Journal of Behavioral and Experimental Economics 122 (2026) 102556

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Stereotypes or prejudice: Behavioral evidence of gender discrimination from rural India[☆]

Rahul Mehrotra

Graduate Institute of International and Development Studies, Chemin Eugene-Rigot 2, CP 76, CH-1211 Geneva 21, Switzerland

ARTICLE INFO

Dataset link: https://zenodo.org/records/17109143

JEL classification:
J16
C93
D03
O12
O20

Keywords:
Gender
Discrimination
Trust
Social preferences
Lab-in-the-field
India

ABSTRACT

Discriminatory social norms drive high levels of gender inequality in India. However, there is a paucity of evidence on how gender discrimination manifests in economic decision-making with real payoffs in a rural, underdeveloped setting. In this paper, I present a lab-in-the-field experiment using incentivized trust and dictator games to distinguish between statistical and taste-based gender discrimination. Negative stereotypes that manifest as lower trust is interpreted as statistical discrimination. Prejudice that manifests as lower trust and social preferences is interpreted as taste-based discrimination. Next, I evaluate whether a behavioral nudge can influence discriminating individuals' preferences over gender versus previous trustworthiness. The evaluation nudge tests whether moving from separate (single-choice) to joint (multiple-choice) evaluation setting triggers a shift from gender-biased to payoff maximizing decision-making. Results indicate that participants demonstrate statistical discrimination. Signaling higher trustworthiness leads to gender unbiased decision-making under joint evaluation, but not under separate evaluation.

1. Introduction

Discriminatory social norms are a prominent reason for high levels of gender inequality, especially among developing countries. Cultural practices including patri-lineality, dowry and male child preference are associated with gender differences in health, education, political and economic outcomes. There is an extensive literature documenting gender inequality in terms of observed economic outcomes (Anderson & Ray, 2010; Anukriti et al., 2022; Sen, 1990; Sukhtankar et al., 2022).[1] Recent experimental literature has highlighted patterns of gender differences in decision-making under different information settings (Barron et al., 2024; Bohren et al., 2023; Delavande & Zafar, 2019; Grossman et al., 2019).[2] However, there is a paucity of lab-in-the-field evidence on how discrimination manifests in economic decision-making in a rural, underdeveloped setting characterized by poor gender equality indicators: first, how do individuals' discriminate in terms of two prominent drivers of economic interactions - trust and social preferences? Second, can discriminating individuals' preferences over gender versus trustworthiness be influenced?

In this paper, I use a field experiment conducted in rural India to investigate whether observed gender discrimination is characterized by statistical discrimination (identified by lower trust) or taste-based discrimination (identified by lower trust and lower social preferences). Statistical discrimination is the result of incomplete information whereby, when there is a lack of knowledge about a person's abilities, stereotypes about their demographic group are used (Aigner & Cain, 1977; Arrow, 1971; Phelps, 1972). Meanwhile, taste-based discrimination is driven by prejudice towards interacting with a specific demographic group. Discriminating individuals are willing to sacrifice economic benefits, like wages or profits, to cater to their aversion (Becker, 1971). Distinguishing between taste-based and statistical

☆ Pre-analysis plan is registered here: https://www.socialscienceregistry.org/trials/1100. Lab documentation, data and program files for replication are available here: https://doi.org/10.5281/zenodo.17109143. I am grateful to the Associate Editor and two anonymous reviewers for their helpful comments. Sanjay Prasad and Purna Chandra Mahanta provided invaluable field research assistance.

E-mail address: rahul.mehrotra@graduateinstitute.ch.

[1] For a review of relevant literature, see Jayachandran (2015).

[2] The bulk of this literature uses laboratory games with university students in developed countries, with the Delavande and Zafar (2019) study in urban Pakistan a notable exception. Bertrand and Duflo (2017) provide an extensive review and highlight the need for more field-based evidence, especially focusing on interventions aimed at weakening gender discrimination.

https://doi.org/10.1016/j.socec.2026.102556
Received 16 January 2025; Received in revised form 10 March 2026; Accepted 18 March 2026
Available online 26 March 2026

discrimination can inform the selection of effective policies to reduce inequality.

Next, I explore whether an evaluation nudge can influence discriminating individuals' preferences over gender versus previous trustworthiness (Bicchieri & Dimant, 2019; Bohnet et al., 2016; Thaler & Sunstein, 2008). Previous research has argued that human beings have two distinct cognition modes: the intuitive and automatic System 1 and the reflective and analytical System 2. Studies have shown that the design and presentation of available choices can influence decision-making by triggering one of these distinct cognition modes. When separately evaluating a single default option, the lack of comparative information leads to System 1 cognition characterized by reliance upon prior beliefs and biases (Bazerman et al., 1998; Kahneman & Miller, 1986). However, joint evaluation of multiple options triggers analytical System 2 cognition by providing additional data-points for comparison (Bazerman & Moore, 2013). I analyze how salient personal characteristics, including gender and trustworthiness, impact the selection of partners in economic interactions. Specifically, I test whether evaluating a single person for selection or rejection triggers System 1 cognition that relies on gender stereotypes. Relatedly, I test whether jointly evaluating multiple people to select a single option triggers System 2 cognition that relies on analytical comparisons of trustworthiness. This evidence can support the design of behaviorally-informed policy interventions to ameliorate discrimination under certain settings.

This study was conducted in 12 villages of Chhattisgarh, a state (province) in central India. State-level poverty rate stood at 40% (double the national average) while female education attainment and maternal mortality rates are also significantly worse than the national average for India (World Bank, 2016). This north-central region comprising of the contiguous states of Bihar, Chhattisgarh, Madhya Pradesh and Uttar Pradesh is characterized by low per capita incomes, high population density and patriarchal norms. Consequently, the selection of Chhattisgarh offers broad external validity for the study's findings across this region.

The experimental design comprises of two phases: in the first phase, I use a within-subjects design whereby all participants play incentivized lab games executed under distinct treatment settings. I use an experimental test to distinguish between statistical or taste-based gender discrimination as proposed by Fershtman and Gneezy (2001). The trust game (as per Berg et al. 1995) provides a behavioral measure of *sender's* trust in the *receiver*. Trust is driven by strategic beliefs regarding *receiver's* trustworthiness and *sender's* risk and social preferences.[3] Therefore, any discrimination towards female *receivers* in the trust game can be driven either by negative gender stereotypes affecting the *sender's* strategic beliefs or prejudice driving lower social preferences. The dictator game (as per Forsythe et al. 1994) helps to distinguish between these competing explanations by eliminating the strategic role of the *receiver*.[4]

In the second phase, *senders* are randomly sorted into two treatment groups and asked to make an evaluation decision under two distinct cognitive settings. The first group is randomly matched with one *receiver* (separate evaluation group), while those in the second group are randomly matched with two *receivers* (joint evaluation group). The matched *receivers'* salient characteristics differ along two dimensions: gender-identity (male or female) and previous trustworthiness (high or low). Under separate evaluation, *senders* can accept or reject the matched *receiver*. Under joint evaluation, *senders* can accept one of the matched *receivers* or reject both. If the match is rejected, the *sender* is paired with an anonymous, randomly drawn *receiver*. This setting allows me to explore how *receivers'* characteristics influence *senders*' decision-making.[5]

Phase I results show that, on average, participants discriminate against female *receivers* only in terms of trust. *Senders* demonstrate lower trust towards randomly matched female *receivers* and this negative impact remains statistically significant even after controlling for *senders*' social preferences and risk aversion. There is no evidence of gender discrimination in terms of social preferences measured in the dictator game. This pattern of results is consistent with statistical discrimination. However, observed statistical discrimination appears rational in this context since the share of transfers returned by female receivers is also lower.[6] Taste-based discrimination is observed among a smaller proportion of lab participants. Thirty three participants (7% of the sample) discriminate towards female *receivers* in terms of both, trust and social preferences. Overall, the average behavior across the full sample is indeed consistent with statistical discrimination.

Phase II results indicate that, on average, both male and female *senders* prefer high trustworthiness partners under joint evaluation as predicted by System 2 cognition. Under separate evaluation characterized by System 1 cognition, results are consistent with own-group bias among women and a status quo bias among men, i.e. women prefer other women, while men do not deviate from their default match. Overall, these findings are consistent with the rejection of taste-based discrimination in the full sample, i.e. on average, there is no aversion among *senders* to engage in transactions with female *receivers*.[7] The observed own-group bias can be explained by prevailing patriarchal norms and low social trust in this study location due to which women may prefer to avoid financial interactions with an unknown, male person. The demographic survey results confirm that a significant majority of the sample agreed with discriminatory statements, while only one-fourth responded positively to a standard survey question on generalized trust in strangers. Related research from this region similarly finds own-group bias among women in decision-making roles, for example: female heads of village councils, who give preference to issues raised by other women (Beaman et al., 2009; Chattopadhyay & Duflo, 2004).

I argue that differentiating between taste-based and statistical discrimination can help inform policy-makers' interventions to reduce gender inequality. Economics research has broadly focused on two types of interventions: quotas and education, while there is also increasing evidence on behavioral interventions from organizational settings. Quotas are effective in promoting equal representation and alleviating discriminatory norms through prolonged exposure over the long-term, however they can also result in negative unintended consequences.[8] On the other hand, educational interventions such as television programs with progressive gender roles and school-based gender sensitization

[3] The behavioral definition of trust states, *an individual trusts if (s)he voluntarily places resources at the disposal of the trustee without any legal commitment, but with an expectation that it will increase the trustor's payoff* (Fehr, 2009). Individual preferences driving trust include risk preferences (Eckel & Wilson, 2004), betrayal aversion (Bohnet & Zeckhauser, 2004), inequity aversion and altruism (Ashraf et al., 2006).

[4] Social preferences measured in the dictator game are shown to be driven by preferences for altruism (Andreoni & Miller, 2002; Charness & Rabin, 2002), inequality aversion (Bolton & Ockenfels, 1998; Fehr & Schmidt, 1999), and 'warm glow' giving effects (Eckel & Grossman, 1996).

[5] This analysis relies only on *senders'* evaluation decision since behavior in the following lab games is endogenous.

