higher when there was a

reputation system may be the limited number of observations per condition, caused by measuring inequality at the group level.

I also provided another test of the "compensation hypothesis." Earlier studies found that the presence of reputational information reduced the reliance on demographic information, and concluded that reputation systems must therefore reduce intergroup inequality. I did not find that the presence of reputational information reduced the importance of demographic information; this ran contrary to the findings of earlier experiments (Abrahao et al., 2017; Ert et al., 2016), one study with field data (Tjaden et al., 2018), and one field experiment (Cui et al., 2020), but similar to another field study (Kas et al., 2021). The disadvantage experienced by Indian trustees was not smaller for trustees with a good reputation than for trustees with no reputation. Part of the explanation for the difference between my findings and the findings of some of the earlier studies may be that earlier studies did not look at ethnicity or nationality but at other demographic characteristics that are associated with less strong prejudices and that can more easily be overruled with better information. (Abrahao et al., 2017). Other studies did not look at actual trusting decisions but at behavior that did not have real consequences, such as the number of clicks on an offer (Tjaden et al., 2018). Finally, one other study used reputation scores that were experimentally varied to increase their variance (Ert et al., 2016), which artificially carries users over the initial hurdle of getting the first positive review and inflates the relative informativeness of ratings.

This study is the first to compare the effect of the presence of a reputation system on intergroup inequality when demographic information is available. The experimental setting allowed me to compare the situation in which there is a reputation system with a situation without a reputation system. This would not be possible when using data from real platforms, since platforms either have or do not have a reputation system, and there are too many differences between platforms to compare inequality across them. The subjects who participated in the experiment likely behaved differently than real platform users. They may have responded differently to in- and outgroup members and they may have attached different value to reputational information. Similarly, the choice for American and Indian subjects likely affected the magnitude and type of discrimination. In the current study American trustors placed more trust in American trustees than in Indian trustees, but Indian trustors did not differentiate based on the nationality of the trustee. There are few studies in which discrimination of Indians by Americans and vice versa is studied. In an earlier study, the majority of US-born Indian Americans indicate that they have been discriminated the last year (Badrinathan, Kapur, Kay,

& Vaishnav, 2021). However, most of them also believe that other groups in the American society (e.g. African Americans, Latino Americans) are discriminated more than they are. Choosing other groups (e.g. white and Black Americans) would possibly have resulted in higher levels of initial discrimination.

However, this does not mean that we cannot draw any conclusions about real-life platforms and about other (ethnic) groups from the results. While the preferences and behaviors of individual subjects may be different, the theory is not about these preferences per se but rather about the dynamics that lead to a reinforcement of or a decrease in inequality. These dynamics are a product of individual behavior. The theory is applicable to different types of discrimination (i.e., statistical and taste-based discrimination), different combinations of groups (i.e., discriminatory patterns may be different when studying different nationalities) and to discrimination based on different characteristics (e.g., gender, ethnicity, or age). In the current study Indian trustors did not have a strong preference for Indian or American trustees, and they based their decision only to a very limited extent on the reputations of the trustees. Based on simulations, Kas et al. (2021) theorized that reputation systems may only lead to an increase in inequality if subjects strongly rely on reputational information and when initial differences between individuals with different demographic characteristics are large enough. The limited inequality between Indian and American trustees in groups with more Indian than American trustors may be a consequence of the limited reliance of Indian trustors on reputational or demographic information. Replicating this study with groups of respondents who are less trustful, who discriminate more and who rely more on reputation information may lead to a stronger reinforcement of inequality.

Overall, the experimental results suggest that reputation systems do not reduce intergroup inequality between individuals with different demographic backgrounds. Intergroup differences may be further reinforced by the establishment of norms legitimizing or mimicking the existing social order (Berger, Vogt, & Efferson, 2021; Costa-Lopes, Dovidio, Pereira, & Jost, 2013) and by the common use of ranking algorithms that favor individuals with a better reputation, such as the algorithm used on Airbnb (Fradkin, Grewal, & Holtz, 2018). Platforms that wish to create equal opportunities for users with different backgrounds should not only try to have an effective reputation system, but also reduce initial differences in the chances for individuals with different backgrounds to participate. They could, for example, do so by making demographic information less visible in the initial booking process – as Airbnb recently started doing and for which there is some evidence that it is effective (Mohammed, 2017). In

general, platforms should be aware that reputation systems are an effective way of increasing interpersonal trust but that they do not benefit every user to an equal extent. Reputation systems help those who obtained at least one review but may negatively affect those individuals who have not participated in interactions before. This means that reputation systems may reinforce initial differences between individuals. In the best case scenario, these initial differences will be random and the resulting inequality will also be randomly distributed among individuals and groups (Frey & Van De Rijt, 2016). In a more realistic scenario, initial differences will be based on structural differences between individuals, such as their demographic characteristics, and the presence of a reputation system may perpetuate these differences.

References

Abrahao, B., Parigi, P., Gupta, A., & Cook, K. S. (2017). Reputation offsets trust judgments based on social biases among Airbnb users. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(37), 9848–9853.

Ahuja, R., & Lyons, R. C. (2019). The Silent Treatment: LGBT Discrimination in the Sharing Economy. *Oxford Economic Papers*, *71*(3), 564–576.

Akerlof, G. (1970). The market for lemons: Qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*(3), 235–251.

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, *21*(1), 99–131.

Arrow, K. (1973). The Theory of Discrimination. *Discrimination in Labor Markets*, *3*(10), 3–33.

Badrinathan, S., Kapur, D., Kay, J., & Vaishnav, M. (2021). *Social Realities of Indian Americans: Results From the 2020 Indian American Attitudes Survey*.

Becker, G. S. (1957). *The Economics of Discrimination.* Chicago: University of Chicago Press.

Berger, J., Vogt, S., & Efferson, C. (2021). Pre-existing fairness concerns restrict the cultural evolution and generalization of inequitable norms in children. *Evolution and Human Behavior*.

Boero, R., Bravo, G., Castellani, M., & Squazzoni, F. (2009). Reputational cues in repeated trust games. *The Journal of Socio-Economics*, *38*(6), 871–877.

Bol, T., de Vaan, M., & van de Rijt, A. (2018). The Matthew effect in science funding. *PNAS*, *115*(19), 4887–4890.

Bolton, G. E., Katok, E., & Ockenfels, A. (2004). How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation. *Management Science*, *50*(11), 1587–1602.

Buskens, V., & Raub, W. (2002). Embedded trust: Control and learning. *Group Cohesion, Trust and Solidarity*, *19*, 167–202.

Carol, S., Eich, D., Keller, M., Steiner, F., & Storz, K. (2019). Who can ride along? Discrimination in a German carpooling market. *Population, Space and Place*, *25*(8), 1–15.

Charness, G., Du, N., & Yang, C. L. (2011). Trust and trustworthiness reputations in an investment game. *Games and Economic Behavior*, *72*(2), 361–375.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Cook, K. S., Hardin, R., & Levi, M. (2005). *Cooperation Without Trust?* New York: Russell Sage Found.

