Reputation offsets trust judgments based on social biases among Airbnb users

Bruno Abrahao[a,1], Paolo Parigi[b], Alok Gupta[c], and Karen S. Cook[a,1]

[a]Department of Sociology, Stanford University, Stanford, CA 94305; [b]Institute for Research in the Social Sciences, Stanford University, Stanford, CA 94305; and [c]Data Science Division, Airbnb, San Francisco, CA 94103

Contributed by Karen S. Cook, July 20, 2017 (sent for review March 15, 2016; reviewed by David Lazer and Chris Snijders)

To provide social exchange on a global level, sharing-economy companies leverage interpersonal trust between their members on a scale unimaginable even a few years ago. A challenge to this mission is the presence of social biases among a large heterogeneous and independent population of users, a factor that hinders the growth of these services. We investigate whether and to what extent a sharing-economy platform can design artificially engineered features, such as reputation systems, to override people's natural tendency to base judgments of trustworthiness on social biases. We focus on the common tendency to trust others who are similar (i.e., homophily) as a source of bias. We test this argument through an online experiment with 8,906 users of Airbnb, a leading hospitality company in the sharing economy. The experiment is based on an interpersonal investment game, in which we vary the characteristics of recipients to study trust through the interplay between homophily and reputation. Our findings show that reputation systems can significantly increase the trust between dissimilar users and that risk aversion has an inverse relationship with trust given high reputation. We also present evidence that our experimental findings are confirmed by analyses of 1 million actual hospitality interactions among users of Airbnb.

online trust | reputation systems | sharing economy | social biases | risk

A new wave of companies, emerging under the banner of the sharing economy (1), is profoundly altering the way we interact and exchange with one another. These Internet-based services are driving a major change in our cultural and technological landscapes and have achieved astounding success, enabling users to share their own personal resources, such as their vehicles, real estate properties, time, or skills. A growing number of individuals trust the sharing economy with a variety of services to satisfy their needs, to generate income, or, more simply, to meet new people. Examples of sharing-economy transactions include hiring a "tasker" from Task Rabbit to run errands, sharing a "couch" with a perfect stranger through CouchSurfing, hiring a "driver" on Uber, or staying in someone's home while traveling using Airbnb.

Users in the sharing economy seek to connect with others engaged in activities on the same platform. Compared with exchanges via traditional e-commerce companies, where transactions are relatively anonymous, the sharing economy exposes us to the more personal character of such interactions. This inevitably prompts attention to the users' sociodemographic characteristics as factors that drive selection.

As a consequence, social biases figure as major hurdles to the growth of sharing-economy services, as they influence users' perceptions of trust and risk. To enable trust between strangers so that everyone can exchange with anyone, beyond cultural and social boundaries, these companies face daunting obstacles in their attempts to minimize these biases.

In this study, we investigate whether and to what extent a sharing-economy platform can design technological features to counteract natural behavioral tendencies that may lead to social biases. This question is of central importance in the social sciences more broadly, but also in the engineering of platforms that aim to enable trust.

Social biases are a result of a number of mechanisms that are difficult to measure. In this work, we make social biases amenable to investigation by focusing on a form of social bias that naturally maps into a quantifiable interpretation and that we expect to be at work in these environments. At the same time, this source of bias is well understood in the social sciences so that we can rely on previous literature, instead of opening up a new dimension of complexity. To this end, we focus on homophily (2–6), the higher likelihood that people trust others who are similar to themselves.

McPherson (4) proposed a theory of how homophily structures modern societies using a construct of social space defined in Blau's theory of preferences (6). Each individual occupies a position in the social space whose coordinates are a function of his or her sociodemographic characteristics. The more features two individuals share in common, the more likely they are to form relationships based on mutual trust.

To operationalize homophily in a structured way, we use Blau's construct of social space to induce and measure the effect of homophily in an experimental setting whose volunteers are active members of the sharing economy. (At the time of writing, the online experiment is accepting participants for demonstration purposes at stanfordexchange.org.)

Building on this baseline, the heart of our experiment is the measurement of the extent to which another source of information that can be artificially engineered could potentially alter the

Significance

We investigate the extent to which artificial features engineered by sharing-economy platforms, such as reputation systems, can be used to override people's tendency to base judgments of trustworthiness on social biases, such as to trust others who are similar (i.e., homophily). To this end, we engaged 8,906 users of Airbnb as volunteers in an online experiment. We demonstrate that homophily based on several demographic characteristics is a relatively weak driver of trust. In fact, having high reputation is enough to counteract homophily. Using Airbnb data, we present evidence that the effects we found experimentally are at work in the actual platform. Lastly, we found an inverse relationship between risk aversion and trust in those with positive reputations.

Author contributions: B.A., P.P., and K.S.C. designed research; B.A., P.P., and A.G. performed research; B.A. and P.P. contributed new reagents/analytic tools; B.A., P.P., A.G., and K.S.C. analyzed data; and B.A., P.P., and K.S.C. wrote the paper.