[6] This finding provides a important complement to previous evidence from Delavande and Zafar (2019) who find evidence of taste-based gender discrimination among an urban population in Pakistan. These disparate results indicate that local norms influence the type and intensity of observed discrimination in economic decision-making.

[7] This study supplements related laboratory evidence on the impact of an evaluation nudge tested on a sample of college students (Bohnet et al., 2016).

[8] Thernstrom and Thernstrom (2009) show voters may dislike quotas which restrict their choices and resent female leaders. Goldin (2014) argues that quotas can also be perceived to violate social norms and reduce the value of traditionally male activities. As a result, quotas may precipitate a backlash against female leaders and even strengthen taste-based discrimination (Boisjoly et al., 2006; Rudman & Fairchild, 2004).

education can improve negative gender stereotypes (Dhar et al., 2017). However, discriminatory prejudices are less malleable (Banerjee et al., 2013; Beaman et al., 2009; Jensen & Oster, 2009). Therefore, contexts where taste-based discrimination dominates may be more suited for quotas to ensure equal representation while slow-moving norms evolve. Conversely, education and behavioral nudges are more relevant where negative stereotypes dominate. Decision-making in organizational settings can become prone to personal biases, for example: hiring and promotion evaluations. This field research adds to existing lab evidence showing mandates for joint evaluation of multiple candidates using common, performance-related criteria can ameliorate gender biases (Bohnet et al., 2016).

The remaining paper is structured as follows: Section 2 describes the study location, lab-in-the-field games and experimental design. Next, Section 3 reports summary statistics from the lab games and presents the econometric analysis. Section 4 includes robustness tests to verify that main findings are robust to alternate regression specifications and estimation methodologies. Finally, Section 5 concludes with a discussion of the main findings and their implications for policy design. Complete experimental protocols are included in Appendix.

2. Experimental methodology

In this section, I describe the experimental design, lab games and randomized treatments used in this research. The lab experiments are conducted in the state of Chhattisgarh, India. The state-level poverty rate stood at 40%, compared to the national rate of 22%, and it performs poorly on gender indicators. Female secondary education attainment is very low at 16%, while child sex ratio (969 females per 1000 males) and female labor force participation (55% of eligible females work) is also low by global standards. Maternal mortality is significantly higher than the national average equaling 221 deaths per 100,000 live births (World Bank, 2016).

This research aims to analyze patterns of gender discrimination in this rural, under-developed setting. Accordingly, Rajim — a rural town located within few hours' travel-time from the state's capital city Raipur, is selected as the base for the research team. From this central location, the research team visited 12 villages in 3 adjoining districts: Dhamtari, Gariyabandh and Raipur (district name and capital city name is identical). A local contact-person is recruited to assist the team in: (1) advertisement of research study among men and women from all available caste and professional clusters in each village, and (2) identifying an easily accessible and private location for the study within the village.[9] An equal ratio of male and female heads of household or their spouses were targeted for participation. The finalized study locations are close but not contiguous in order to minimize travel time and opportunities for social learning.[10] Fig. 1 provides a spatial visualization of the study locations.

The lab sessions relied on private, paper-based interviews because the participants' low literacy levels preclude the use of computer-based lab protocols.[11] Interviewers used a fixed script to conduct each individual session. Player pairs are formed by randomly matching two individuals across different villages. All individuals from one village participate as *receivers* using the strategy method, while the participants from all 11 remaining villages are *senders*. This one-to-many matching of a single *receiver* with eleven *senders* is required to increase the number of observations and negate inter-personal relationships influencing behavior in the lab. The empirical analysis is based solely on *sender*-level observations. One village-session is conducted per day, with the *receiver*-village participating on the first day followed by *sender*-villages.[12] All participants in each village completed the experimental protocol and demographic survey in a single continuous sitting with their assigned interviewer. In total, the complete study timeline spanned 12 days. Payments to *senders* and *receivers* are explained in Section 2.2.3 after describing the experimental design.

Four hundred and seventy two individuals participated as *senders* and 44 participated as *receivers*. Ex-ante statistical power calculations were used to estimate desired sample sizes required to identify treatment effects across both phases of the study. The standard assumptions used include: statistical power equal to 0.8 and confidence level of 0.05. Estimates from previous studies using similar experimental designs are used as proxies for variance in outcome variables and minimum detectable effect (MDE) sizes. Results from (Fershtman & Gneezy, 2001), Fershtman et al. (2005), and Castillo and Petrie (2010) are used as proxies to estimate sample size for Phase I focusing on anonymous and gender-salient lab games. Results from (Bohnet et al., 2016) are used as proxies to calculate optimal sample size in Phase II focusing on separate and joint evaluation treatments.

A simple graphical representation of the experimental design is provided in Fig. 2 below. The complete experiment protocols are available in the Online Appendix.

2.1. Lab-in-the-field games

This study relies on standard Trust and Dictator Games played under distinct treatment settings between randomly matched players. In addition, all participants participate in a simple risk lottery to calculate their risk aversion. The games are briefly described below:

Trust Game: All participants in the 11 *sender* villages are assigned the role of trustors, while 44 participants from the single *receiver* village are assigned the role of trustees. Both players start the game with an equal endowment of Rs 50. Next, the trustors are asked to allocate their endowment between themselves and the trustees using multiples of Rs 5 (0; 5; 10; 15;...; 45; 50). The trustee receives triple the amount sent ($3X$) and can send back any amount Y between 0 and $3X$. The trustor earns $(50 - X + Y)$ and the trustee earns $(50 + 3X - Y)$. The trustee uses the strategy method to indicate a contingent amount to return for each potential amount which can be received by them.[13] The amount sent by the trustor, i.e. Rs X, is interpreted as a measure of *trust*, while the amount returned by the trustee, i.e. Rs Y, is interpreted as a measure of their *trustworthiness*.

Dictator Game: Same as before, all participants from each of the 11 *sender* villages are assigned the role of dictators, while the 44 participants from the first village are the *receivers*. In this game, only the dictators receive a fixed endowment of Rs 50. Next, the dictators are asked to allocate their endowment of Rs 50 between themselves and the trustees using multiples of Rs 5 (0; 5; 10; 15;...; 45; 50). The *receiver* receives triple the amount sent ($3X$) and then the game ends for both players. The dictator earns $(50 - X)$ and the *receiver* earns $(3X)$. The transfer sent by the dictator, i.e. Rs X, is interpreted as a measure of their *social preferences*.

[9] The contact-person is usually the local grocery-shop owner who is paid a nominal fee for his assistance.

[10] Chaudhuri et al. (2006) show that communication between previous and future participants generates a process of social learning and influences future participants' experimental behavior.

[11] Participants' literacy was measured using a survey question "Can you read and write a simple text?" Only 55% of participants responded positively. Summary statistics for the complete demographic survey of participants are discussed in Section 2.3 and reported in Table 8 in Appendix A.

[12] The strategy method is used to form player-pairs and *receivers*' real earnings are calculated at the end of the experiment, i.e. after all *senders*' have completed their participation. Therefore, it is not possible for the *receivers* to determine the actual amounts sent to them and adjust their responses between rounds.

[13] Casari and Cason (2009) show that using the strategy method is linked to a significantly lower rate of trustworthiness, compared to the game method. However, since my empirical hypotheses are not based on trustees' behavior this does not influence the results of this study.

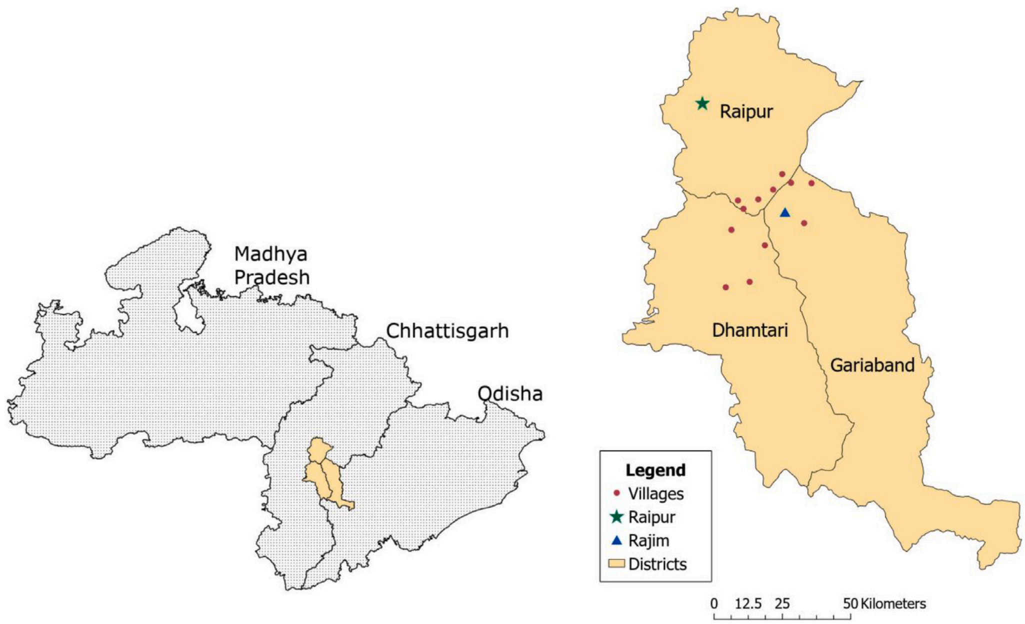

Fig. 1. Study Location.
Notes: This figure provides a cartographic representation of the study location. The sample includes 12 villages from three adjoining districts: Dhamtari, Gariyabandh and Raipur in Chhattisgarh state. Rajim, a rural town situated within few hours' travel time from the state's capital city Raipur, is the study team's central base location.