Costa-Lopes, R., Dovidio, J. F., Pereira, C. R., & Jost, J. T. (2013). Social psychological perspectives on the legitimation of social inequality: Past, present and future. *European Journal of Social Psychology*, *43*(4), 229–237.

Cui, R., Li, J., & Zhang, D. J. (2020). Reducing Discrimination with Reviews in the Sharing Economy. *Management Science*, *66*(3), 1071–1094.

Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and dynamics of Mechanical Turk workers. In *WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining* (pp. 135–143).

Dubois, D., Willinger, M., & Blayac, T. (2012). Does players' identification affect trust and reciprocity in the lab? *Journal of Economic Psychology*, *33*(1), 303–317.

Duffy, J., Xie, H., & Lee, Y. J. (2013). Social norms, information, and trust among strangers: theory and evidence. *Economic Theory*, *52*(2), 669–708.

Edelman, B., & Luca, M. (2014). *Digital Discrimination: The Case of Airbnb.com* (Harvard Business School NOM Unit Working paper No. 14–054).

Edelman, B., Luca, M., & Svirsky, D. (2017). Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics*, *9*(2), 1–22.

Ert, E., Fleischer, A., & Magen, N. (2016). Trust and Reputation in the Sharing Economy:

The Role of Personal Photos on Airbnb. *Tourism Management*, *55*, 62–73.

Fehrler, S., & Przepiorka, W. (2013). Charitable giving as a signal of trustworthiness: Disentangling the signaling benefits of altruistic acts. *Evolution and Human Behavior*, *34*(2), 139–145.

Fradkin, A., Grewal, E., & Holtz, D. (2018). *The Determinants of Online Review Informativeness: Evidence from Field Experiments on Airbnb* (No. 2939064). *SSRN Electronic Journal*.

Frey, V., & Van De Rijt, A. (2016). Arbitrary Inequality in Reputation Systems. *Scientific Reports*, *6*(1), 1–5.

Gambetta, D. (2009). Signaling. In *The Oxford Handbook of Analytical Sociology* (pp. 168–194). Oxford: Oxford University Press.

Ge, Y., Knittel, C. R., MacKenzie, D., & Zoepf, S. (2016). *Racial and gender discrimination in transportation network companies* (No. w22776). *NBER Working Paper Series*.

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, *18*(12), 1192–1217.

Jaeger, B., & Sleegers, W. (2020). *Racial discrimination in the sharing economy: Evidence from Airbnb markets across the world* (Working paper). PsyArXiv. https://doi.org/10.31234/osf.io/qusxf

Jaeger, B., & van Beest, I. (2019). The effects of facial attractiveness and trustworthiness in the sharing economy. *Journal of Economic Psychology*, *75*, 102125.

Jiao, R., Przepiorka, W., & Buskens, V. (2021). Reputation effects in peer-to-peer online markets: A meta-analysis. *Social Science Research*, *95*, 102522.

Kas, J., Corten, R., & van de Rijt, A. (2021). The role of reputation systems in digital discrimination. *Socio-Economic Review*.

Katz, V. (2015). Regulating the Sharing Economy. *Berkeley Technology Law Journal*, *30*(4), 11–29.

Kuwabara, K. (2015). Do Reputation Systems Undermine Trust? Divergent Effects of Enforcement Type on Generalized Trust and Trustworthiness. *American Journal of*

Sociology, *120*(5), 1390–1428.

Laouénan, M., Rathelot, R., Autor, D., Bagues, M., Boustan, L., De Chaisemartin, C., … Zinovyeva, N. (2017). *Ethnic Discrimination on an Online Marketplace of Vacation Rental Ethnic Discrimination on an Online Marketplace of Vacation Rentals*.

Lauterbach, D., Truong, H., Shah, T., & Adamic, L. (2009). Surfing a web of trust: Reputation and reciprocity on Couchsurfing.com. In *2009 International Conference on Computational Science and Engineering* (Vol. 4, pp. 346–353).

Macy, M. W., & Skvoretz, J. (1998). The Evolution of Trust and Cooperation between Strangers: A Computational Model. *American Sociological Review*, *63*(5), 638–660.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415–444.

Merton, R. K. (1968). The Matthew Effect in Science. *Science*, *159*(3810), 55–63.

Mohammed, A. R. S. (2017). *Designing for Racial Impartiality: The Impact of Relocating Host Photos on the Airbnb Website* (Working paper).

MTurk. (2019). Retrieved August 26, 2020, from https://www.mturk.com/

Murphy, L. (2016). *Airbnb's Work to Fight Discrimination and Build Inclusion: A Report Submitted to Airbnb*.

Nunley, J. M., Owens, M. F., & Howard, R. S. (2011). The effects of information and competition on racial discrimination: Evidence from a field experiment. *Journal of Economic Behavior & Organization*, *80*(3), 670–679.

Parigi, P., Santana, J. J., & Cook, K. S. (2017). Online Field Experiments: Studying Social Interactions in Context. *Social Psychology Quarterly*, *80*(1), 1–19.

Pope, D. G., & Sydnor, J. R. (2011). What's in a Picture? Evidence of Discrimination from Prosper.com. *Journal of Human Resources*, *46*(1), 53–92.

Przepiorka, W., & Berger, J. (2017). Signalling theory evolving: signals and signs of trustworthiness in social exchange. In *Social Dilemmas, Institutions and the Evolution of Cooperation.* (pp. 373–392). Berlin: De Gruyter Oldenbourg.

Przepiorka, W., & Diekmann, A. (2013). Temporal embeddedness and signals of trustworthiness: Experimental tests of a game theoretic model in the United Kingdom, Russia, and Switzerland. *European Sociological Review*, *29*(5), 1010–1023.

Raub, W. (2004). Hostage Posting as a Mechanism of Trust: Binding, Compensation, and Signaling. *Rationality and Society*, *16*(3), 319–365.

Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation systems. *Communications of the ACM*, *43*(12), 45–48.

Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in internet transactions: Empirical analysis of eBay' s reputation system. *Advances in Applied Microeconomics*, *11*(2), 127–157.

Robbins, B. G. (2017). Status, identity, and ability in the formation of trust. *Rationality and Society*, *29*(4), 408–448.

ter Huurne, M., Ronteltap, A., Corten, R., & Buskens, V. (2017). Antecedents of trust in the sharing economy: A systematic review. *Journal of Consumer Behaviour*, *16*(6), 485–498.

ter Huurne, M., Ronteltap, A., Guo, C., Corten, R., Buskens, V., ter Huurne, M., … Buskens, V. (2018). Reputation Effects in Socially Driven Sharing Economy Transactions. *Sustainability*, *10*(8), 2674.

Teubner, T. (2017). The web of host-guest connections on Airbnb - A social network perspective. *Journal of Systems and Information Technology*, *20*(3), 262–277.