Reviewers: D.L., Northeastern University; and C.S., Eindhoven University.

Conflict of interest statement: A.G. is a data scientist at Airbnb who performed the experiments that rely on the company's private data. Paolo Parigi began working at Uber after the research design, experiment execution, data analysis, and writing of the study were completed.

Freely available online through the PNAS open access option.

Data deposition: The data necessary to replicate our experimental results are available through the Stanford Exchange Project at stanfordexchange.org/project.

[1]To whom correspondence may be addressed. Email: abrahao@cs.stanford.edu or kcook@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1604234114/-/DCSupplemental.

perception of trust structured by homophily and counteract this natural tendency. To this end, we focus on the reputation system (7, 8), which platforms use to allow users to review and to "rate" the behavior of other members (7).

The premise of reputation systems is that the aggregate rating associated with a person is an indicator of the quality and the risk entailed in potential transactions with that individual. We hypothesize that it is because of these reputation systems that the lack of direct experience in interacting with unknown and distant alters does not generate paralyzing uncertainty.

Nonetheless, previous research presents only weak evidence that reputation systems serve as safeguards of opportunistic behavior, which would result in increased trust (9). Moreover, there has been limited quantification of the extent to which reputation systems have the capacity to increase trust between those with different degrees of dissimilarity in social space. This study directly measures interpersonal trust structured by the interplay between homophily and reputation.

Researchers have extensively studied reputation systems in online platforms in auction markets (10), crowdsourcing (10, 11), and the sharing economy (12–14). The latter case is of particular interest, as the user population, as well as the pool of services they offer, is too large and diverse to be standardized, while users cannot rely directly on preexisting institutional arrangements to inform their decisions.

Measuring trust as a function of the influence of social distance or reputation directly on a sharing-economy platform represents a major research challenge that is not amenable to direct manipulation by researchers. The platform exposes users to features of the alternatives that are confounded with trust. Features, such as color preference, attractiveness, etc., are difficult to isolate, categorize, or quantify. These factors are highly heterogeneous, and users infer them subjectively and indirectly through photos or other signals, as opposed to through structured data that the platform displays. Moreover, when the platform presents users with alternatives, we observe the outcome of the user's thought process by their selections, but it is difficult to capture and quantify their preferences between every pair among the available alternatives without making the selection process unnecessarily complex.

Due to such challenges, we designed a large-scale online laboratory in collaboration with Airbnb, one of the world's most successful sharing-economy companies, with >2 million hospitality listings in >190 countries. We engaged 8,906 Airbnb users as volunteers to participate in an experiment external to Airbnb's platform, with the aim of collecting behavioral data on decisions that involve trust, while isolating other confounding factors (15).

Traditionally, buyers and sellers may develop trust and engage in multiple exchanges over time. However, relationships in the sharing economy usually result in one-time transactions. To mimic this scenario while allowing for comparability with previous literature, we drew from research on behavioral economics and social psychology to use standard methods to measure interpersonal trust. To this end, we designed a variation of the widely studied investment game (16, 17).

In the experiment, participants played the role of Investor and started with a number of credits. We showed them five profiles, which we presented as belonging to other randomly selected Airbnb users who had entered their information in a previous round. These other users were playing the role of receivers with incentives to accumulate credits. We displayed the receivers' demographics and reputation features, resembling the way the platform presents users with potential partners via search results. We then gave incentives for participants to seek favorable investment outcomes in a single-shot interaction with the receivers by offering potential rewards.

The credits participants invested in a profile were multiplied by three and given to that receiver. In the spirit of the prisoner's dilemma game (18), the receiver could cooperate with the participant and return a good portion of the credits invested in them (to help the participant increase her credits) or defect and keep most or all of these credits (in which case the participant loses credits). As participants put themselves in a vulnerable position through investments, the amount of investment in each receiver served as an indirect proxy for how much trust they placed in the receivers to return credits.

While we presented the profiles to participants as other Airbnb users also participating in the experiment, we actually generated these profiles in advance. Our goal was to understand how participants assign trust and the relative importance of each facet, homophily or reputation, as we varied demographic and reputation features experimentally.

We present evidence, first, that homophily is at play to a significant extent when participants make trust decisions. Second, we show that the reputation system makes possible the construction of expectations that counterbalance the tendency toward homophily. This results in the extension of trust to dissimilar users in social space. Third, we show a strong inverse relationship between risk aversion and trust in the case of high reputations. Lastly, we used insights gained through the experiment to guide our analysis of Airbnb data containing 1 million real-world interactions. We show evidence that the diversity of users who select others with whom to have a hospitality interaction increases as the reputation of the partners gets higher.

Experimental Design

We sent invitations to 100,000 Airbnb users who identified as US residents, of which 8,906 responded and registered to participate (6,714 completed their entire participation). Table S1 presents the univariate distributions of the demographics over the participants, and Tables S2–S4 analyze self-selection bias. Each participant built a short profile by providing us with four pieces of information reflecting their demographics—namely, age, gender, marital status, and home state. We chose these features for their simplicity, allowing us to conveniently operationalize Blau's construct experimentally while reducing participation attrition.