Risk Lottery: Each participant has to select either a risky gamble or a certain amount across four rows of choices offered to them. The risky gamble is a 50% chance of winning Rs 100 or nothing, and remains constant in each successive row. The certain amount declines in each successive row and consists of Rs 50, 40, 30 or 20 respectively. One row is selected using a random draw. If the risky gamble was selected then the outcome is decided by a coin-toss. The number of risky gambles rejected by each participant is used as a simple measure of their *risk aversion*.[14]

2.2. Experimental design

In this section, I will describe the experimental protocols and the specific empirical hypotheses that are evaluated. Each lab session starts with registration of participants, introduction to the study and recording of informed consent to participate. The session consists of three phases: Phase I comprises of two rounds of lab games played under distinct treatment settings. The round order is kept fixed. Potential carry-over effects are addressed by including a risk lottery and gender priming exercise after the first round of lab games. Player-pairs are randomly re-matched for each round of lab games. In Phase II, *senders* are randomly split into two treatment groups. Each treatment group plays one round of lab games. Finally, Phase III consists of a simple household survey used to record participants' demographic characteristics before the session concludes with payment of real earnings.

2.2.1. Phase I: Statistical vs. Taste-based discrimination

The implementation of Phase I is described below:

Gender-Anonymous Round: Each *sender* is randomly matched with a *receiver* using a random-draw of an anonymous and unique serial number. Both players participate under double-blind setting whereby no information regarding their co-player's identity or geographic location is revealed. After the match is completed and recorded, the first round of trust and dictator games are executed.

Risk Lottery and Gender Priming: All senders' participate in a risk lottery, as described previously, and a simple priming task (as per Shih et al. 1999) before the gender-salient round. The objective of this exercise is to increase participants' sensitivity to gender identities by highlighting commonly held gender roles in the study area.[15] Experimenters read aloud four statements. Participants respond to each statement using the scale: *strongly agree, agree, disagree,* or *strongly disagree*. The statements for male participants are: (i) *As a man, I help my family by earning an income*; (ii) *As a man, I am strong in order to protect myself or my family against outsiders*; (iii) *As a man, I make the important investment decisions in my household*; (iv) *As a man, I know how to operate machinery*. The corresponding statements for female participants are: (i) *As a woman, I help my family by taking care of the house*; (ii) *As a woman, I maintain good relations with relatives and neighbors*; (iii) *As a woman, I make the daily, household decisions*; (iv) *As a woman, I know how to sew clothes and cook food.*

Gender-Salient Round: After priming their own gender identity, the *sender* is next matched with a new *receiver* using a random draw. In this round, the random draw includes the serial number, first-name, and gender of the corresponding *receiver*, which is revealed to *sender*.[16] The *receiver's* first-name and gender is read aloud by the interviewer in the lab instructions. After the instructions are understood by the *sender*, the second round of trust and dictator games are executed.

[14] Multiple variations of these lab games have been proposed in the literature to observe more sophisticated measurements of these behavioral parameters. For this study context, I use the basic variations to maximize successful implementation between experimenters and study participants.

[15] The design of this priming exercise is based on advance field-testing of the experimental protocol. All statements reflect gender roles which are uniformly understood in this rural study region.

[16] The caste affiliation of individuals can be inferred from their last names, therefore only *receiver's* first-name is revealed. This negates the influence of caste affiliations on *sender's* behavior while still making *receiver's* gender-identity salient. *Receivers* provide their consent for their names to be recorded and revealed in this round. As described earlier, player-pairs are matched across distinct, non-contiguous villages whose names are kept anonymous.

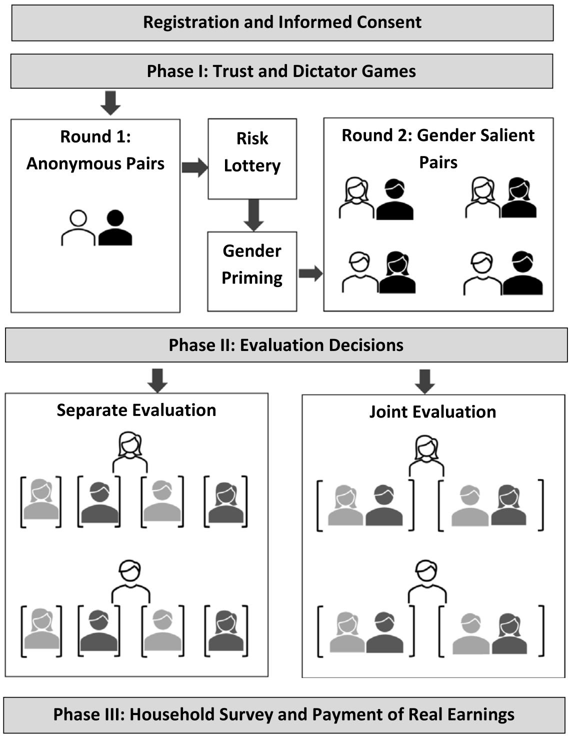

Fig. 2. Within-subjects Experimental Design.
Notes: This illustration denotes the experimental design. The anthropometric figures indicate whether the participant is anonymous, male or female. Transparent figures indicate the *senders*, while solid figures indicate the *receivers*. Light-grey figures indicate high trustworthiness *receivers* and dark-grey figures indicate low trustworthiness *receivers*. After recording informed consent, Phase I comprises of two rounds of trust and dictator games, respectively. In the first round, an anonymous player-pair is randomly matched. Next, there is a risk lottery and a gender beliefs elicitation exercise. For the second round of games, player-pairs are randomly re-matched and their gender identity is revealed. In Phase II, *senders* are randomized into either a Separate Evaluation or Joint Evaluation group. Under Separate Evaluation, each *sender* is randomly matched with one of the following *receiver*-types: [*female-high trustworthiness*], [*female-low trustworthiness*], [*male-high trustworthiness*] or [*male-low trustworthiness*]. *Senders* either select or reject this match in favor of an anonymous receiver. Under Joint Evaluation, each *sender* is randomly matched with two *receivers* of opposing types: [*female-high trustworthiness* and *male-low trustworthiness*] or [*male-high trustworthiness* and *female-low trustworthiness*]. *Senders* can select one of the matches or reject both in favor of an anonymous receiver. Another round of lab games is implemented after pairs are matched across both evaluation groups. Finally, Phase III concludes with a household survey to record demographic characteristics and payment of earnings.

Overall, gender salience in this study is interpreted as a combination of priming the *sender's* own gender identity plus the revelation of the *receiver's* gender identity. This experimental design allows me to test two empirical hypotheses: (1) randomly matched *receiver's* salient gender identity has no impact on *sender's* trust; (2) randomly matched *receiver's* salient gender identity has no impact on *sender's* social preferences. Taste-based discrimination is inferred if both hypotheses are rejected, i.e. *senders* discriminate against female *receivers* in terms of both the belief component of trust as well as social preferences. Statistical discrimination is inferred if only the second hypothesis is rejected. If both hypotheses fail to be rejected, we can exclude the presence of gender discrimination in the study sample.

2.2.2. *Phase II: Separate vs joint evaluation*

In Phase II, *senders* are randomized into either a Separate Evaluation or Joint Evaluation group. The objective is to test the impact of *receivers'* gender and trustworthiness on *sender's* willingness to transact under different cognitive settings. The implementation proceeds as follows:

Pre-selection of *receivers*: To introduce controlled variation in *receiver* characteristics, I select a group of 4 *receivers*, who vary according to gender (2 males and 2 females) and previous trustworthiness (2 with average level of trustworthiness and 2 with below-average trustworthiness). This criterion is used because if matched *receiver's* trustworthiness is significantly below the group average, then *sender's* selection decision will be affected by loss aversion. Similarly, if matched *receiver's* trustworthiness exceeds the group average, there will be no incentive to consider the anonymous option.

Receivers' trustworthiness observed under gender-anonymous setting is used for this selection. Average trustworthiness observed across the pool of *receivers* is equal to Rs 3, i.e. on average, *receivers* return Rs 3 more than the amount sent to them by *sender*. Meanwhile, the below-average amount used is Rs 1, i.e. on average, the low trustworthiness *receiver* returns Rs 1 more than the amount sent to them by *sender*.[17]

Accordingly, the group of pre-selected *receivers* will consist of the following individuals: (i) *High Trustworthiness-Male*: Male whose trustworthiness in the gender-anonymous round equals the average trustworthiness (Rs 3); (ii) *High Trustworthiness-Female*: Female whose trustworthiness in the gender-anonymous round equals the average trustworthiness (Rs 3); (iii) *Low Trustworthiness-Male*: Male whose trustworthiness in the gender-anonymous round is below the average trustworthiness (Rs 1); (iv) *Low Trustworthiness-Female*: Female whose trustworthiness in the gender-anonymous round is below the average trustworthiness (Rs 1).

Separate Evaluation: *Senders* in this group are randomly matched with one of the pre-selected *receivers* using a random draw: *High Trustworthiness-Male*, *Low Trustworthiness Female*, *High Trustworthiness-Female*, *Low Trustworthiness-Male*. The experimenter then reveals the following information-set: (1) gender of matched *receiver*, (2) matched *receiver's* trustworthiness in the gender-anonymous round, and (3) average trustworthiness of all *receivers* in the gender-anonymous round. The *sender* is asked to either accept or reject the matched *receiver*. If the *sender* accepts the match, then they play the third round with the selected *receiver*. However, if the *sender* rejects the match, they have to randomly draw another anonymous *receiver* from the pool to play the third round.

Joint Evaluation: *Senders* in this group are randomly matched with two of the pre-selected *receivers*. Since the intention is to introduce variation in gender and trustworthiness, there are two possible pairs: (*High Trustworthiness-Male*, *Low Trustworthiness-Female*), or (*High Trustworthiness-Female*, *Low Trustworthiness-Male*). The experimenter then reveals the following information-set: (1) gender of both matched *receivers*, (2) male *receiver's* trustworthiness, and (3) female *receiver's* trustworthiness, and (4) average trustworthiness of all *receivers*. The *sender* is asked to select one of the two matched players for the final round of lab games, or reject them both. If the *sender* rejects both players, they have to randomly draw another anonymous *receiver* from the pool.