Tjaden, J. D., Schwemmer, C., & Khadjavi, M. (2018). Ride with me - Ethnic discrimination in Social Markets. *European Sociological Review*, *34*(4), 418–432.

Weigelt, K., & Camerer, C. (1988). Reputation and corporate strategy: A review of recent theory and applications. *Strategic Management Journal*, *9*(5), 443–454.

Wozniak, D., & MacNeill, T. (2020). Racial discrimination in the lab: Evidence of statistical and taste-based discrimination. *Journal of Behavioral and Experimental Economics*, *85*, 1–16.

Wu, J., & Jin, M. (2018). Signaling Peer Trust in Accommodation-sharing Services: Effects

of Similarity and Reviews on Listing Sales. In *Wuhan International Conference on e-Business*.

Wu, J., Ma, P., & Xie, K. L. (2017). In sharing economy we trust: the effects of host attributes on short-term rental purchases. *International Journal of Contemporary Hospitality Management*, *29*(11), 2962–2976.

Appendix A: Additional descriptive statistics

All reported tests are conducted at the level of the group. An independent samples T-test shows that there is neither a significant difference in the average trust rate between the control condition and the reputation condition (M_{Control} = 50.4, $M_{\text{Reputation}}$ = 55.8, $z(115) = 1.08$, p = 0.278), nor a difference in trustworthiness rate (M_{Control} = 50.3, $M_{\text{Reputation}}$ = 56.9, $z(112) = 1.04$, p = 0.301). This suggests that the reputation system did not lead to an overall increase in trust and trustworthiness. However, this does not imply that reputation did not affect decision making at the individual level, as will be shown in the additional exploratory analyses. A potential explanation for the lack of an overall effect of the presence of a reputation system on trust and trustworthiness may be that trust was already relatively high in the control condition, possibly because of the presence of information about the trustee's nationality. Another reason may be that the number of rounds in the game was limited, limiting the shadow of the future and therefore the increasing effect of the reputation system on trust and trustworthiness. Both explanations imply that in situations where the reputation system has a larger effect on trust, the reputation system may have a stronger effect on inequality – thereby both reducing the relevance of demographic information, and increasing the prevalence of reputation cascades. Finally, this result is based on the most conservative test possible. Although the effect goes in the expected direction, the difference is not significant, possibly because of the limited sample size due to reporting T-tests at the group level.

Table A1 shows the average trust level for American and Indian trustees and the average inequality per group in the different conditions. Table A2 shows per condition the average trustworthiness level and success rate (the fraction of interactions in which trust is placed and honored) for American and Indian trustees. The average fraction of rounds in which trustees are trusted across all conditions does not significantly differ between American and Indian trustees (Table A1, bottom row, $z(232) = -1.91$, p = 0.057), but American trustees are on average more likely to honor trust (Table A2, bottom row, $z(193) = -4.68$, $p < 0.001$). Overall, this results in a much higher success rate for American trustees than Indian trustees (Table A2, bottom row, $z(232) = -3.82$, $p < 0.001$), but not in a difference in the average payoff of American trustees and Indian trustees in a group (M_{US}= 34.4, M_{India}= 34.6, $t(106) = -0.274$, p = 0.785), because abusing trust yields trustees a higher (immediate) payoff than honoring trust. American and Indian trustors are equally likely to place trust (M_{US}= 54.3, M_{India}= 55.1, $z(232)$ = 0.18, p = 0.856).

Table A1: Average trust rate per group for Indian and American trustees, inequality rate per group, by condition

Condition		**N**	**Trust rate US trustees**	**Trust rate Indian trustees**	**N**	**Inequality**
Control	*Majority trustors US*	29	32.8 (17.2)	21.1 (17.7)	28	0.18 (0.17)
	Majority trustors India	28	23.2 (15.9)	23.7 (17.5)	24	0.24 (0.18)
	Total	57	28.1 (17.1)	22.4 (17.5)	52	0.21 (0.18)
Reputation	*Majority trustors US*	32	29.3 (23.9)	26.6 (25.9)	29	0.25 (0.18)
	Majority trustors India	28	31.7 (23.9)	24.1 (25.0)	26	0.27 (0.19)
	Total	60	30.4 (23.7)	25.4 (25.3)	55	0.26 (0.18)
Total		117	29.3 (21.7)	23.9 (21.8)	107	0.23 (0.18)

Note: Standard deviations in parentheses.

Table A2: Average trustworthiness and success rate for Indian and American trustees per group, by condition

Treatment		N	TW rate US trustees	N	TW rate Indian trustees	N	Success US trustees	Success Indian trustees
Control	*Majority trustors US*	28	60.7 (42.5)	25	26.4 (36.3)	29	23.3 (21.8)	6.9 (11.4)
	Majority trustors India	24	63.5 (42.9)	22	35.4 (37.2)	28	14.7 (14.9)	8.9 (10.7)
	Total	52	62.0 (42.3)	47	30.6 (36.6)	57	19.1 (19.1)	7.9 (10.9)
Reputation	*Majority trustors US*	27	51.6 (43.7)	25	41.4 (42.9)	32	19.9 (24.8)	15.2 (22.6)
	Majority trustors India	23	71.4 (37.3)	21	26.7 (40.7)	28	25.0 (24.3)	9.8 (21.1)
	Total	50	60.7 (41.7)	46	34.7 (42.1)	60	22.3 (24.5)	12.7 (21.9)
Total		102	61.4 (41.8)	93	32.6 (39.3)	117	20.7 (22.0)	10.4 (17.5)

Note: Standard deviations in parentheses.

Appendix B: Robustness checks

There are various ways to quantify inequality. I tested if the result of the main hypothesis test is robust to using other commonly used inequality measures. The results change neither when using the Herfindahl index (M_{Control} = 0.65, $M_{\text{Reputation}}$ = 0.70, $z(105) = 1.46$, $p = 0.145$), nor when a using the logged difference in the number of rounds that Indian and American trustees were expected to be selected, obtained from a negative binomial regression of the counts of trust placed in each trustee (M_{Control} = 4.06, $M_{\text{Reputation}}$ = 5.81, Wilcoxon rank-sum test: $z = -1.299$, $p = 0.194$).

In the main analyses I used an absolute measure of inequality. This measure does not make a distinction between groups in which American trustees were advantaged and groups in which Indian trustees were advantaged. When using a signed inequality measure (0.5 – fraction of trust placed in Indian trustees) the results neither changed when looking at all groups together (M_{Control} = 0.072, $M_{\text{Reputation}}$ = 0.072, $t(112) = 0.003$, $p = 0.997$), nor when only looking at groups with an American majority (M_{Control} = 0.117, $M_{\text{Reputation}}$ = 0.054, $t(58) = 0.855$, $p = 0.396$), nor when looking at groups with an Indian majority (M_{Control} = 0.026, $M_{\text{Reputation}}$ = 0.094, $t(52) = -0.757$, $p = 0.453$).