We required that the experiment reflect, as much as possible, the way users make decisions on the platform, except for factors that are external to trust. In the platform, users are pressed to make the best possible decisions, as it is imperative to eliminate risks associated with critical factors, such as their safety, while maximizing satisfaction and minimizing cost. Thus, a major challenge in the design of our online experiment was to engage users so that they would attempt to make the best use of their judgment when making decisions involving trust. To capture attention and provide incentives for the exertion of good judgment, we offered 100 prizes, each for 100 US dollars (USD). The chances of winning were proportional to the number of credits accumulated in the investment game.

We generated each of the potential receivers according to prescribed rules. We placed the receivers' profiles at social distance d from the participant, defined in the context of Blau's social space as the number of features on which two individuals differ (6). This is equivalent to the mathematical definition of Hamming distance. Accordingly, distance $d = 0$ meant that the receiver matched all of the demographic attributes of the participant (e.g., the same age group, the same gender, the same marital status, and the same US state). In turn, $d = 1$ meant that one randomly selected feature's category differed from that of the participant. The profile at $d = 2$ is strictly farther from the participant by having one additional randomly selected feature changed to a different category. Lastly, $d = 4$ meant that the profile had all of the demographic features in a different category from those of the participant. The receivers were placed at distances $d = 0, 1, 2$, and two of them were placed at $d = 4$. We showed the five profiles simultaneously on the participant's screen in random order.

In addition to demographics, the generated profiles included two reputation features—namely, the average number of star ratings and the number of reviews on Airbnb. The star rating

SOCIAL SCIENCES

is a postinteraction subjective evaluation of an alter. It consists of the assignment of zero to five stars, where the number of stars is proportional to the degree of positiveness. The ratings a member receives are averaged over all of their raters, rounded to the half unit, and presented in the member's profile on the platform. Similarly, an interaction grants the two parties the opportunity to mutually provide free-form written reviews. Due to the difficulty of manipulating textual contents of reviews experimentally, we restricted our attention to the number of reviews a user received.

We manipulated these two dimensions in a structured way to study their effects on trust. Among the five profiles participants saw on the screen, four had reputation features with similar values, chosen independently at random for each participant's session, which we refer to as the baseline reputation. These were the profiles at social distances $d = 0, 1, 2$ and one of the profiles at $d = 4$. The other generated profile at distance $d = 4$ had one of the reputation features randomly selected to be switched to either a better or a worse value than baseline (see *Game Design Details* for how we manipulated the numerical values of reputation). For convenience, we refer to the profile that has a different reputation feature than the baseline as being at distance $d = 5$.

We randomly assigned users to two possible worlds. In world 1, the profile at $d = 5$ not only had the largest distance from the participant, but also a weaker reputation than all other profiles (the baseline reputation). In this case, reputation did not compete with the tendency toward homophily. In world 2, the profile at $d = 5$ had a better reputation than the baseline reputation. This induced a tension between placing trust in the most distant profile with a better reputation or in the other profiles closer to the participant in social space. Fig. S1 shows a partial view of the screen users see in the experiment, and Fig. S2 shows a diagram that exemplifies the structure of a user's session.

We gave participants a single "wallet" with 100 credits, which they could keep or invest in receivers in whatever way they chose. Therefore, participants could gain or lose credits through their investments. Because this was a one-time game, it was easy to show that the Nash equilibrium was not to invest any amount, since the dominant strategy for receivers was not to return any amount. (Nevertheless, we observed such rational behavior only in rare instances.)

It is argued that risk is a component of trust in general, and some definitions of trust include risk (8). Even though previous research has attempted to relate trust and risk, the empirical evidence of the connection between risk attitudes and trust has been weak (17). Moreover, research that has addressed this question has been limited to laboratory experiments or small datasets.

Given the opportunity to study this question using a large population, we introduced a risk-assessment question before the investment game. We worded the question as: "A lottery ticket costs 100 (USD) and people win with 50% chance. How much should the prize be for you to choose to buy a ticket?" Players could enter any numerical value, which corresponded to the minimum reward that would make the participant take the risk of buying a ticket. The prize value 200 (USD) had the expected value of net gain equal to zero (after paying off the ticket) and corresponded to the minimum rational value. Thus, values >200 (USD) measured risk aversion proportional to their magnitude. In *Risk Assessment Question*, we summarize the distribution of answers (Table S1) and argue that our measure captures risk behaviors in accordance with previous research (Table S5) (19, 20).

Multilevel–Multivariate Analysis

We had five measurements (investments) on each observational unit (participant). As a result, the five investments were correlated, which we accounted for by nesting investments within subjects in a multilevel model. We fitted the model using a multivariate regression with 10 independent variables, one for each investment in the combination (d, w) of profile distance $d : \{0, 1, 2, 4, 5\}$ and world $w : \{1, 2\}$. The investments a participant made had different sources of mutual correlation. For instance, the sum of the investments had to be at most 100 credits. We accounted for these by computing the model fit with an unconstrained covariance structure that learned from the data the correlations and independent variances across measurements (21).