The observed behavior across both groups is used to analyze the impact of randomly matched *receivers*' gender and trustworthiness on *senders*' selection decision. Specifically, the separate evaluation group is used to test the following single hypotheses: (1) randomly matched *receiver*'s gender has no impact on the selection-decision by *senders*; (2) randomly matched *receiver*'s trustworthiness has no impact on the selection-decision by *senders*. If these hypotheses are rejected, we can determine *sender*' preferences over the receivers' salient characteristics. Failure to reject these hypotheses under separate evaluation will be

[17] Since *receivers* use the strategy method to indicate the amount returned for each contingent amount sent to them, the formula used to calculate mean trustworthiness is: $Mean\ Trustworthiness = \frac{Total\ Amount\ Returned - Total\ Amount\ Received}{Number\ of\ Contingent\ Choices}$.

consistent with the default-choice effect, i.e. participants prefer to stick with the offered choice. Next, the joint evaluation group is used to test a hypothesis that combines both parameters of interest: (3) randomly matched *receiver*-pair's gender and trustworthiness have no joint impact on the selection decision by *senders*. Failure to reject this joint hypothesis will indicate that the evaluation nudge does not have its intended impact.[18]

2.2.3. Payments to lab participants

All participants contribute time that is normally spent on other productive activities to participate in this study. The average individual session duration is approximately 45 min, while the total study timeline spans 12 days, i.e. one day per village. All *senders* are informed they will be paid their real earnings from one row of the risk lottery and one randomly selected round of lab games. *Senders* use a lottery to randomly select the round of lab games to be used to calculate their earnings after their session. Amounts earned from each game are explained to all individuals. Mean earnings for *senders* are Rs 145 (minimum = Rs 70 ; maximum = Rs 210). The mean equals half the daily agricultural wage-rate in the study region.

Real earnings for *receivers* - who participate using the strategy method on day 1 - cannot be calculated until all 11 *sender*-villages have participated. However, it is not possible to invite *receivers* to participate based solely on trust in the study team to return approximately 2 weeks later with their earnings. Accordingly, the *receivers* are explained the study timeline and informed they will be paid a minimum participation fee of Rs 150 on the same day, plus any excess amount earned in the lab games after approximately 2 weeks. The fee amount is jointly determined with the local contact-person to encourage participation. At the end of the study, total earnings are calculated using a random lottery to select matched senders and game rounds. Amounts earned from each game are explained. If total lab earnings are more than Rs 150, this excess amount is paid as well. Mean payments to *receivers* equaled Rs 157 which corresponds to 54% of the daily agricultural wage-rate. Maximum amount paid equaled Rs 210 and minimum was Rs 150 by design. This payment protocol for *receivers* is necessary to boost sample size of sender-level observations without matching player-pairs within the same village. Interpersonal transactions involving real payoffs risk future repercussions between known participants, especially targeting vulnerable groups of women and underprivileged caste groups. All empirical hypotheses evaluated in this study focus on *senders*' decisions.[19]

Omissions in *Senders*' Lab Instructions: Recent experimental literature and professional surveys argue against a binary definition whereby any omission from explaining the complete experimental protocols to all participants equals deception (Charness et al., 2022; Krawczyk, 2019). Accordingly, the lab instructions for *senders* omit a description of the strategy method, associated matching probabilities and *receivers*' payment protocol. These omissions are implemented because: first, simplified lab instructions are critical to ensure uniform comprehension of lab instructions across this multi-phase experiment (survey data confirmed only 55% can read or write a simple text). Simplification reduces cognitive load among both participants and human experimenters who repeat multiple individual sessions per day. Second, these omissions are evaluated by the study team to confirm there are no negative impacts on participants' earnings and experimental control, either in this study or future studies with the same study population. Finally, each individual receives complete feedback on their personal outcomes at the end of their participation.

2.3. Phase III: Demographic survey

In the final Phase III, a short survey is used to record all participants' demographic characteristics. The total sample comprises of an equal ratio of male and female participants with household heads or their spouses comprising 93% of *senders*. The average age is 41 years and a large majority of participants are married. A little more than half are literate, i.e. able to read and write a simple text. A significant majority of *senders* belong to the other backward castes (OBC) category. This is characteristic of the overall population in the study region which predominantly consists of OBC population. Similarly, the agricultural sector is the major source of employment with 73% of *senders* working as either wage-labourers or self-employed in agriculture. Land ownership is low equaling 1.2 acres on average. Only 27% of *senders* responded positively to a survey question on generalized trust in strangers, indicating a relatively low level of social capital among the sample. On average, *senders* rejected 1.41 risky gambles in the risk lottery which indicates that the sample is risk neutral. Finally, 61% of the sample agreed with a survey measure of self-assessed gender discriminatory attitudes. These descriptive statistics are reported in Table 8 in the Online Appendix, alongside a balance check between *senders* randomly matched with female vs. male *receivers*. There are no statistically significant differences in these observed characteristics across both groups.

3. Empirical analysis

In this section, I report the lab games' summary statistics and empirically test the study hypotheses using regression analysis. The summary statistics for Phase I lab games are reported in Table 1. As described in the previous section, player-pairs are formed at the start of both rounds by randomly matching 472 *senders* with 44 *receivers* who participate using the strategy method. All analyses focus on the sender-level observations. Columns (1) and (2) report the number of player-pairs and mean (with standard deviations) amounts sent in the gender-anonymous setting. Columns (3) and (4) report the same information for the gender-salient setting. Column (5) reports the results from a two-tailed t-test for significant differences in means. Trust (Rs) refers to the mean amounts sent in the trust games for the full sample of 472 player-pairs. The following rows report the means for sub-samples disaggregated by the randomly matched player-pairs' gender. Similarly, Social Preferences (Rs) refers to the mean amounts sent in the dictator games for the full sample and rows below report the sub-sample means disaggregated by the player-pairs' gender.

Summary statistics indicate there is no significant difference in the mean amount sent by *senders* to *receivers* in the trust games. However, there is a statistically significant reduction in mean trust amount sent by both male and female *senders* to female *receivers* after their gender-identities are made salient (Rs 9.69 vs Rs 7.70, p-value=0.00). In particular, male *senders* send less to female *receivers* (Rs 10.21 vs Rs 7.80, p-value=0.03). There are no significant differences in the dictator games. The lab data indicates that 118 *senders* (57 men and 61 women) discriminate in the trust game only, i.e. send a lower amount when gender identity is made salient. Meanwhile, 57 *senders* (27 men and 30 women) discriminate in both trust and dictator games. The frequency distribution of amounts sent in both lab games are illustrated in Figure 3 in the Online Appendix. In summary, the figures indicate right-skewed distributions, i.e. a significant proportion of *senders* prefer to send low amounts across both games. The distribution of shares returned to *senders* by *receivers* indicate a bi-modal distribution, i.e. most *receivers* return either 1/3rd or 2/3rd of the tripled amounts sent to them.

[18] Note that the total set of unique combinations include 6 possible player-pairs that differ by gender, trustworthiness or both. However, a considerably larger sample size is required to identify significant differences in means across additional sub-groups with a reliable level of statistical power. This study opts for a simpler, joint comparison across two most relevant pairs. Future research with larger sample sizes can build on these findings.

[19] If the minimum participation fee is excluded from the calculation, the real earnings of *receivers* from their randomly selected round of lab games equaled Rs 126 on average (minimum = 75, maximum = 210).

Table 1
Summary statistics: Phase I lab games, by *Sender–Receiver* gender.

	Round One: Gender-anonymous		Round Two: Gender-salient		*t-test* for
	Number of pairs (1)	Mean (Std. dev.) (2)	Number of pairs (3)	Mean (Std. dev.) (4)	difference in means (*p*-value) (5)
Trust (Rs.)	472	9.58 (9.36)	472	8.78 (9.27)	0.19
Female-Female	121	9.17 (8.67)	119	7.61 (6.95)	0.12
Male-Female	118	10.21 (8.85)	116	7.80 (7.73)	0.03**
Both-Female	239	9.69 (8.76)	235	7.70 (7.32)	0.00***
Female-Male	118	9.66 (9.38)	120	10 (10.89)	0.80
Male-Male	115	9.26 (10.56)	117	9.70 (10.69)	0.75
Both-Male	233	9.46 (9.96)	237	9.85 (10.77)	0.68
Social Preferences (Rs.)	472	4.82 (7.01)	472	4.38 (7.17)	0.34
Female-Female	121	4.38 (6.97)	119	3.53 (5.50)	0.30
Male-Female	118	4.87 (5.93)	116	4.48 (6.99)	0.65
Both-Female	239	4.62 (6.47)	235	4.00 (3.19)	0.29
Female-Male	118	4.15 (5.96)	120	4.21 (7.48)	0.95
Male-Male	115	5.91 (8.80)	117	5.30 (8.39)	0.59
Both-Male	233	5.02 (7.53)	237	4.75 (7.94)	0.70

Notes: This table reports the Phase I lab games summary statistics. Columns (1) and (3) report the number of randomly matched player-pairs. Columns (2) and (4) report the mean and standard deviations from the anonymous and gender-salient rounds respectively. Column (5) reports results from a t-test for significant differences in means. Trust (Rs) refers to the mean amounts sent in the trust games for the full sample of 472 player-pairs. The following rows report the means for sub-samples disaggregated by the gender identities of the randomly matched player-pairs. Similarly, Social Preferences (Rs) refers to the mean amounts sent in the dictator games for the full sample and the following rows report the sub-sample means disaggregated by gender identities. Player-pairs are formed at the start of both rounds by randomly matching 472 *senders* with 44 *receivers* who participate in the lab games using the strategy method. All analysis in this study focuses on *sender*-level observations.

Next, I report the summary statistics for Phase II lab games in Table 2. As described in the previous section, the full sample of 472 *senders* are now randomly divided between two groups: separate evaluation group (N=235) and joint evaluation group (N=237). The separate evaluation group can either select a matched *receiver* whose gender and trustworthiness is revealed, or else reject them in favor of an anonymous *receiver*. Column (1) reports the number of player-pairs formed by implementing this procedure and column (2) reports the mean (with standard deviations) amounts sent in the following lab games. The joint evaluation group can select either one out of two matched *receivers*, else reject both and play with an anonymous *receiver*. Column (3) reports the number of player-pairs formed by implementing this procedure in the joint evaluation group and column (4) reports the mean (with standard deviations) amounts sent in the following lab games. Column (5) reports the results from a two-tailed t-test for significant differences in means.