The number of rounds in which trust is placed differs across groups. In groups in which trust is placed in an uneven number of rounds, Indian and American trustees cannot be trusted an equal number of times. The number of rounds in which trust is placed thus affects the range of values the inequality variable can take. To account for this, I constructed a corrected inequality variable. For groups in which trust was placed an even number of times there is no difference between the original and the corrected inequality measure. For groups in which trust was placed an uneven number of times, the corrected inequality variable took a value of 0 if the fraction of trust placed in Indian trustees could not be closer to 0.5. For example, in groups in which trust was placed five times, the corrected inequality variable was 0 when Indian trustees were trusted two or three times. When the fraction of trust placed in Indian trustees did not fall in this range, corrected inequality was calculated as the distance between the fraction of trust placed in Indian trustees and the closest boundary of that range, divided by the size of the range of values that were considered to be unequal (2 * 0.4 in the case of groups in which trust was placed in five rounds). The results do not change when using this new corrected inequality variable instead of the original inequality variable: there is no significant difference in inequality between two conditions (M_{Control} = 0.18, $M_{\text{Reputation}}$ = 0.23, $z(105) = 1.21$, $p = 0.227$).

The inequality measure that is used in the main hypothesis test is based on all rounds in which trust was placed in a group. This is a rather conservative calculation of inequality, because it takes time for individuals to build a reputation, so the effect of the reputation system on inequality is expected to be stronger in later rounds. When constructing the inequality variable without the first round in which trust is placed in a group, the difference between the control condition and the reputation condition is not significant ($M_{\text{Control}} = 0.21$, $M_{\text{Reputation}} = 0.26$, $z(105) = 1.35$, $p = 0.176$). Also when the second round (MControl = 0.26, MReputation = 0.33, $z(82) = 0.334$, p = 0.123) and third round ($M_{\text{Control}} = 0.34$, $M_{\text{Reputation}} = 0.36$, $z(71) = 0.55$, $p = 0.585$) in which trust is placed are excluded, there is still not more inequality in groups in the reputation condition. When excluding more rounds, the number of observations is not sufficient to draw any conclusions.

Reputation cascades are not expected to emerge when trust is abused. When only including rounds in which trust was placed and honored in the calculation of inequality within groups, inequality is higher than when also including rounds in which trust was not honored (mean difference = 14.4 percent point, z(101) = 3.98, $p < 0.001$), suggesting the cascading of reputation led to inequality between trustees of different nationality. Average inequality between trustors who honored trust is 0.35 in the control condition and 0.41 in the reputation condition. This difference is not significant at the 95% confidence level ($z(86) = 1.54$, $p = 0.122$).

Overall, these results suggest that the finding that reputation systems do not reduce inequality is robust to using different measures of inequality.

Appendix C: Full results conditional logit models

The full results of the conditional logit models with the choice of the trustor as the dependent variable are reported in Table C1.

Table C1: Results conditional logit models with the choice of the trustor as dependent variable. Estimated logged odds (logit). Clustered

	American trustor				**Indian trustor**			
	Control	**Reputation**			**Control**	**Reputation**		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Main effects								
Trustee India	-0.77***	-0.82***	-0.87**	-0.84***	0.15	0.26	0.25	0.11
	(0.22)	(0.22)	(0.30)	(0.23)	(0.21)	(0.24)	(0.29)	(0.25)
Number of times trustee	-	0.81***	0.76**	0.81***	-	0.30*	0.29	0.28
honored trust		(0.16)	(0.25)	(0.17)		(0.14)	(0.21)	(0.15)
Number of times trustee	-	-2.70***	-2.72***	-2.83**	-	0.42	0.42	-0.30
abused trust		(0.71)	(0.72)	(0.89)		(0.22)	(0.22)	(0.54)
Trustee India * Number of	-	-	0.10	-	-	-	0.02	-
times trustee honored trust			(0.32)				(0.30)	
Trustee India * Number of	-	-	-	0.47	-	-	-	0.97
times trustee abused trust				(1.32)				(0.54)
Second trustee								
Game number	0.09	1.28**	1.27**	1.28**	-0.76	-1.36*	-1.37*	-1.32*
	(0.51)	(0.49)	(0.49)	(0.49)	(0.65)	(0.54)	(0.54)	(0.55)
Round	0.24	0.07	0.06	0.07	-0.05	0.04	0.04	0.03
	(0.13)	(0.14)	(0.15)	(0.14)	(0.13)	(0.11)	(0.12)	(0.12)
Majority trustors India	0.84	-1.26*	-1.25*	-1.26*	-0.47	-0.40	-0.39	-0.32
	(0.65)	(0.61)	(0.62)	(0.61)	(0.63)	(0.61)	(0.63)	(0.61)
Constant	-1.50	-2.25*	-2.21*	-2.23*	1.92	2.35*	2.35*	2.19*
	(0.98)	(0.99)	(0.99)	(0.98)	(1.26)	(1.13)	(1.13)	(1.10)

(Continued on next page)

Table C1: continued

	American trustor				Indian trustor			
	Control	**Reputation**			**Control**	**Reputation**		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Third trustee								
Game number	0.20	0.32	0.33	0.32	-0.84	-0.92	-0.92	-0.83
	(0.46)	(0.66)	(0.67)	(0.66)	(0.61)	(0.49)	(0.49)	(0.49)
Round	0.26*	0.11	0.11	0.11	-0.01	0.15	0.15	0.11
	(0.11)	(0.17)	(0.17)	(0.17)	(0.11)	(0.11)	(0.11)	(0.11)
Majority trustors India	0.46	-1.08	-1.08	-1.08	-0.67	-0.19	-0.19	-0.17
	(0.76)	(0.84)	(0.83)	(0.84)	(0.70)	(0.63)	(0.62)	(0.61)
Constant	-1.42	-1.79	-1.79	-1.81	2.21	1.08	1.08	0.95
	(1.11)	(1.34)	(1.33)	(1.34)	(1.34)	(1.11)	(1.11)	(1.06)
Fourth trustee								
Game number	0.43	1.01	0.99*	1.02*	-1.1	-0.01	-0.01	0.14
	(0.57)	(0.51)	(0.50)	(0.51)	(0.79)	(0.62)	(0.62)	(0.63)
Round	0.22	0.25	0.24	0.25	0.26*	0.23	0.23	0.22
	(0.12)	(0.14)	(0.15)	(0.14)	(0.12)	(0.12)	(0.12)	(0.12)
Majority trustors India	1.37	0.25	0.26	0.24	-0.67	-1.02	-1.02	-0.87
	(0.81)	(0.64)	(0.65)	(0.64)	(0.73)	(0.62)	(0.63)	(0.63)
Constant	-2.22*	-3.08**	-3.02**	-3.11**	0.66	-0.37	-0.37	-0.76
	(1.09)	(1.13)	(1.11)	(1.14)	(1.59)	(1.33)	(1.32)	(1.40)
Number of decisions	480	548	548	548	440	524	524	524
Number of trustors	120	137	137	137	110	131	131	131

Note: Standard error in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

The effect of online reputation systems on intergroup inequality

Judith Kas

Supplementary materials

Preregistration report & Instructions participants

Preregistration report

Submitted to: OSF Registries at September 11, 2019 (also available at: https://osf.io/mkcxh/?view_only=d4d57caa29fc4d0db680156d188b0a69)

Title

The reinforcement of ethnic inequality through reputation systems.