As a first-order approximation, we fitted the empty model (i.e., without explanatory variables) with 10 intercepts. The five intercepts for each world corresponded to the average distribution of investments among the five profiles across all participants (complete pooling). Fig. 1 shows a plot of the mean estimates, together with the mean number of credits saved, for worlds 1 and 2. Table S6, model 1 shows the numerical estimates from the model fit.

We were mainly interested in the additive effect of the number of different coordinates between two individuals' feature vectors, or their Hamming distance. However, any real-world sociodemographic feature inevitably produces heterogeneous effects on trust (e.g., gender may affect investments more than marital

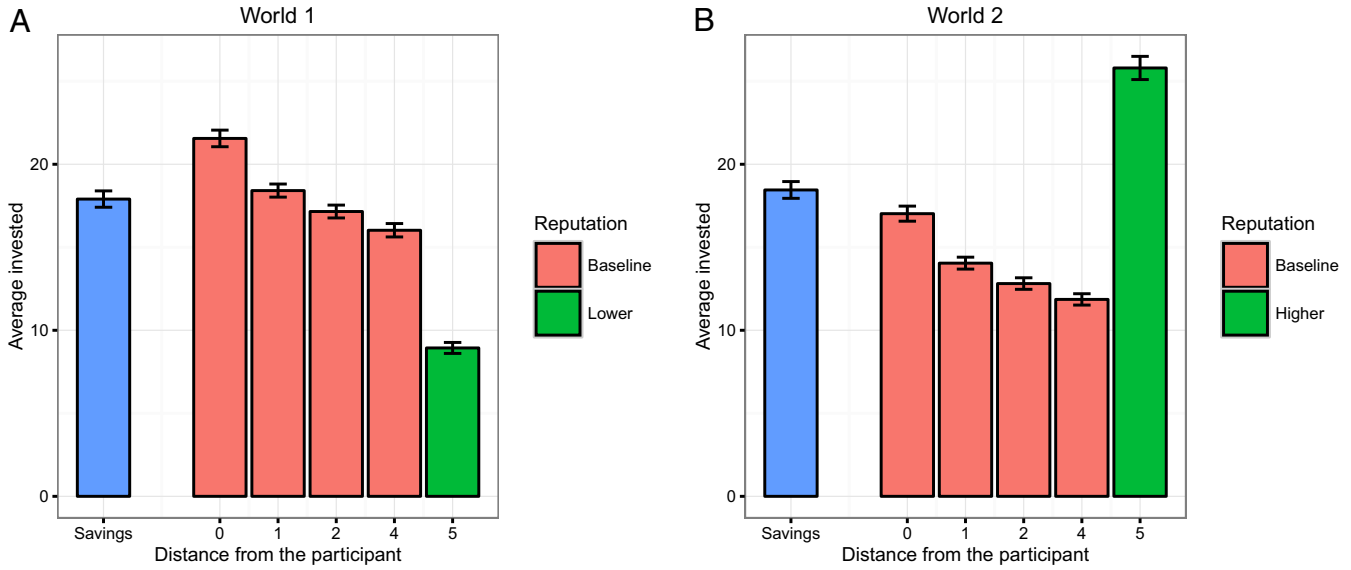

Fig. 1. Empty model estimates of average investment in profile at distance *d* and average savings. (*A*) In world 1, the second profile at distance $d = 4$ (here identified as $d = 5$) has a worse reputation than baseline. (*B*) In world 2, the profile at distance $d = 5$ has a better reputation than the baseline.

status does), and Hamming distance by itself may not explain all of the variance in the investments. Thus, to take these effects into account, we extended the empty model by including explanatory variables.

Table S7 shows a list of the inputs we used to form these covariates. They can be categorized into three sets structured in a multilevel model as: (*i*) level 1 variables corresponding to the profile's characteristics, annotated with "P"; (*ii*) level 2 variables corresponding to the participant's (or subject's) characteristics, annotated with "S"; and (*iii*) the cross-level interactions between the level 1 and 2 variables.

We expected that the demographic features we used to increase social distance could have resulted in effects rooted in preferences, which are not necessarily biases, such as preferences for "female," "married," or "older" as indicators of perceived trustworthiness. Thus, the cross-level interactions aimed to control for these effects.

The multivariate model estimated the effects associated with the covariates specifically for each dependent variable (d, w). This allowed us to show the effects of each explanatory variable on trust in each of the five profiles (in each of the worlds) separately. In the case of cross-level interactions, this was not always possible due to the symmetries in the participant's session. For example, all profiles at $d = 0$ exactly matched the participant's gender, marital status, and region. In these cases, we estimated joint effects on the investments in the five profiles simultaneously (by world). In Table S8 we present an alternative analysis of the data based on McFadden's choice model (22).