At this stage, it is useful to note the primary objective of Phase II is to analyze the causal impact of randomly matched *receivers* gender and trustworthiness on *senders*' selection decision under distinct evaluation settings. This is accomplished using regression analysis in the following section. The amounts sent in the Phase II lab games, as reported in Table 2, are endogenous and driven by the selection decisions of *senders*.

With this caveat, the Phase II summary statistics indicate a broadly similar pattern of mean amounts sent by both evaluation groups. However, male *senders* who select anonymous *receivers* demonstrate significantly lower trust (Rs 11.43 vs Rs 4.06, p-value= 0.01) and social preferences (Rs 8.93 vs Rs 2.19, p-value= 0.01). This indicates that, on average, men who reject both matches under joint evaluation are associated with lower generalized trust and social preferences in other people. Female *senders* who select male *receivers* demonstrate higher social preferences under joint evaluation (Rs 3.71 vs Rs 8.27, p-value= 0.02), indicating that on average, women demonstrate higher altruism towards male *receivers* when offered multiple choices. Fifty three *senders* (out of 235) reject their matched *receiver* under separate evaluation. Only 31 (out of 237) reject the matched *receivers* under joint evaluation. This provides descriptive evidence that rejection rates are lower when multiple options are presented for evaluation.

Finally, the distribution of strategies chosen by participants in both Phase II lab games are illustrated in Figure 4 in the Online Appendix. Similar to Phase I, these figures indicate right-skewed distributions whereby most *senders* prefer to send relatively low amounts in both lab games.

3.1. Statistical or taste-based gender discrimination?

Next, I estimate linear regression models using ordinary least squares (OLS). The following regression equation is estimated separately on cross-sectional data from each of the first two rounds:

$$Y_{ijv} = \beta_0 + \beta_1 F_{ijv} + \beta_2 X_{ijv} + \gamma_j + \delta_v + \epsilon_{ijv} \quad (1)$$

where Y_{ijv} is the outcome variables of interest from the lab-in-the-field games for *sender* i, interviewed by interviewer j in village v (i.e. trust amount sent, trustworthiness transfers received, social preferences amount sent), β_0 is the constant term, and F_{ijv} is a binary variable equal to 1 if *sender* i is randomly matched with a female *receiver*, and 0 otherwise. The coefficient β_1 is expected to be negative and statistically significant if participants discriminate against female co-players. X_{ijv} is a vector of participant i's demographic characteristics and risk aversion.

Table 2
Summary statistics: Phase II lab games, by *Sender–Receiver* gender.

	Separate evaluation		Joint evaluation		*t-test* for
	Number of pairs (1)	Mean (Std. dev.) (2)	Number of pairs (3)	Mean (Std. dev.) (4)	difference in means *p*-value (5)
Trust (Rs.)	235	10.79 (11.18)	237	10.80 (11.25)	0.98
Female-Female	56	9.82 (9.62)	83	10.18 (11.59)	0.82
Male-Female	47	12.55 (11.08)	50	13.80 (13.68)	0.62
Female-Male	35	11.71 (12.00)	26	12.31 (8.39)	0.82
Male-Male	44	10.34 (11.88)	47	10.85 (10.75)	0.83
Female-Anonymous	25	6.80 (10.69)	15	8.67 (7.19)	0.51
Male-Anonymous	28	11.43 (9.98)	16	4.06 (5.84)	0.01***
Social Preferences (Rs.)	235	4.81 (7.76)	237	4.73 (8.00)	0.93
Female-Female	55	5.45 (10.24)	83	3.15 (5.53)	0.13
Male-Female	47	4.68 (6.95)	50	7.10 (11.48)	0.21
Female-Male	35	3.71 (5.05)	26	8.27 (8.71)	0.02**
Male-Male	44	3.98 (5.96)	47	3.83 (6.44)	0.91
Female-Anonymous	25	2.00 (5.20)	15	5.33 (8.34)	0.18
Male-Anonymous	28	8.93 (9.46)	16	2.19 (5.15)	0.01***

Notes: This table reports the Phase II lab games summary statistics. The total study sample of 472 senders are randomly assigned between two groups: separate evaluation group (N=235) and joint evaluation group (N=237). The separate evaluation group can either select or reject one randomly matched receiver whose gender and previous trustworthiness is revealed. If the match is rejected, senders play the lab games with an anonymous receiver. The joint evaluation group is offered a selection of two randomly matched receivers who differ in terms of gender and previous trustworthiness. Senders can select one of two matched receivers, else reject both and play the lab games with an anonymous receiver. Columns (1) and (3) report the number of player-pairs formed by implementing this procedure for separate evaluation and joint evaluation groups, respectively. Columns (2) and (4) report the mean and standard deviations of the amounts sent in the lab games for both groups, respectively. Column (5) reports results from a t-test for significant differences in means. Trust (Rs) refers to the mean amounts sent in the trust games. The following rows report the means for sub-samples disaggregated by the identities of the player-pairs. Similarly, Social Preferences (Rs) refers to the mean amounts sent in the dictator games and the following rows report the sub-sample means disaggregated by identities. All analyses focus on the *sender*-level observations.

The equation includes interviewer fixed effects γ_j and village fixed effects δ_v (which correspond to session fixed effects) to control for any interviewer, village, and session-specific effects. Finally, ϵ_{ij} is the error term. Huber-White heteroskedasticity-robust standard errors are estimated for inference.[20] I also report p-values from wild-bootstrap and randomization inference procedures respectively.

Table 3 reports estimation results for Phase I. Panels A and B report the impact of being randomly matched with a female *receiver* in Round 1: Gender Anonymous and Round 2: Gender Salient, respectively. Panel C uses a pooled regression with lab data from both rounds and reports the interaction between random match with a female *receiver* and an indicator for the gender-salient round. In columns (1) and (2), I report the impact of trustee gender being female on trust, i.e. amount sent in the trust game. In columns (3) and (4), I report the impact of trustee being female on trustworthiness, i.e. share of the amount received by the trustee returned to the trustor. This sample only includes observations where a positive amount was sent in the trust game because trustworthiness cannot be calculated if *senders* did not send any money to *receivers* in the trust game. In columns (5) and (6) I report the impact of trustee being female on social preferences, i.e. amount sent in the dictator game. All even-numbered columns include *senders'* demographic characteristics and risk aversion as covariates. One participant did not respond to the demographic survey question on generalized trust in strangers, therefore they are dropped from the sample when covariates are included in the estimations.

In Panel A, there is no statistically significant impact of being matched with a female *receiver* on the amounts sent in the trust and dictator games. This is the expected result and simply confirms there is no spurious treatment effects when *receivers* are anonymous. Next in panel B, there is evidence of discrimination towards women in terms of trust when gender is made salient in columns (1) and (2). After controlling for risk preferences and demographic covariates, the estimated coefficient indicates that participants send Rs 1.90 less to women (equivalent to 22% of the mean trust amount sent). The coefficients in columns (5) and (6) are negative but not statistically significant, allowing me to confirm that there is no impact of revealing the gender on participants' social preferences. The unilateral transfer by *senders* in the dictator game is low on average (Rs 4.60 across both rounds), however 45% of participants send a positive amount and there is no significant discrimination towards female *receivers*. Finally, the results in columns (3) and (4) indicate that female *receivers* also return significantly lower amounts to the sender.

Finally, Panel C confirms the presence of gender discrimination in terms of trust, also when comparing outcomes across both gender-anonymous and gender-salient rounds of lab games. In column (1), the estimated coefficient on the interaction-term indicates that participants send Rs 2.52 less to women (equivalent to 27% of the mean across both rounds). The coefficient narrowly loses statistical significance at

20 Clustering standard errors is not recommended in this setting since treatment is allocated at the individual level. Furthermore, there are too few *sender* villages to obtain unbiased standard errors.

Table 3
Impact of random match with female receiver.

	Dependent variables: Trust games Amount sent (Rs)		Trust games Share returned (%)		Dictator games Amount sent (Rs)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Gender Anonymous Round						
Matched with Female *Receiver*	0.51	0.10	−0.10***	−0.09***	−0.34	−0.44
	(0.84)	(0.87)	(0.02)	(0.03)	(0.61)	(0.65)
	{0.54}	{0.92}	{0.00}	{0.00}	{0.58}	{0.47}
	[0.53]	[0.91]	[0.00]	[0.00]	[0.58]	[0.49]
Observations	472	471	357	356	472	471
R^2	0.07	0.11	0.09	0.13	0.11	0.14
Mean Dependent Variable	9.58	9.54	0.45	0.45	4.82	4.79
Panel B: Gender Salient Round						
Matched with Female *Receiver*	−1.90**	−1.91**	−0.08***	−0.08***	−0.60	−0.60
	(0.83)	(0.87)	(0.02)	(0.02)	(0.63)	(0.64)
	{0.02}	{0.03}	{0.00}	{0.00}	{0.36}	{0.34}
	[0.02]	[0.02]	[0.00]	[0.00]	[0.35]	[0.35]
Observations	472	471	335	334	472	471
R^2	0.09	0.11	0.07	0.11	0.09	0.14
Mean Dependent Variable	8.78	8.72	0.40	0.40	4.38	4.34
Panel C: Pooled Regression using Both Rounds						
Matched with Female *Receiver*	3.06	2.47	−0.13**	−0.13**	−0.02	−0.13
	(1.87)	(1.87)	(0.05)	(0.05)	(1.39)	(1.39)
	{0.11}	{0.19}	{0.01}	{0.02}	{0.99}	{0.93}
	[0.11]	[0.19]	[0.01]	[0.01]	[0.99]	[0.93]
Gender-Salient Round	0.46	0.29	−0.07***	−0.07***	−0.29	−0.33
	(0.93)	(0.92)	(0.03)	(0.03)	(0.68)	(0.70)
	{0.62}	{0.75}	{0.00}	{0.01}	{0.68}	{0.62}
	[0.59]	[0.73]	[0.00]	[0.01]	[0.64]	[0.59]
Matched with Female *Receiver*	−2.52**	−2.23*	0.03	0.03	−0.31	0.24
× Gender-Salient Round	(1.18)	(1.18)	(0.03)	(0.03)	(0.89)	(0.88)
	{0.04}	{0.06}	{0.36}	{0.38}	{0.73}	{0.77}
	[0.04]	[0.06]	[0.34]	[0.38]	[0.73]	[0.78]
Observations	944	942	692	690	944	942
R^2	0.07	0.11	0.08	0.10	0.09	0.13
Mean Dependent Variable	9.18	9.13	0.42	0.42	4.60	4.57
Demographic Covariates	No	Yes	No	Yes	No	Yes
Village & Interviewer Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimation. Panels A and B report the impact of being randomly matched with a female *receiver* in Round 1: Gender Anonymous and Round 2: Gender Salient, respectively. Panel C uses a pooled regression and reports the interaction between random match with a female *receiver* and an indicator for the gender-salient round. Columns (1) and (2) report the impact of treatment (= random match with female receiver) on trust, columns (3) and (4) on trustworthiness, and columns (5) and (6) on social preferences. Even-numbered columns include covariates: *risk preferences, household head, scheduled caste, scheduled tribe, other backward caste, age, married, literate, agriculture self-employed, agriculture wage labor, non-agriculture self-employed, non-agriculture wage labor, salaried, land ownership, generalized trust in strangers, and discriminatory attitudes.* One player is dropped from the sample when covariates are added due to non-response to the survey question on generalized trust in strangers. Columns (3) and (4) only includes observations where a positive amount was sent by the *senders* in the trust game. All columns include village fixed effects and interviewer fixed effects. Heteroskedasticity-robust standard errors are reported in parentheses. P-values from wild-bootstrap inference are reported in curly brackets. P-values from randomization inference test are reported in square brackets. *** significant at 1 percent, ** significant at 5 percent, * significant at 10 percent.