Authors

Judith Kas

Description

Discrimination is a widely spread and persistent problem that is increasingly studied in the context of online peer-to-peer markets (e.g. Edelman & Luca, 2014). Recent studies suggest that reputation systems may decrease or even eliminate discrimination, because they decrease the importance of more diffuse demographic information (Abrahao, Parigi, Gupta, & Cook, 2017; Cui, Li, & Zhang, 2016; Ert, Fleischer, & Magen, 2015; Mohammed, 2017; Tjaden, Schwemmer, & Khadjavi, 2017). However, these studies overlook an important assumption underlying this conclusion, namely that equally trustworthy individuals are equally likely to obtain reviews. As Frey and van de Rijt (2016) show, reputation tends to cascade: individuals who already have a positive review are more likely to be selected for new transactions and will as a consequence accumulate more reviews. Initial differences accumulate over time, resulting in inequality between equally trustworthy individuals. These initial differences may be random, but could also be the result of a difference in the probability to receive reviews between individuals of different ethnicity. Since reviews can generally only be written after a completed transaction, individuals who have characteristics that are more preferred by others are more likely to be selected for transactions and thus more likely to obtain reviews. If we consider that individuals with different ethnicity/nationality differ with respect to the probability to receive a first review, and that a first review is critical for acquiring more reviews, the probability to be selected for a transaction for renters of different nationality may diverge over time. In the current study I aim to answer the research question: Under what conditions do reputation systems exacerbate discrimination?

Hypotheses

On the one hand reputation systems may thus decrease the importance of demographic information, thereby decreasing overall discrimination. Hypothesis 1: The difference in trust received between individuals of different nationality is smaller when there is a reputation system than when there is no reputation system. On the other hand, initial differences between users of different ethnicity may translate into differences in reputation, which in turn affect future possibilities for interaction on the platform. Hypothesis 2: The difference in trust received between individuals of different nationality is larger when there is a reputation system than when there is no reputation system

Design Plan

Study type

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Is there any additional blinding in this study?

The research will not directly interact with the study subjects, since the experiment will be conducted online.

Study design

To answer the research question, an online experiment that is an extension of the experiment of Frey and Van De Rijt (2016). Subjects play one or two trust games of eight rounds in groups of eight. Half of them plays in the role of trustor, the other half in the role of trustee. The trustors take turns. When it is his or her turn, the trustor can choose to place trust in one of four trustees. When a trustee is selected, (s)he can choose to honor trust or abuse trust. The trustor's payoff is highest (50 points) when (s)he places trust and when the trustee honors trust, but lowest when the trustee abuses trust (0 points). The trustee's payoff is highest when (s)he abuses trust (80 points). Trustors that are not active in a round and trustees that are not selected get a payoff of

30 points. When the trustors places no trust, all players get a payoff of 30 points. Under the assumption of rationality, trustors anticipate that selfish trustees will always abuse trust, and will therefore be motivated to select the most trustworthy trustee. The experiment has a between-subjects design. Subjects participate either in the control condition or in the reputation condition. In the control condition, no information about past decisions of the other subjects is available. In the reputation condition full information about the decisions of the other subjects in the session is available. In the study of Frey and Van De Rijt (2016), trustors had no prior information about the trustees, so the inequality that they observed was based on initial random choices of trustors in early rounds of the experiment. In the current study I extend their design by providing some basic demographic information about the trustees to evaluate to what extent reputation systems may reinforce differences in outcomes between individuals with a different nationality. In all groups half of the trustees has the American nationality, and half has the Indian nationality. The distribution of the nationality of the trustors is counterbalanced across the groups: In half of the games American subjects are in the majority, with three of four trustors from the United States, while the fourth trustor is Indian. In the other half of the games Indian subjects are in the majority, with three trustors from Indian, while the fourth trustor is from the United States.

Randomization

The experiment consists of two conditions. The experiment will be ran online. The setup of the experiment requires a specific composition of the groups in terms of the nationality of the subjects. To reduce waiting times for the subjects and to avoid that a large number of subjects cannot be matched, only one condition will be ran at the time. Conditions are randomly assigned to times.

Sampling Plan

Registration prior to creation of data

Data collection procedures

The subjects participating in the experiment will be recruited through Amazon Mechanical Turk. All subjects that complete the experiment receive $2.00 plus a bonus that varies between $1.80 and $6.40 per game. The size of the bonus depends on the decisions made by the subject and its group members. Only MTurk workers from the US and India can participate in the study. MTurk workers must be age 18 or above. Only workers who have completed 1000 HITs or more and with an HIT acceptance rate of 90% or higher can participate in the experiment.

Subjects can only participate once. The experiment will be conducted in the fall of 2019. Using an experiment allows us directly compare a situation with and a situation without a reputation system and to draw causal conclusions about the effect of reputation on demographic inequality based on multiple sequences of reputation accumulation. Compared to a student sample, there is more demographic variation among workers on MTurk. We make use of the fact that the large majority of MTurk workers is from the US, followed by India (Difallah, Filatova, & Ipeirotis, 2018). Two pilot sessions have been conducted (one per condition) to test the software and to answer the following questions: -How many subjects drop out during the session. -How long does it take subjects to complete the experiment (necessary to determine a fair payment).

Sample size

Per condition (control and reputation), 75 groups play the game. Each group consists of eight players and each subject plays two games, so in total I will recruit 300 subjects per condition (600 subjects in total).

Sample size rationale

No response

Stopping rule

No response

Variables

Manipulated variables

I will manipulate whether trustors in the experiment can see previous decisions of trustors and trustees, other than their own decisions and the decisions of the trustees they selected (control versus reputation condition). Furthermore, I vary the composition of the trustor group. In half of the groups there are three American trustors and one Indian trustor. In the other half of the groups there are three Indian trustors and one American trustor.