Note that the intercepts of the full model are consistent with those in the empty model, up to estimation errors (Table S6, models 1, 2, and 3). Thus, we used this estimate of the distribution of mean investment over the profiles as a starting point and studied how the explanatory variables changed these values.

Fig. 2 presents the effects of these covariates (integer-valued variables were centered and standardized to make all coefficients comparable). Negative values reduce average investments, whereas positive values increase them. Our main goal was to show that the heterogeneity of the features did not significantly alter the main effects we observed on average investments as a function of d in the empty multivariate model.

Results

Fig. 1*A* shows that homophily dominated investment decisions. That is, the farther away the profile was on the demographic dimensions from the participants, the lower the investment they received, on average. Furthermore, the profile at $d = 5$ with worse reputation received less investment on average than the equivalent alternative with respect to social distance (i.e., the profile at $d = 4$). Quite strikingly, Fig. 1*B* shows that reputation builds trust beyond homophily. The average investment in the profile at $d = 5$, possessing the best reputation, was significantly higher than the average invested in all of the closest profiles. Note that despite the strong influence of the reputation system in world 2, the magnitude of the investments in the profiles with baseline reputation was still driven by homophily.

The explanatory variables exhibited variance beyond that explained by social distance, which implies that there are differences in investment behavior by demographic group and their interactions. However, as we argue next, the changes in the average investments (model intercepts) that these effects produced in the multivariate model were not strong enough to significantly alter the conclusions regarding homophily and reputation that we previously derived from the empty model.

Homophily Is at Work. The covariate "profile distance" was by far the dominant one with respect to variance explained (F value 5668.8, $P < 0.001$). This was followed by the number of reviews with a much smaller F value (26.1, $P < 0.001$).

The dashed lines in Fig. 2 have the values ± 1.37 and correspond, in the most conservative way, to the smallest difference in average investment between two profiles with baseline reputation, minus two standard errors. That is, a coefficient that

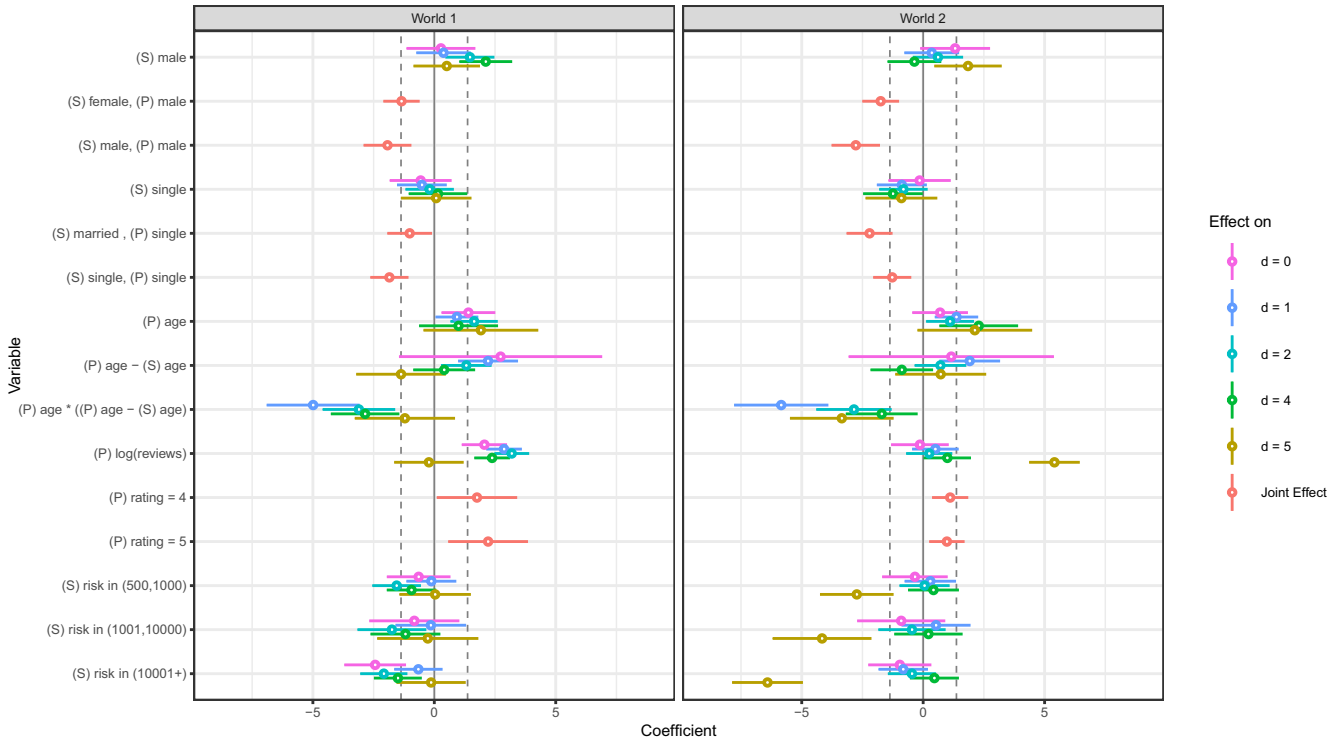

Fig. 2. The effects of the covariates associated with the participant (S) and profiles (P) in the multivariate multilevel model. The dashed lines have the values ± 1.37, which correspond to the smallest average investment difference between two profiles with baseline reputation, minus two standard errors.

exceeds these boundaries potentially produces an effect that could alter the conclusions we derived from the empty model. A first glance at Fig. 2 reveals that most of the coefficients are contained within these boundaries.