conventional levels (p=0.06) in column (2) after controlling for risk aversion and demographic covariates, however the magnitude remains stable (equivalent to 24% of the mean across both rounds). Columns (5) and (6) confirm that there is no impact of revealing gender on participants' social preferences. Finally, the results in columns (3) and (4) indicate that female *receivers* return significantly lower shares (less 13%) to the sender in the trust game. All *receivers* return a lower share on average when gender is made salient (less 7%), relative to the gender anonymous round. In Section 4.1 below, I further distinguish between the underlying mechanisms driving the observed treatment effects on trust.

Overall, these results indicate that participants behave differently towards women when they have a strategic role in the trust game, i.e. when they have to rely upon their co-operation in order to increase their pay-off. According to the experimental test proposed by Fershtman and Gneezy (2001), this allows me to conclude that participants are statistically discriminating against female co-participants. However, it is useful to note that the statistical discrimination appears rational since female *receivers* also return less to the *senders*.

3.2. Separate vs. Joint evaluation

In this section, I evaluate the impact of matched *receivers*' gender and trustworthiness on the senders' decision to accept or reject their randomly matched *receivers* in Phase II.[21,22]

[21] Table 9 in the Online Appendix reports the balance check between the separate and joint evaluation groups and confirm there are no statistically significant differences.

[22] The multi-phase design of this experiment makes it possible that *senders*' perception of this study's gender focus from Phase I influences their Phase II decisions. I mitigate this concern by: (1) re-randomizing treatments across

Table 4
Phase II: Separate evaluation group.

	Dependent variable: Matched *Receiver* is Selected			
	(1)	(2)	(3)	(4)
Match with Female Receiver	0.14**	0.12**	0.02	0.03
	(0.05)	(0.06)	(0.08)	(0.08)
	{0.01}	{0.04}	{0.83}	{0.73}
	[0.01]	[0.05]	[0.82]	[0.74]
Match with High Trustworthiness Receiver	0.10*	0.11*	−0.02	−0.01
	(0.06)	(0.06)	(0.08)	(0.08)
	{0.07}	{0.05}	{0.80}	{0.91}
	[0.06]	[0.06]	[0.78]	[0.92]
Match with Female Receiver x Female Sender			0.24**	0.19*
			(0.11)	(0.12)
			{0.02}	{0.09}
			[0.03]	[0.09]
Match with High Trustworthiness Receiver x Female Sender			0.26**	0.25**
			(0.11)	(0.11)
			{0.02}	{0.03}
			[0.02]	[0.03]
Female Sender			−0.19*	−0.15
			(0.11)	(0.12)
			{0.09}	{0.22}
			[0.06]	[0.18]
Observations	235	234	235	234
R^2	0.13	0.20	0.18	0.23
Matched Receivers Selected	0.77	0.77	0.77	0.77
P-value joint significance	0.01	0.03	0.00	0.01
P-value High Trustworthiness Match = Female Match	0.63	0.91		
Demographic Covariates	No	Yes	No	Yes
Village & Interviewer Fixed Effects	Yes	Yes	Yes	Yes

Notes: Linear probability model estimated using OLS. Dependent variable equals 1 if randomly matched *receiver* is selected (else 0). Same as before, covariates are included in even-numbered columns (2) and (4). One player is dropped from the sample when covariates are added due to incomplete data. Heteroskedasticity-robust standard errors are reported in parentheses. P-values from wild-bootstrap inference are reported in curly brackets. P-values from randomization inference test are reported in square brackets. All columns include village fixed effects and interviewer fixed effects. *** significant at 1 percent, ** significant at 5 percent, * significant at 10 percent.

First, I estimate the following linear probability model using OLS to test the impact of randomly matched *receiver's* salient characteristics under separate evaluation:

$$P_{ijv} = \beta_0 + \beta_1 F_{ijv} + \beta_2 T_{ijv} + \beta_3 X_{ijv} + \gamma_j + \delta_v + \epsilon_{ijv} \quad (2)$$

$$P_{ijv} = \beta_0 + \beta_1 F_{ijv} + \beta_2 S_{ijv} + \beta_3 F_{ijv} \times S_{ijv} + \beta_4 T_{ijv} + \beta_5 T_{ijv} \times S_{ijv} + \beta_3 X_{ijv} + \gamma_j + \delta_v + \epsilon_{ijv} \quad (3)$$

In Eq. (2), the binary dependent variable P_{ijv} equals 1 if *sender* i, interviewed by interviewer j in village v selects the randomly matched *receiver* (0 if rejects). The explanatory variables of interest are two binary indicators: F_{ijv} which equals 1 if the matched *receiver* is female (0 if male) and T_{ijv} which equals 1 if the matched *receiver* is of high-trustworthiness (0 if low-trustworthiness). X_{ijv}, γ_j, δ_v and ϵ_{ij} are defined similarly as in Eq. (1) previously. Huber-White heteroskedasticity-robust standard errors are reported alongside p-values from wild-bootstrap and randomization inference procedures. Eq. (3) is defined the same as above, but also includes new interaction terms between the explanatory variables of interest (F_{ijv} and T_{ijv}) and *sender's* gender S_{ijv} to evaluate heterogeneous treatment effects. Most covariates are binary variables, therefore OLS functions similarly to probit and logit in this context while imposing the least structure on the data.

Columns (1) and (2) in Table 4 reports estimation results for Eq. (2), while columns (3) and (4) report the results for Eq. (3) including interaction effects. Covariates including risk preferences and demographic characteristics are included in even-numbered columns (2) and (4), same as in previous estimations. The results in column (2) indicate that *senders* are on average 12% more likely to select the *receiver* if she is female, and 11% more likely to select a high-trustworthiness *receiver*. A Wald test confirms that the treatment indicators are jointly statistically significant. Moreover, the coefficient of the interaction term in column (4) shows that female senders are primarily driving this results since they are 19% more likely to select a female *receiver* and 25% more likely to select a high-trustworthiness *receiver*, compared to male *senders*. The direct coefficients for male *senders* are not statistically significant, indicating that men are not switching from their default choice.

Next, I estimate the following linear probability models using OLS to analyze impact of *receiver-pairs*' characteristics under joint evaluation:

$$P_{ijv} = \beta_0 + \beta_1 T_{ijv} + \beta_2 X_{ijv} + \gamma_j + \delta_v + \epsilon_{ijv} \quad (4)$$

$$P_{ijv} = \beta_0 + \beta_1 T_{ijv} + \beta_2 S_{ijv} + \beta_3 T_{ijv} \times S_{ijv} + \beta_4 X_{ijv} + \gamma_j + \delta_v + \epsilon_{ijv} \quad (5)$$

Eq. (4) represents a linear probability model where the dependent variable P_{ijv} corresponding to *sender* i, interviewed by interviewer j in village v can now take three values depending upon the *sender's* decision. Therefore, each outcome is coded as a binary variable in relation to the others, i.e. (1) $P_{ijv} = 1$ if male *receiver*, else 0, (2) $P_{ijv} = 1$ if female *receiver* is selected, else 0; or (3) $P_{ijv} = 1$ if anonymous *receiver* is selected, else 0.[23] A linear probability model can now be estimated separately for each outcome using OLS. This approach imposes least structure on the data and offers a straightforward interpretation, in comparison to the alternative modeling approach using multinomial logit or probit.

each round of lab games, (2) focusing the empirical hypotheses in Phase II on distinct outcomes, i.e. *senders*' selection decision.

[23] Since these binary variables are coded in relation to each other they sum up to 1 for each participant.

Table 5
Phase II: Joint evaluation group.

