THE GRADUATE INSTITUTE GENEVA

INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

Graduate Institute of International and Development Studies
International Economics Department
Working Paper Series

Working Paper No. HEIDWP19-2024

Perceived abilities and gender stereotypes within the household: experimental evidence from Bangladesh

Carlotta Nani
Geneva Graduate Institute

Chemin Eugène-Rigot 2
P.O. Box 136
CH - 1211 Geneva 21
Switzerland

Perceived abilities and gender stereotypes within the household. Experimental evidence from Bangladesh*

Carlotta Nani
Geneva Graduate Institute

September 2024

Abstract

Is it possible to improve women's agency by providing information about their abilities? Using a lab experiment in the field, I study how perceived abilities and gender stereotypes shape intra-household dynamics. I use an incentivized decision-making game with 525 married couples from 42 rural villages in Bangladesh to investigate whether women are discriminated against because they are perceived to be less skilled than their husband, and whether it is possible to reduce this gender bias within households. During the game, I provide information on women's abilities and I observe how beliefs and decisions change. The empirical analysis shows that the less capable women are perceived compared to men, the less they are involved in decision-making. After the information treatment, husbands with the lowest regard for their wife's skills are 20 percent more likely to make allocations in her favour. The treatment has a larger impact on younger couples, on men with stronger control preferences and on risk-averse women. This brings further evidence of the inability of spouses to observe each other's skills. Two weeks after the experiment, women in treated couples report being more involved in household decisions. These results suggest that gender discrimination within households has a statistical component that can be corrected by increasing skills' observability.

Keywords: Bangladesh, field experiment, gender discrimination, intra-household dynamics
JEL Codes: D82, D83, D91, J12, J16

*Please do not cite or circulate without the author's permission.

I thank Ashraf Mian for his incredible work as project manager; Mamun Rashid, Anindya Barai and the DRI Bangladesh team for their professionalism during the project implementation, Munshi Sulaiman for his recommendations and Sara Moioli for her help during the pilot. I thank the Swiss National Science Foundation for funding this project; Patricia Arnold, Dominic Eggel and Henri Wacongne for their administrative support before and during the fieldwork activities. I am grateful for their helpful comments to Ingvild Almås, Marianna Battaglia, Tessa Bold, Martina Björkman Nyqvist, Leonardo Bursztyn, Paula Bustos, Julia Cajal-Grossi, Lucia Corno, Pierre Dubuis, Didi Egerton-Warburton, Raquel Fernandez, Catalina Franco Buitrago, Xavier Giné, Esther Gehrke, Jessica Goldberg, Paula Gonzalez-Martinez, Marcela Ibanez Diaz, Cynthia Kinnan, Andreas Madestam, Jonathan Morduch, Arne Nasgowitz, Ugo Panizza, Sahar Parsa, Natalia Rigol, Charlotte Ringdal, Anusuya Sivaram, Vincent Someville, Miri Strjyan, Nathan Sussman, Bertil Tungodden, Christopher Woodruff, the participants of the 2023 Nordic Conference in Development Economics, of the 2023 German Development Economics Conference, of the 21st EUDN PhD Workshop, of the 8th Swiss Young Researchers Workshop in Behavioral Economics and Experimental Research and of the Rare Voices in Economics research clusters. I am extremely grateful to my advisor, Lore Vandewalle, for her guidance and support throughout the project.

1 Introduction

According to the 2022 Global Gender Gap Report, it will take 132 years to reach full gender parity (World-Economic-Forum, 2022). Closing gender gaps in health, education and bargaining power could increase GDP levels, since unequal access to opportunities on the basis of gender means that human capital is being misallocated. This is particularly relevant for low-income countries where women face more and greater barriers that exacerbate disparities (Jayachandran, 2015). Despite interventions to boost women's bargaining power, restrictive gender norms that shape intra-household dynamics remain a major obstacle to women's independence (Afzal et al., 2022, Bursztyn et al., 2020, Chang et al., 2020). When the bargaining power is concentrated in the hands of one individual, their beliefs and preferences shape households' decisions. Because social norms are differently perceived and internalized by male and female household members, appointing the main decision-maker on the basis of gender, rather than actual abilities, can lead to sub-optimal outcomes (Afzal et al., 2022, Buchmann et al., 2023, Cassidy et al., 2022).