Fig. 2 shows that participant's gender "(S) male" in both worlds had small positive effects on all profiles. Marital status "(S) single" had effects that were not significantly different from zero. For age, the older the profiles ["(P) age"], the more credits they received. One SD (14 y) above the mean (39.7) had positive effects for all of the profiles with coefficients ranging from 0.93 (0.44) to 2.29 (0.81). The effects associated with region had small values that varied together across different profiles (omitted in Fig. 2 for clarity).

As these effects changed the investments roughly uniformly across the profiles, these effects did not cause significant changes in the differences between the investment means.

We note that the preceding effects did not change homophily trends due to the inclusion of interaction effects between participants' characteristics and those of the profile in the model. In Fig. 2, these variables are labeled with both S and P, such as "(S) female, (P) male" for gender. Recall that we included these interactions to capture preferences that are not necessarily biases. Indeed, in both worlds, male profiles received on average up to 3.38 (0.50) fewer credits than females, while not married profiles received on average up to 2.50 (0.40) fewer credits than married profiles. Age difference exhibited a nonlinear relationship. As the profiles got older than the participant, homophily came into play, and the positive effect of the profile's age decreased significantly, as indicated by the interaction of profile's age with the age difference between the profile and the subject.

Without controlling for these preferences (no interaction effects), the model exhibited effects associated with demographic features that canceled out the homophily effects produced by social distance in the case of males or singles. For illustration, in Table S6, we included the effects of gender and marital status for the models that included the interactions (model 3) and that with interactions removed (model 2).

As the group effects of investment behavior were not large enough to alter the trends produced by profile distance, we show evidence that homophily figures as a major driving force, structuring decisions of whom to trust with investments.

Trust via Reputation. We first focus on the effects of reviews in Fig. 2. In world 1, an increase in the log-transformed number of reviews, "(P) reviews (log)" resulted in a statistically indistinguishable increase in mean investment in profiles with baseline reputation, between 2.03 (0.47) and 3.20 (0.36) credits. Although the profile at $d = 5$ in this world always had fewer reviews than baseline, the variation in its number of reviews did not affect the average investment it received. In contrast, in world 2, an increase in the number of reviews increased the mean investment in the profile at $d = 5$, with the best reputation by 5.42 (0.52) credits. Symmetric to world 1, the change in the number of reviews of the baseline reputation did not affect the average investment in these profiles.

Comparing the effects of number of reviews between the two worlds, we see that high reputation resulted in larger investment increases in cases in which holding the best reputation was an exception among the alternatives (world 2). Surprisingly, these exceptions were the profiles that were the farthest away from the participants in the social space.

The coefficient of the joint effect estimated for variable "(P) rating = 4" represented an increase of 1.74 (0.83) and 1.13 (0.37) credits for worlds 1 and 2, respectively, in mean investment (reference "no rating available"). The corresponding increases for profiles with five-star ratings, "(P) rating = 5," was 2.21 (0.82) and 0.99 (0.36) for worlds 1 and 2, respectively. This shows that varying between 4 and 5 stars did not cause a significant difference in average investment, as participants may have considered them equally high.

In the full model of world 1, the difference between the mean investments comparing the profile with baseline reputation receiving the smallest average investment ($d = 4$) and the profile at $d = 5$ with the lowest reputation was 5.73 (1.89). In world 2, the difference between the profile with baseline reputation receiving the largest mean investment ($d = 0$) and the profile at $d = 5$ with the best reputation was 15.46 (1.68). In Fig. 2, we see that none of the effects were large enough to cancel out the shifts produced by reputation and alter our conclusion with respect to trust increases (world 2) or reductions (world 1). This shows evidence that the reputation system is a strong signal that shifts trust beyond homophily, thereby overriding the effects of assessments of social distance.

Risk. Fig. 2 shows the effects of answers to the risk-assessment question on the investments in each profile. We grouped responses by ranges, where the higher the range, the more risk-averse we classified a participant to be. These are the covariates with prefix "(S) risk in range," where the reference level is the range $[200, 400]$, the low end of rational values.

In world 1, we saw little or no effect associated with risk attitudes on the investments in any of the profiles, except for small negative effects on the investments in those with baseline reputation and weak similarity with the participant ($d = 2$ and $d = 4$). The effects ranged from a reduction of -1.74 (0.61) to -2.85 (0.48) in average investments, with slightly stronger effects proportional to the level of risk aversion.