	Dependent variable: Matched *Receiver* is selected Female *Receiver* Selected		Male *Receiver* Selected		Anonymous *Receiver* Selected	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Impact of matched Receiver-pairs' characteristics on senders' selection decision						
Match with Female-High Trustworthiness *Receiver*	0.29***	0.29***	−0.26***	−0.26***	−0.03	−0.03
	(0.07)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)
	{0.00}	{0.00}	{0.00}	{0.00}	{0.50}	{0.47}
	[0.00]	[0.00]	[0.00]	[0.00]	[0.51]	[0.49]
Observations	237	237	237	237	237	237
R^2	0.12	0.23	0.13	0.22	0.09	0.15
Panel B: Treatment interaction with Senders' Gender						
Match with Female-High Trustworthiness *Receiver*	0.17*	0.18*	−0.20**	−0.19**	0.03	0.02
	(0.10)	(0.10)	(0.09)	(0.10)	(0.07)	(0.07)
	{0.08}	{0.08}	{0.04}	{0.05}	{0.70}	{0.80}
	[0.07]	[0.07]	[0.02]	[0.03]	[0.67]	[0.79]
Female *Sender*	0.12	0.13	−0.15	−0.16	0.03	0.03
	(0.10)	(0.10)	(0.10)	(0.10)	(0.06)	(0.07)
	{0.22}	{0.19}	{0.12}	{0.12}	{0.61}	{0.66}
	[0.19]	[0.22]	[0.06]	[0.09]	[0.62]	[0.67]
Match with Female-High Trustworthiness *Receiver* x Female *Sender*	0.21*	0.21	−0.10	−0.11	−0.11	−0.10
	(0.13)	(0.13)	(0.11)	(0.13)	(0.09)	(0.09)
	{0.09}	{0.11}	{0.38}	{0.36}	{0.22}	{0.31}
	[0.10]	[0.12]	[0.38]	[0.36]	[0.23]	[0.31]
Observations	237	237	237	237	237	237
R^2	0.18	0.24	0.18	0.22	0.10	0.15
Mean of Dependent Variable	0.56	0.56	0.31	0.31	0.13	0.13
Demographic Covariates	No	Yes	No	Yes	No	Yes
Village & Interviewer Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Linear probability models estimated using OLS. In columns (1) and (2), the dependent variable is a binary indicator equal to 1 if the female *receiver* is selected (else 0). In columns (3) and (4) the dependent variable is equal to 1 if male *receiver* is selected (else 0), and in columns (5) and (6), the dependent variable is equal to 1 if the anonymous *receiver* is selected (else 0). Same as before, covariates including *senders*' risk preferences and demographic characteristics are included in the even-numbered columns (2), (4), and (6). All estimations include village fixed effects and interviewer fixed effects. Heteroskedasticity-robust standard errors are reported in parentheses. P-values from wild-bootstrap inference are in curly brackets. P-values from randomization inference test are in square brackets. *** significant at 1 percent, ** significant at 5 percent, * significant at 10 percent.

The explanatory variable of interest is a binary treatment indicator T_{ijv} which equals 1 if the matched *receiver*-pair = (*High Trustworthiness-Female*, *Low Trustworthiness-Male*), or 0 if matched with the alternative (*High Trustworthiness Male*, *Low Trustworthiness Female*). X_{ijv}, γ_j, δ_v and ϵ_{ij} are defined similarly as in Eq. (1) previously. Huber-White heteroskedasticity-robust standard errors are reported alongside p-values from wild-bootstrap and randomization inference procedures. Eq. (5) is defined the same as above, but also includes new interaction terms between the explanatory variables of interest (T_{ijv}) and *sender's* gender S_{ijv} in order to evaluate heterogeneous effects.

Panels A and B in Table 5 reports results for Eqs. (4) and (5) respectively. In columns (1) and (2), the dependent variable is a binary variable equal to 1 if the female receiver is selected (equal to 0 if either female or anonymous receiver is selected). Similarly, in columns (3) and (4) the dependent variable is equal to 1 if male receiver is selected, and in columns (5) and (6), the dependent variable is equal to 1 if the anonymous receiver is selected. The estimates in panel A show that when *senders* are matched with an high-trustworthiness female *receiver*, she is approximately 29% more likely to be selected compared to the low-trustworthiness man and the anonymous option. The corresponding estimate for low-trustworthiness male receiver in column (4) indicates that he is 26% less likely to be selected.

The direct coefficients in panel B indicate male *senders'* being 18% more likely to select an high-trustworthiness female, compared to low-trustworthiness man and the anonymous option. Male *senders* are also 19% less likely to select a low trustworthiness man when there is a high-trustworthiness woman also available. Overall, these results show that participants select *receivers'* with high trustworthiness, irrespective of their gender under joint evaluation. This evidence supports the hypothesis that under joint evaluation, participants make analytic decisions in order to maximize their expected pay-off.

4. Robustness tests

In this section, I test the sensitivity of the main findings using alternate regression specifications and estimation methods.

4.1. Alternate specification: Control for social preferences

Results in Table 3 indicate that the Phase I treatment (= random match with female *receiver*) has a negative and statistically significant impact on trust when gender identity is made salient. Previous literature has shown that trust measured using lab games is driven by social preferences and risk aversion (Ashraf et al., 2006; Eckel & Wilson, 2004.)[24] Therefore, I re-estimate Eq. (1) after adding social preferences, i.e. amounts sent in the corresponding dictator games, into the vector of covariates. Specifically, I control for amount sent in the anonymous dictator game as a control when regressing amounts sent in the anonymous trust game on the treatment variable and covariates. Similarly, I control for amount sent in the gender salient dictator game when regressing amounts sent in the gender salient trust game on the treatment variable and covariates. This specification allows me to test

[24] Researchers have identified more distinct components of trust, including betrayal aversion (Bohnet & Zeckhauser, 2004) and inequity aversion (Ashraf et al., 2006). However, these additional drivers are beyond the scope of this study which does not include these measurements.

Table 6
Alternate specification: Impact of random match with female receiver.

	Dependent variable: Amount sent (Rs) in trust games					
	Gender anonymous		*Gender salient*		*Both rounds*	
	(1)	(2)	(3)	(4)	(5)	(6)
Matched with Female *Receiver*	0.73	0.38	−1.49**	−1.50**	3.04*	2.56
	(0.75)	(0.77)	(0.71)	(0.75)	(1.65)	(1.66)
	{0.35}	{0.64}	{0.04}	{0.05}	{0.08}	{0.13}
	[0.32]	[0.62]	[0.04]	[0.04]	[0.06]	[0.11]
Social Preferences	0.65***	0.63***	0.70***	0.69***	0.67***	0.65***
	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)
	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}	{0.00}
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Gender-Salient Round	–	–	–	–	0.66	0.50
					(0.82)	(0.82)
					{0.43}	{0.55}
					[0.35]	[0.49]
Matched with Female *Receiver* × Gender-Salient Round	–	–	–	–	−2.31**	−2.07**
					(1.03)	(1.03)
					{0.03}	{0.04}
					[0.03]	[0.04]
Observations	472	471	472	471	944	942
R^2	0.28	0.30	0.35	0.36	0.31	0.32
Mean Dependent Variable	9.58	9.54	8.78	8.72	9.18	9.13
Demographic Covariates	No	Yes	No	Yes	No	Yes
Village & Interviewer Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimations. This table reports the impact of being randomly matched with a female *receiver* on amounts sent in the trust games. Social preferences, i.e. amounts sent in the dictator game, is added as a covariate. Columns (1) and (2) report the results for gender anonymous setting and columns (3) and (4) for the gender salient setting. Column (5) and (6) use pooled data from both rounds and report the interaction effects between random match with female *receiver* and an indicator for the gender-salient round. Even-numbered columns include previously discussed covariates. One player is dropped from the full study sample when covariates are added due to non-response to the survey question on generalized trust in strangers. All columns include village fixed effects and interviewer fixed effects. Heteroskedasticity-robust standard errors are reported in parentheses. P-values from wild-bootstrap inference are reported in curly brackets. P-values from randomization inference tests are in square brackets. *** significant at 1 percent, ** significant at 5 percent, * significant at 10 percent.

whether the treatment effect on trust remains statistically significant after controlling for both: social preferences and risk aversion, measured using the risk lottery.

The results are reported in Table 6. Columns (1) and (2) report the results for gender anonymous setting and columns (3) and (4) for the gender salient setting. All columns include controls for social preferences, interviewer and village-session fixed effects. Even-numbered columns include risk aversion and previously discussed demographic covariates. The coefficients indicate that treatment effect on trust remains negative and statistically significant when gender is made salient, while the magnitudes are relatively smaller as expected due to the additional control variable. Social preferences are a robust predictor of amounts sent in the trust game under both anonymous and gender-salient settings, which is consistent with the previous literature. The coefficient magnitude indicates that, on average, 1 rupee sent in the dictator games is associated with an increase of 0.6–0.7 rupees sent in trust games. These results are robust to alternate inference methods, including heteroskedasticity-robust standard errors, wild-bootstrap inference and randomization inference.

4.2. Alternate estimator: Tobit model for censored variables

Next, it is useful to note that amounts sent in lab games, i.e. outcome variables in Eq. (1), are bounded between the range Rs 0–50. While this empirical design is standard in experimental research with lab games, it raises the concern that an underlying assumption for valid OLS estimation regarding normally distributed standard errors is violated. Therefore, I re-estimate Eq. (1) using Tobit regressions that account for censored variables by using maximum likelihood estimation. Since the outcomes of interest are units of money, I only account for censoring at the upper-limit of the distribution. Correcting for censoring at the lower-limit is not required since the variables cannot be negative and observed zero-valued observations represent real outcomes, not missing data.

The results are reported in Table 7. The coefficients represent the impact of being randomly matched with a female *receiver* under gender anonymous and gender-salient settings, respectively. Columns (1) and (3) report the impact on trust while columns (2) and (4) report the impact on social preferences. All columns include covariates, interviewer and village-session fixed effects. Same as previously, there is no statistically significant treatment effect when gender identities are anonymous in Panel A, as expected. In Panel B, when gender identities are made salient, the treatment coefficient on amounts sent in trust games remain similar in magnitude as before and statistically significant. The coefficients are also robust to alternate inference methods, including heteroskedasticity-robust standard errors and randomization inference. Overall, this indicates that the main findings in Table 3 remain comparable in magnitude and statistical significance when correcting for censored outcomes.

5. Concluding remarks and implications for policy

Discriminatory norms are one of the most important drivers of unequal socio-economic participation by women in India. Average female labor-force participation has persisted at a level below half the same rate for men (below 30% for women vs. above 70% for men). Growth in female labor-force participation rates has stalled in urban regions and declined in rural areas since 1999–2000 leading to 35 million 'missing' women in the labor-force (World Bank, 2014). Gender prejudice is also evidenced in the preference for male children. Sex ratio decreased from 927 girls per 1000 boys to 914 girls per 1000 boys under age 7 between the last two rounds of the decennial census of India. Infant mortality rates for girls are significantly higher across all major states.[25]

This paper contributes new evidence on the behavioral patterns of gender discrimination from rural India. Economic theory shows

[25] According to data from the National Sample Survey Organization and Census of India, 2011.