While there is consensus among policymakers that empowering women should be a top priority, there is limited evidence in the literature of what strategies work, and why gender gaps persist (Chang et al., 2020). In this paper, I take a step back and investigate the roots of the gender discrimination observed in intra-household dynamics. My main hypothesis is that the lack of information on women's abilities pushes individuals to form opinions through gender stereotypes that represent women as less capable. To test this hypothesis, I conduct a lab experiment with married couples in Bangladesh. First, I focus on whether women are discriminated against and have a low willingness to be involved in decision-making because they are perceived (and perceive themselves) as less capable than men. Then, I study whether perceptions are based on gender stereotypes that arise from information gaps about women's skills. Finally, I test whether filling these gaps alleviates stereotypes and increases women's involvement in decision-making.

The data collected from 525 married couples across 42 rural villages show that women are perceived to be less capable than men in problem-solving and money management. This *perceived skills gender gap* is correlated with women's involvement in decision-making: the wider the gap (the less skilled a woman is perceived compared to her husband), the less involved she is in household decisions. I design an incentivized decision-making game to investigate whether this gender discrimination is statistical. Namely, whether it arises from a lack of information. The game consists of three rounds played simultaneously by spouses. Each spouse answers eight Raven coloured progressive matrices (RCPM) and earns points for each correct answer. The final earnings of the couple depend on both spouses' performances. The main feature of the game is that one-third of the matrices are worth more points than the remaining two-thirds. Individuals decide privately how many high and low-payoff matrices they and their spouse will answer. Throughout the game, I elicit individuals' beliefs about their own and their spouse's performance. Couples are randomly assigned to a treatment and a control arm: couples in the treatment arm are informed about the

wife's cognitive skills; couples in the control arm receive no update. The outcomes of interest are the number of high-payoffs assigned to women, and the size of the difference between the guessed number of correct answers given by the husband and by the wife (the belief gender gap).

After the information treatment, individuals who initially perceived the husband as more skilled increase the number of high-payoffs allocated to the wife. The same occurs in more recent couples, who have been exposed to each other's abilities for a shorter period of time. Perceptions of women's skills in housework and childcare, which are more observable within the household, do not affect the game outcomes. These results further support the hypothesis that unobservable abilities lead to biased perceptions. Social and family pressure, risk aversion and control preferences of men are usually identified as the main drivers of household decisions, and the reason why norms are hard to change (Ashraf, 2009, Bursztyn et al., 2020, Dhar et al., 2022). In my study, the information provision has a larger impact on men with stronger control preferences and on risk-averse women. The treatment has no impact based on gender attitudes (measured using self-reported measures and a non-verbal Implicit Association Test) and social pressure. When looking at real-life outcomes, two weeks after the lab experiment, women in the treated group report being more involved in household decisions.

This study contributes to the literature on gender discrimination in three important and innovative ways. First, I explore the origins of gender discrimination in household decision-making. Especially in environments with patriarchal norms, there is a negative correlation between men's perception of norms and women's agency. Women seem to be able to work only if it is necessary to sustain the family (Bernhardt et al., 2018). Resources are diverted towards the activity of male household members, and even when women retain some control over productive assets, the revenues from these activities are managed by men (Bernhardt et al., 2019, De Mel et al., 2009a, Fafchamps et al., 2014, Roy et al., 2015). In some cases, men choose less optimal investments to limit their wife's access to resources (Mani, 2020). When they are excluded from decision-making, women opt for costly and usafe financial instruments to conceal earnings and keep control over resources (Almås et al., 2018, Castilla, 2015, Jakiela and Ozier, 2016, Riley, 2022, Schaner, 2015, 2017). I analyse women's low involvement in decision-making as a result of information gaps: when skills are harder to observe, individuals turn to gender stereotypes that depict women as unfit to work or manage resources, reinforcing gender norms and exacerbating discrimination. Previous research has highlighted the importance of perceptions about abilities in intra-household dynamics. To my knowledge, I am the first to investigate explicitly whether gender discrimination within households is statistical and arises from stereotypes and perceived abilities.