The most striking results were related to world 2. In this case, risk attitudes did not correlate with the average investments in any of the profiles, except in the profile at $d = 5$ (with significance $P < 0.001$). These effects were among the strongest we found (Fig. 2, *Right*, the bottom three items). The decrease in mean investment ranged between 3.91 (0.75) and 8.16 (0.71) and was inversely proportional to the degree of risk aversion. This shows

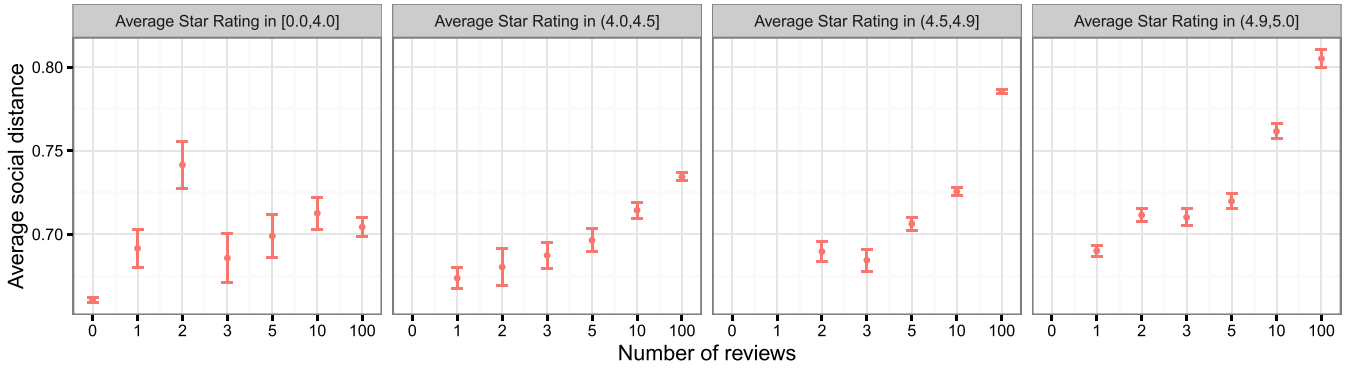

Fig. 3. Real-world data from Airbnb show that an increased reputation of the host in the form of rating (graph) and number of reviews (*x* axis) results in greater diversity of guests who selected them (*y* axis).

that risk aversion was not correlated with reduced trust in general; restricted to the case of high reputation, trust had a strongly negative correlation with risk aversion. The more risk-averse the participant, the less they trusted the positive information provided by the reputation system. Interestingly, risk aversion did not seem to correlate with distrust in negative reputations.

Real-World Data Analysis. The intuitions we gained from the experiment suggested that reliance on reputation may reduce the user's attention to the number of dimensions in which their partner's demographic characteristics differ when they select a host or a guest. In the experimental data, we had access to every option the participant considered and the degree of preference for every pair of them. Although this is not possible to observe from Airbnb's internal database of historical interactions, we sought to use the insights gained through the experiment to guide a real-world, large-scale data analysis and extract the same intuition.

On Airbnb, guests are the active participants in the social selection process. They select partners through searching and making a request. We studied 1 million requests to stay by guests taking place over the same period as the study. For purposes of this analysis, we considered two dimensions of social distance: age and gender. (Airbnb does not collect marital status, and the hospitality interactions usually occur between users from distinct locations.)

For each of these demographic features, we coded distance as 0 if values for hosts and guests were available and equal, and 1 otherwise. We considered two ages equal if they were within 10 y of each other. (We repeated the analysis using different age thresholds within which we considered two people as belonging to the same age group, namely 3, 5, 10, and 20 y. Across the different experiments, the absolute values of social distance changed, but the trends did not.)

Fig. 3 shows how the average social distance between guests and hosts varied, conditioned on the number of reviews and on the star rating of the host at the time of booking request. Strikingly, the intuition we derived in the experiment held in the actual platform. We saw a trend for the average social distance to increase, which became clearer as the number of reviews (within each graph in Fig. 3) and ratings (across graphs in Fig. 3) changed. Note that this effect is not simply explained by an increased number of interactions—we did not see a significant increase in social distance for hosts with a large number of reviews and low ratings (Fig. 3, first graph). This shows that our experimental findings were not simply an artifact of our online laboratory, but that our main conclusions generalized to patterns found in real interactions. That is, as high reputation tends to shrink social distance, we saw higher tolerance for individuals at farther social distances between guests and their selected hosts as the reputation of the host got better.

Discussion

Companies operating in the sharing economy are predicated on trust, but cannot rely directly on preexisting institutional arrangements. Our work shows evidence that the reputation system of Airbnb, and by extension of sharing-economy sites—the star ratings and the number of reviews—may operate to bridge the gap between institutionally generated trust and the organically grown trust present in social platforms. Although we gathered evidence for the tendency of individuals to trust similar others, by trusting the reputation system, participants in our study were willing to extend trust to those who exhibited a high degree of dissimilarity in the social space.