Measured variables

The main outcome variable is inequality in the average level of trust placed in American and Indian trustees. I will calculate per group the fraction of trust placed in ethnic majority trustees, which is the total number of times trust was placed in ethnic majority trustees divided by the

total number of times trust was placed in a group. Considering that in each group two of the trustees are American and the other two are Indian, on average 50% of the trust would be placed in American trustees when trustors would randomly select one of the trustees, without taking their nationality into account. The more the fraction of trust placed in American trustees deviates from 0.5, the higher the inequality, so inequality is calculated as the absolute difference between the fraction of trust placed in American trustees and 0.5 . Groups in which trust was never placed are excluded from the analyses. To further increase our understanding of the effect of reputation systems on inequality I will break down the analyses for trustors with a different nationality. Per group I calculate for American and Indian trustors the fraction of rounds in which they placed trust in American trustees, out of the total number of rounds in which American or Indian trustors placed trust in that group. By comparing choices of American and Indian trustors across conditions we get a better understanding of the preferences of trustors with different nationalities and we improve our understanding of reputation systems on the choices of the trustors. By calculating the scores on the group level I control for differences in overall trust between groups. Groups in which trust was never placed are excluded from the analyses. At the round level I will record the ethnicity of the trustor who is at play, as well as the nationalities and reputation scores of the trustees she has to choose from. I also record if a trustor placed trust, and if so, which trustee was selected and what the decision of that trustee was. Reputation is measured by the number of times a trustee has honored and abused trust in the past. I will also construct a variable that indicates whether the trustor and trustee share the same nationality. Another variable indicates whether the subject belongs to the ethnic majority in a group. A last variable indicates if a trustor has selected the trustee that has been selected on the most recent turn the trustor could observe, provided that trust had been honored (Frey et al., 2016). This variable is only constructed for rounds in which trustors could observe a prior trust placement and honoring decision.

Indices

No response

Analysis Plan

Statistical models

I will use an independent samples T-test to test both hypotheses. I compare: 1) if the level of inequality between majority and minority trustees differs between the control condition and the reputation condition, and 2) to test if the level of inequality between American and Indian trustees differs between the two conditions. By doing so I account for different types of discrimination: homophily (which is expected to result in a overall advantage for the majority) and a preference for one of the two nationalities (which results in an overall advantage for one of the nationalities, regardless of the majority in a group). I will also use an independent samples T-tests to compare the choices of American and Indian trustors across the conditions. This also allows me to test if trustors indeed place more trust in trustees who have the same nationality as they have (homophily).

Transformations

No response

Inference criteria

No response

Data exclusion

Attention checks are used in the instructions of the game. Only subjects who give the right answers to five comprehension questions that also included questions about the history table (which varies across the condtions) can participate in the experiment. At the end of the experiment, subjects are asked what they believe to be the goal of the experiment.

Missing data

For the analyses on the group level, only data from groups that complete the entire game will be included in the dataset. For the analyses on the individual level all datapoints will be included in the analyses.

Exploratory analysis

In addition, I will perform a number of exploratory analyses to learn more about the mechanisms underlying the effect of the reputation system on inequality. First, I use T-tests to compare the overall levels of trust and trustworthiness between the conditions. Second, because reputation tends to cascade, the first trustor who places trust in the reputation condition may have a large influence on the choices of future trustors. I test if the difference difference in

inequality (between the majority and the minority, and between American and Indian trustees) between the conditions is affected by the nationality of the first trustor that placed trust by using a T-test at the group level. Next, I use a conditional logit model to analyze how characteristics of the trustor at play and the trustees she can choose from affect the choice of the trustor. The choice for one of four trustees is the outcome variable. For each trustee, I include their nationality, as well as a variable that indicates if the trustee has the same nationality as the trustor that makes a choice in the specified round as independent variables. I also included the nationality of the trustor, as well as the round number and a variable that indicates which nationality is the majority in the group. I include random intercepts for trustors, because they make multiple decisions across the game(s) and these decisions are not independent. The model was run separately for the control condition and the reputation condition, and for the groups in which the American versus Indian trustors were the majority. If the outcomes do not depend on which nationality is the majority, all of these groups will be analyzed together. In that case a variable indicating the majority condition will be added to the models. To test if trustors indeed place more trust in trustees who have a better reputation, I include the reputation of the trustee as an independent variable in the models for the reputation condition. I will further extend the models of the reputation conditions to also test if the difference in the level of trust placed between trustees with a different nationality is smaller for trustees with a more positive reputation. In a first model I include the interaction between the reputation and the nationality of the trustee. IA second model will include the interaction between the reputation of the trustee and the variable that indicates if the trustee and the trustor have the same nationality. Lastly, I will analyze the prevalence of reputation cascade by estimating the probability that a cascade continued using a logistic regression with random intercepts for trustors and the variable indicating if the trustor selected the last trustee that last honored trust she was aware of.

Other

Abrahao, B., Parigi, P., Gupta, A., & Cook, K. S. (2017). Reputation offsets trust judgments based on social biases among Airbnb users. Proceedings of the National Academy of Sciences of the United States of America, 114(37), 9848–9853.

Cui, R., Li, J., & Zhang, D. J. (2016). Discrimination with Incomplete Information in the Sharing Economy: Evidence from Field Experiments on Airbnb.

Edelman, B., & Luca, M. (2014). Digital Discrimination: The Case of Airbnb.com. Harvard Business School.

Ert, E., Fleischer, A., & Magen, N. (2015). Trust and Reputation in the Sharing Economy: The Role of Personal Photos on Airbnb. Available SSRN 2624181.

Frey, V., & Van De Rijt, A. (2016). Arbitrary Inequality in Reputation Systems. Scientific Reports, 6(1), 38304.

Mohammed, A. R. S. (2017). Designing for Racial Impartiality: The Impact of Relocating Host Photos on the Airbnb Website. Working Paper.

Tjaden, J. D., Schwemmer, C., & Khadjavi, M. (2018). Ride with Me—Ethnic Discrimination, Social Markets, and the Sharing Economy. European Sociological Review, 34(4), 418–432.

Instructions participants

Consent form

The goal of this experiment is to better understand how people make decisions in groups. This experiment is conducted by researchers from Utrecht University, the Netherlands.

If you agree to be in this study, you will play two games in which you will be paired with seven other individuals. The decisions that you and the other participants make in the games will determine how much you earn. If you complete the experiment, you will receive a payment of $2.00 plus a bonus that varies between $3.60 and $12.80. You will receive your payment as a bonus payment on to the HIT on MTurk that you used to sign up for the study. It may happen that you finish reading the instructions, but after that cannot be assigned to a group to play the game. In that case you will receive a payment of $2.

There are no known risks associated with your participation in this research beyond those of everyday life. Taking part in this study is voluntary. Confidentiality of your research records will be strictly maintained. We collect your Amazon Mechanical Turk ID to calculate your bonus payment. After the study has been completed, we will separate your MTurkID from the experimental data so answers cannot be traced back to you. The data from the study will be kept at least 10 years.

If there is anything about the study or taking part in it that is unclear or that you do not understand, you may contact the researcher (Judith Kas, j.kas@uu.nl).

Instructions for participants

Please carefully read all of the information that follows.

In this experiment, you will earn points by making decisions in "games". How much you earn depends on your decisions, the decisions of others and on chance. The experiment will last about 45 minutes. At the end of the experiment, you will be paid $2.00, plus a bonus of $1.00 for every 100 points you earned in the games. Please note that this is an interactive experiment. That means that you may have to wait while other participants are making decisions.

We will now explain the rules of the game. After you have read the instructions you will be asked to answer a few questions that help you evaluate your understanding of the game.

The rules of the game

This game is played in groups of 8 participants, four participants in the role of A (A1, A2, A3 and A4) and four participants in the role of B (B1, B2, B3 and B4). Every participant is randomly assigned to a role and number. Throughout the game, all participants keep their role and number.