Table 7
Alternative estimation method: Tobit regressions.

	Dependent variables: Trust games Amount sent (Rs)		Dictator games Amount sent (Rs)	
	(1)	(2)	(3)	(4)
Panel A: Gender Anonymous Round				
Matched with Female *Receiver*	0.50	0.08	−0.34	−0.45
	(0.82)	(0.84)	(0.60)	(0.62)
	[0.54]	[0.92]	[0.58]	[0.49]
Observations	472	471	472	471
Pseudo R^2	0.01	0.02	0.02	0.02
Mean Dependent Variable	9.58	9.54	4.82	4.79
Panel B: Gender Salient Round				
Matched with Female *Receiver*	−1.91**	−1.92**	−0.60.	−0.60
	(0.82)	(0.84)	(0.62)	(0.62)
	[0.02]	[0.02]	[0.35]	[0.35]
Observations	472	471	472	471
Pseudo R^2	0.01	0.02	0.02	0.02
Mean Dependent Variable	8.78	8.72	4.38	4.34
Demographic Covariates	No	Yes	No	Yes
Village & Interviewer Fixed Effects	Yes	Yes	Yes	Yes

Notes: Tobit estimations for censored outcome variables. Panels A and B report the impact of being randomly matched with a female *receiver* in Round 1: Gender Anonymous and Round 2: Gender Salient, respectively. Columns (1) and (2) report the impact of treatment (= random match with female receiver) on trust, columns (3) and (4) on social preferences. Even-numbered columns include all previously discussed covariates. One player is dropped from the sample when covariates are added due to non-response to the survey question on generalized trust in strangers. All columns include village fixed effects and interviewer fixed effects. Heteroskedasticity-robust standard errors are reported in parentheses. P-values from randomization inference test are reported in square brackets. *** significant at 1 percent, ** significant at 5 percent, * significant at 10 percent.

that unequal treatment can be characterized either by taste-based discrimination driven by prejudice or statistical discrimination driven by negative stereotypes. I conduct a simple experimental test using lab measurements of trust and social preferences to distinguish between these competing explanations. My findings indicate that, on average, participants in this study statistically discriminate against women in economic transactions involving trust. Furthermore, it is interesting to note that observed statistical discrimination is indeed rational because female participants also demonstrate lower trustworthiness, on average. It is beyond the scope of this study to reliably determine why both genders differ in terms of trustworthiness, especially since this phenomenon is possibly driven by historically regressive norms that have become internalized.

Research on ameliorating gender discrimination has considered exploiting the existence of dual cognitive modes among humans: System 1 cognition that explains intuitive decision-making, and System 2 cognitive that explains analytical decision-making based. I apply this framework to test an evaluation nudge hypothesized to trigger System 2 cognition through the comparative analysis of multiple expected payoffs. My findings indicate that, on average, participants prefer to select partners based on their trustworthiness, irrespective of gender, when System 2 cognition is triggered.

In the real world, these behavioral parameters are hard to observe for policymakers who instead rely on observable outcomes of discrimination to guide policies to promote gender equality. Efforts have broadly focused on two types of interventions: quotas and education. Quotas are effective in promoting equal representation and alleviating discriminatory norms through prolonged exposure, however they can also result in unintended consequences. In certain contexts, quotas that restrict people's choices and violate strong local norms can lead to resentment and even backlash (Goldin, 2014; Thernstrom & Thernstrom, 2009). On the other hand, educational interventions such as television programs with progressive gender roles and school-based gender sensitization programs can improve negative gender stereotypes (Dhar et al., 2017). However, groups characterized by strong prejudices or falling outside the coverage of educational interventions can prove less malleable (Banerjee et al., 2013; Beaman et al., 2009).

In conclusion, I argue that identifying the dominant mechanisms driving observed gender inequality in specific contexts can help in the design and selection of behaviorally-informed interventions to promote equality. Groups where taste-based discrimination dominates may be more suited for quotas to ensure equal representation while slow-moving norms evolve. Conversely, education is more relevant where negative stereotypes dominate. Decision-making by time-stressed individuals in organizational settings are prone to personal biases. Therefore mandates for joint evaluation of common, performance-related criteria can promote fair comparisons across multiple candidates. The objective of this study is to provide robust evidence of gender discrimination from a representative study sample in a densely populated region characterized by regressive gender inequality and poverty. Similar evidence can be measured more extensively using periodically conducted large-scale demographic surveys and future academic studies.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.socec.2026.102556.

Data availability

Replication package is available here: https://zenodo.org/records/17109143.

References

Aigner, D. J., & Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*, *30*(2), 175–187.
Anderson, S., & Ray, D. (2010). Missing women: Age and disease. *Review of Economic Studies*, *77*(4), 1262–1300.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Anukriti, S., Kwon, S., & Prakash, N. (2022). Saving for dowry: Evidence from rural India. *Journal of Development Economics*, *154*, Article 102750.

Arrow, K. (1971). The theory of discrimination. Working Papers 403, Princeton University, Department of Economics, Industrial Relations Section.

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*(3), 193–208.

Banerjee, A., Duflo, E., Imbert, C., & Pande, R. (2013). Entry, exit, and candidate selection: Evidence from India. Mimeo, 3ie Grantee Final Report.

Barron, K., Ditlmann, R., Gehrig, S., & Schweighofer-Kodritsch, S. (2024). Explicit and implicit belief-based gender discrimination: A hiring experiment. *Management Science*.

Bazerman, M., & Moore, D. (2013). *Judgment in managerial decision making 8th ed.*. John Wiley & Sons, Hoboken, NJ.

Bazerman, M. H., Tenbrunsel, A. E., & Wade-Benzoni, K. (1998). Negotiating with yourself and losing: Making decisions with competing internal preferences. *Academy of Management Review*, *23*(2), 225–241.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: Does exposure reduce bias? *The Quarterly Journal of Economics*, *124*(4), 1497–1540.

Becker, G. S. (1971). *University of chicago press economics books*, *The economics of discrimination*. (9780226041162), University of Chicago Press.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.

Bertrand, M., & Duflo, E. (2017). Field experiments on discrimination. *1*, In *Handbook of economic field experiments* (pp. 309–393). Elsevier.

Bicchieri, C., & Dimant, E. (2019). Nudging with care: the risks and benefits of social information. *Public Choice*.

Bohnet, I., van Geen, A., & Bazerman, M. (2016). When performance trumps gender bias: Joint vs. Separate evaluation. *Management Science*, *62*(5), 1225–1234.

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior and Organization*, *55*(4), 467–484.

Bohren, J. A., Haggag, K., Imas, A., & Pope, D. G. (2023). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics*, 1–45.

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., & Eccles, J. (2006). Empathy or antipathy? The impact of diversity. *The American Economic Review*, *96*(5), 1890–1905.

Bolton, G. E., & Ockenfels, A. (1998). Strategy and equity: An ERC-analysis of the güth–van damme game. *Journal of Mathematical Psychology*, *42*(2), 215–226.

Casari, M., & Cason, T. N. (2009). The strategy method lowers measured trustworthy behavior. *Economics Letters*, *103*(3), 157–159.

Castillo, M., & Petrie, R. (2010). Discrimination in the lab: Does information trump appearance? *Games and Economic Behavior*, *68*(1), 50–59.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 817–869.

Charness, G., Samek, A., & van de Ven, J. (2022). What is considered deception in experimental economics? *Experimental Economics*, *25*, 1573–6938.

Chattopadhyay, R., & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica*, *72*(5), 1409–1443.

Chaudhuri, A., Graziano, S., & Maitra, P. (2006). Social learning and norms in a public goods experiment with inter-generational advice. *Review of Economic Studies*, *73*(2), 357–380.

Delavande, A., & Zafar, B. (2019). Gender discrimination and social identity: Evidence from urban Pakistan. *Economic Development and Cultural Change*, *68*(1), 1–40.

Dhar, D., Jain, T., & Jayachandran, S. (2017). Evaluation of a school-based gender sensitization campaign in India. AEA RCT Registry. March 18..

Eckel, C. C., & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior*, *16*, 181.

Eckel, C. C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior and Organization*, *55*(4), 447–465.

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association*, *7*(2–3), 235–266.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 817–868.

Fershtman, C., & Gneezy, U. (2001). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics*, *116*(1), 351–377.

Fershtman, C., Gneezy, U., & Verboven, F. (2005). Discrimination and nepotism: The efficiency of the anonymity rule. *The Journal of Legal Studies*, *34*(2), 371–396.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*(3), 347–369.

Goldin, C. (2014). A pollution theory of discrimination: Male and female differences in occupations and earnings. In *Human capital in history: the American record* (pp. 313–348). University of Chicago Press.

Grossman, P. J., Eckel, C., Komai, M., & Zhan, W. (2019). It pays to be a man: Rewards for leaders in a coordination game. *Journal of Economic Behavior and Organization*, *161*, 197–215.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics*, *7*(1), 63–88.

Jensen, R., & Oster, E. (2009). The power of TV: Cable television and women's status in India. *The Quarterly Journal of Economics*, *124*(3), 1057–1094.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136.

Krawczyk, M. (2019). What should be regarded as deception in experimental economics? evidence from a survey of researchers and subjects. *Journal of Behavioral and Experimental Economics*, *79*, 110–118.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, *62*(4), 659–661.

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology*, *87*(2), 157.

Sen, A. (1990). More than 100 million women are missing. *The New York Review of Books*.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*(1), 80–83.

Sukhtankar, S., Kruks-Wisner, G., & Mangla, A. (2022). Policing in patriarchy: An experimental evaluation of reforms to improve police responsiveness to women in India. *Science*, *377*(6602), 191–198.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, New Haven, CT.

Thernstrom, S., & Thernstrom, A. (2009). *America in black and white: One nation, indivisible*. Simon and Schuster.

World Bank (2014). *India–Women, Work and Employment*. Washington, DC ; World Bank Group, http://documents.worldbank.org/curated/en/753861468044063804/India–Women–work–and–employment.

World Bank (2016). *India States Briefs - Chhattisgarh*. Washington, DC; World Bank Group.