While we present evidence that these effects are at work in the actual Airbnb platform, our experimental results are limited to the specific population that participated in our study. Moreover, although we found very sizable effects associated with homophily, we emphasize that the inclusion of other demographic characteristics not displayed explicitly by the platform, but that can be inferred indirectly through pictures or other signals—such as nationality, race, class, religion, ethnicity, etc.—could lead to the observation of even greater effects. For example, the literature suggests that racial features play a significant role in determining trust (23, 24).

Materials and Methods

Our experimental methods were reviewed and approved by Stanford University's Internal Review Board (protocol 34470, approved on August 11, 2015). We required invitees to provide us with consent to participate in the study, whose terms we displayed on the entry page of our experiment's website.

See *Supporting Information* for detailed information on our sample, an analysis of self-selection, and more information on our research design and data analysis.

ACKNOWLEDGMENTS. This work was supported by National Science Foundation Grant 1257138 (to B.A., P.P., and K.S.C.).

1. Stein J (January 29, 2015) Baby, you can drive my car, and do my errands, and rent my stuff. *Time*. Available at http://time.com/magazine/us/3687285/february-9th-2015-vol-185-no-4-u-s/. Accessed August 9, 2017.
2. Simmel G (1955) *Conflict and the Web of Group Affiliation* (The Free Press, New York).
3. Lazarsfeld P, Merton R, (1954) Friendship as a social process: A substantive and methodological analysis. *Freedom and Control in Modern Society*, eds Berger M, Abel T, Page C (Van Nostrand, New York), pp 18–66.
4. McPherson M (1983) An ecology of affiliation. *Am Socio Rev* 48:519–532.
5. McPherson M, Smith-lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27:415–444.
6. Blau P (1977) A macrosociological theory of social structure. *Am J Sociol* 83:26–54.
7. Resnick P, Kuwabara K, Zeckhauser R, Friedman E (2000) Reputation systems. *Commun ACM* 43:45–48.
8. Earle T (2010) Trust in risk management: A model-based review of empirical research. *Risk Anal* 30:541–574.
9. Snijders C, Weesie J (2009) Online programming markets. *eTrust: Forming Relationships in the Online World*, The Russell Sage Foundation Series on Trust, (Russell Sage, New York), pp 166–186.
10. Houser D, Wooders J (2006) Reputation in auctions: Theory, and evidence from eBay. *J Econ Manag Strat* 15:353–369.
11. Lin T, Abrahao B, Kleinberg R, Lui J, Chen W (2014) Combinatorial partial monitoring game with linear feedback and its applications. *Proceedings of the 31st International Conference on Machine Learning* (Proceedings of Machine Learning Research, Beijing).
12. State B, Abrahao B, Cook K (2012) From power to status in online exchange. *Proceedings of the 4th ACM Int'l Conference on Web Science* (Association for Computing Machinery, New York).
13. State B, Abrahao B, Cook K (2016) Power imbalance and rating systems. *Proceedings of the 10th International AAAI Conference on Web and Social Media,* Cologne, Germany (AAAI Press, Palo Alto, CA).
14. Parigi P, State B (2014) Disenchanting the world: The impact of technology on relationships. *Social Inform* 1:166–182.
15. Cook K, Cooper R (2003) Experimental studies of cooperation, trust, and social exchange. *Trust and Reciprocity: Interdisciplinary Lessons for Experimental Research*, eds Ostrom E, Walker J (Russell Sage, New York), pp 209–244.
16. Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity, and social history. *Games Econ Behav* 10:122–142.
17. Houser D, Schunk D, Winter J (2010) Distinguishing trust from risk: An anatomy of the investment game. *J Econ Behav Organ* 74:72–81.
18. Poundstone W (1992) *Prisoner's Dilemma* (Doubleday, New York).
19. Weber E, Blais A, Betz E (2002) A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *J Behav Decis Making* 15:263–290.
20. Gneezy U, Potters J (1997) An experiment on risk taking and evaluation periods. *Q J Econ* 112:631–645.
21. Pinheiro J, Bates DM (2000) *Mixed-Effects Models in S and S-PLUS* (Springer, New York).
22. McFadden D (1974) Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, ed Zarembka P (Academic, New York), pp 105–142.
23. Edelman B, Luca M (2005) Digital discrimination: The case of airbnb.com. *SSRN*. (Harvard Business School, Boston) Harvard Business School Working Paper 14-054.
24. Smith S (2010) Race and trust. *Annu Rev Sociol* 36:453–75.
25. Fausto-Sterling A (1992) *Myths of Gender: Biological Theories About Women and Men* (Basic, New York).
26. Zervas G, Proserpio D, Byers J (2015) A first look at online reputation on Airbnb, where every stay is above average. *SSRN*. Available at https://ssrn.com/abstract=2554500. Accessed August 9, 2017.
27. Santana J, Parigi P (2015) Risk aversion and engagement in the sharing economy. *Games* 6:560–573.