We understand that sometimes studies claim that participants will actually interact in real time, but in fact use simulated other people ("bots"). As a policy, we do not do this. In the game you will really be interacting with 7 other people.

The game proceeds in rounds. There are 8 rounds. In each round only one of the As is active and interacts with the Bs. In round 1, it is A1's turn to interact with the Bs; in round 2, it is A2's turn, and so on such that in round 5, it is again A1's turn. How many points the active A and the Bs get depends on their choices.

The active A (simply called "A" in Figure S1) chooses either "RIGHT" or "DOWN."

- If A chooses RIGHT, A gets 30 points and all four Bs get 30 points as well.
- If A chooses DOWN, A must select one of the four Bs.
- If A chooses DOWN and selects one of the Bs, the selected B chooses "RIGHT" or "DOWN."

If the selected B chooses DOWN, A and the selected B get 50 points each. If the selected B chooses RIGHT, A gets nothing (0 points) and the selected B gets 80 points. In either case, the

other Bs – the ones that were not selected by the A at play – get 30 points each. The As who are not active in a round get 30 points in that round, just as the Bs that were not selected.

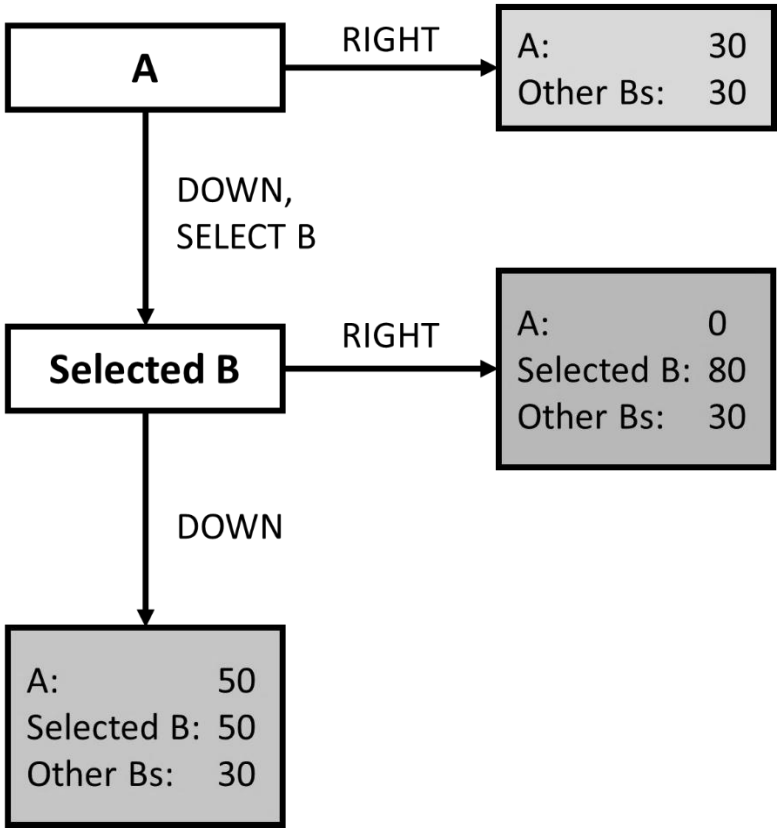

Figure S1: The game

The computer interface

We will now explain what your screen will look like during the game. What you will see on the right-hand side of your screen will be self-explanatory. On the left-hand side of your screen you see an example of a 'history window', which will always be visible during the game. Each of the four columns represents one B-participant and each row represents a round. At the top of the table you see the nationality and gender of each B-participant. The current round - round 5 in the example - is indicated by the bright green row color. In parentheses it is displayed which A is at play in which round. On the next pages we explain the meaning of the colors. The background color of a cell in the table shows what choices were made.

Control condition only:

Each A gets informed about the results of his/her own interactions but not about the results of the interactions of any other A. Bs get informed about all choices.

Reputation condition only:

All As and Bs get informed about all choices of all participants.

In Figure S2 (control condition) / S3 (reputation condition) you see an example screen for an A (A1 in this case). The color of the background shows what choices were made. At the bottom of the table you can find a legend with an explanation of the colors.

Control condition only:

A1 sees question marks (“?”) in the rounds in which one of the other As is at play. The question marks of past rounds are not highlighted in color because A1 never receives information about the choices in these rounds.

On the left you see an example screen for a B. Again, the colors indicate what decisions were made in previous rounds. In the example it's player A1's turn.

Control condition only:

The plus signs in past rounds indicate to a B which past rounds A has information about. Participant A1 only sees a plus sign ("+") in the rows in which he or she is at play (round 1 and 5). All participants in role B see plus signs ("+") and question marks ("?") in the same rounds as the A who is at play (A1 in the example). The plus signs in future rounds show when it's A's turn again. Note, the plus signs and question marks are placed in different rounds if it is a different A’s turn.

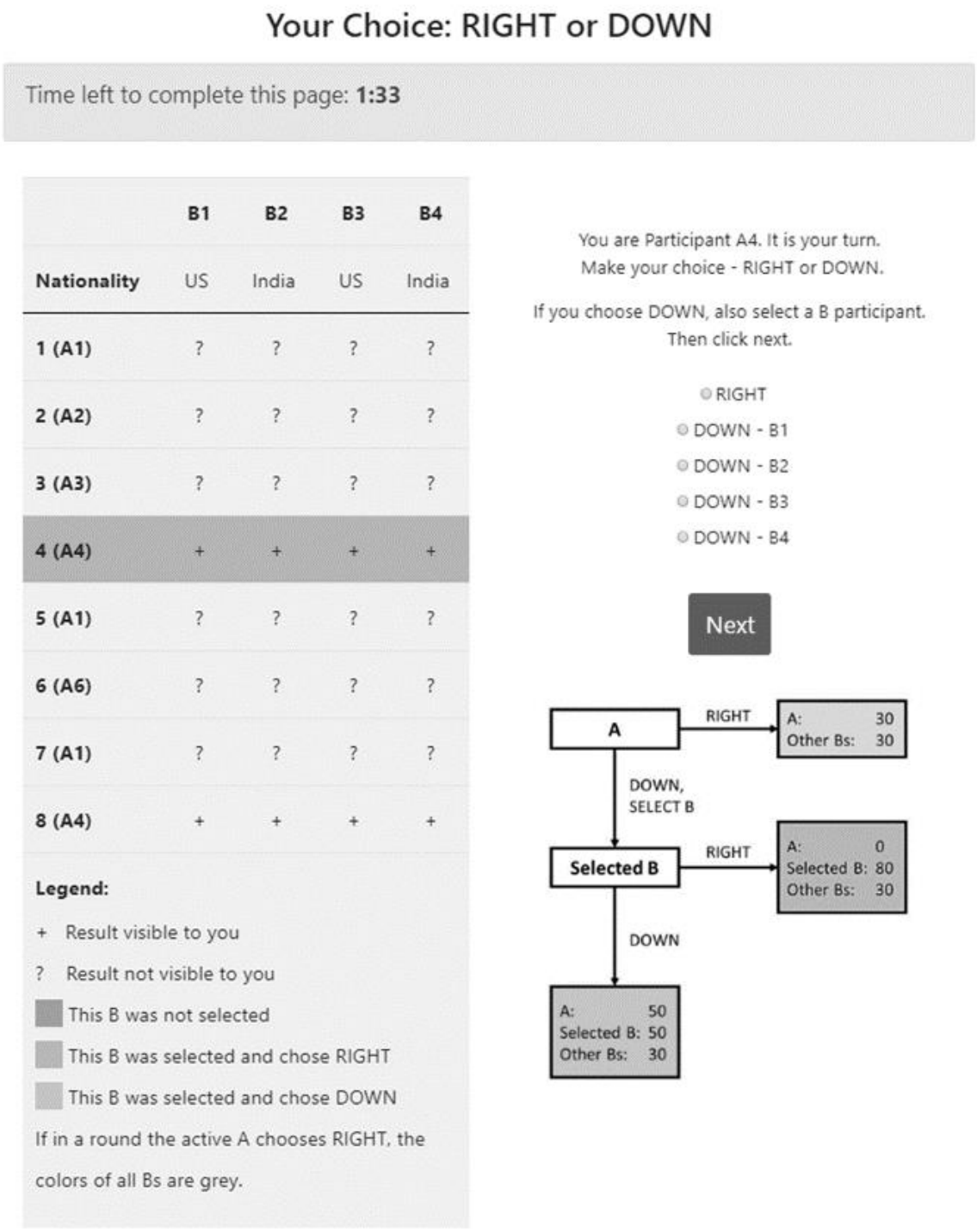

Figure S2: Decision screen for an A-player in the control condition.

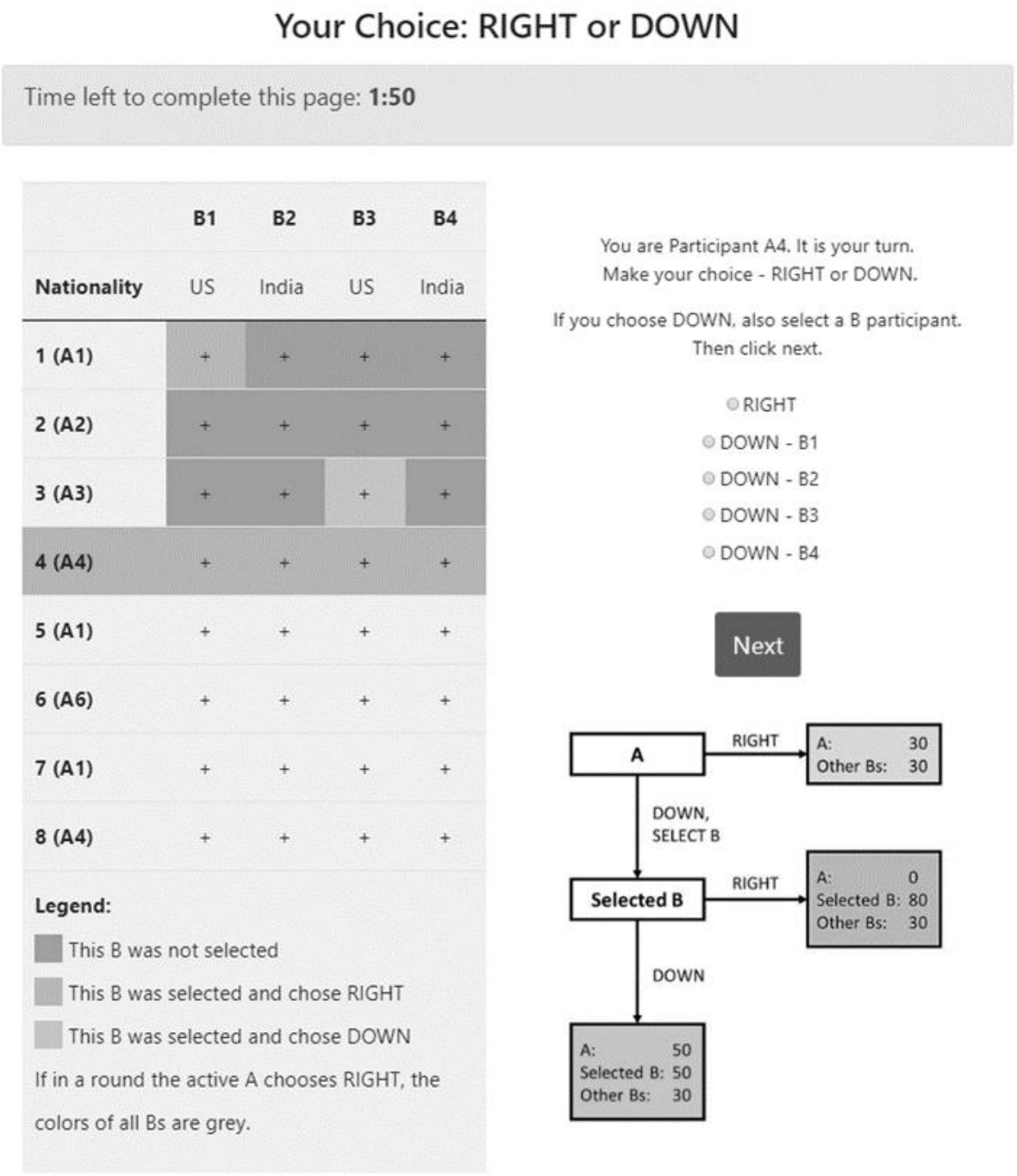

Figure S3: Decision screen for an A-player in the reputation condition**.**

The duration of the game

You will participate in 2 games, each lasting 8 rounds, one game played after the other. For each game you get randomly assigned to a new group of 8 participants and to your role. It is

possible that you are matched with the same other participant in more than one game. However, should this happen, neither you nor the other participant will be able to notice this.

Interactive game

We would like to remind you that this is an interactive game: you are playing the game with 7 real other people. This means that you may have to wait sometimes while other people are making decisions. It also means that other people have to wait while you are making a decision. It is therefore important that you complete this study without interruptions. Please make your decision within the time limit shown on your screen. If you do not make your decision between RIGHT or DOWN within the time limit, the game will end and you will receive no payment and no bonus. The other participants will in that case receive the points that they earned so far, plus the points they would have earned in the remaining rounds (based on their average number of points earned per round). If this happens in the first round of the game (before you had the chance to earn any points) you will receive 30 points per round.

During the study, please do not close this window or leave the study's web pages in any other way. If you do close your browser or leave the task, you will not be able to re-enter and we will not be able to pay you!

To help you make a decision in time, we can send you a desktop notification when it is your turn to make a decision. If you wish to receive these notifications, please click the button below. After that click "Next".