Second, I focus on women's willingness to be involved. I study whether, given the opportunity, women decide to join the decision-making process or prefer to step back. McKelway (2021) shows that giving women more access to resources does not translate into higher agency if they perceive

themselves to be less capable of managing them. According to Jayachandran (2021), women's lack of willingness to pay to conceal earnings from their husbands comes from the belief that women should not be involved in household finances. Abbink et al. (2020) investigate gender bias in intra-household decision-making using a lab experiment with married couples in Bangladesh. Players choose between making a decision involving risky choices or transferring the decision-making to their spouse. Women are more likely to pass the decision to their husbands because of low bargaining power and low perceived financial capabilities. Buchmann et al. (2023) use a dictator game and a market experiment to study how reputation affects household investment decisions in Malawi. By making salient episodes where the wife proved to lack expertise, the authors show that men's decisions are affected by perceptions about their wife's abilities. In the market experiment, less expert women under-invest in new technologies and are willing to pay to hide their mistakes and preserve their reputation within the household.

While the literature has extensively investigated the impact of social norms on women's empowerment, it remains a key question whether exposure to women's abilities can relax norms and change behaviors and beliefs. Field et al. (2021), evaluate the impact of women's wages being disbursed digitally in rural India. Labor force participation and financial autonomy increased for women who received digital payments. In addition, the authors observe changes in perceptions and gender attitudes: treated women are more likely to value women's work and less likely to judge households where women work. Their husbands become less likely to report reputation costs for men whose wives work. Beaman et al. (2009) study whether exposure to female leaders changes public opinion on women. Although voters' opinions are harder to shift and villagers still prefer male leaders, they observe a reduction in gender stereotypes and an increase in women's chances of running for leadership. Bursztyn et al. (2020) highlight the links between misperceptions, social norms and women's labor force participation in Saudi Arabia. Even if men privately agree that women should be allowed to work outside of the house, they forbid it because they believe it is socially unacceptable. Revealing that their peers share their private views translates into an increased likelihood of their wives looking for jobs four months after the intervention.

Discrimination can be categorized as taste-based or statistical. Statistical discrimination occurs when a group is treated differently because of a signalling issue: if individuals have imperfect information on subjects, they use stereotypes and beliefs to form opinions. Taste-based discrimination depends on an intrinsic dislike for a group and therefore is harder to eradicate. Exposure to the real abilities of the stereotyped categories and providing information about unconscious bias can change attitudes and behaviors (Alesina et al., 2018, Beaman et al., 2009, Bertrand and Duflo, 2017, Castillo and Petrie, 2010). However, there are situations where despite the information provision, attitudes do not change. Montoya et al. (2020) focus on gender discrimination in loan applications. The authors inform loan officers about women's higher repayment rates and show that after receiving this information, taste-based officers discriminate even more against women. Wozniak and

MacNeill (2020) set up an experiment where, despite making it costly for individuals to discriminate against a group, participants do not update their beliefs and discrimination persists. While studies on statistical and taste-based discrimination usually involve individuals who do not have intimate relationships (employers-employees, university peers, teachers-students, etc.), I apply this methodology to real-life partners living under the same roof.

In terms of policy contribution, my results suggest that women's exclusion from household decision-making has a statistical component that can be alleviated with interventions to make women's skills more observable.

The remainder of this paper is organized as follows. In Section 2, I describe the setting where the study takes place, the recruitment procedure, the experimental design and the main hypotheses. In Section 3 I present the data collected and the sample of respondents. In Section 4, I provide details on the empirical strategy used in the analysis, the main results and the alternative mechanisms. In Section 5, I present my conclusions.

2 Experimental Design

2.1 Country Context

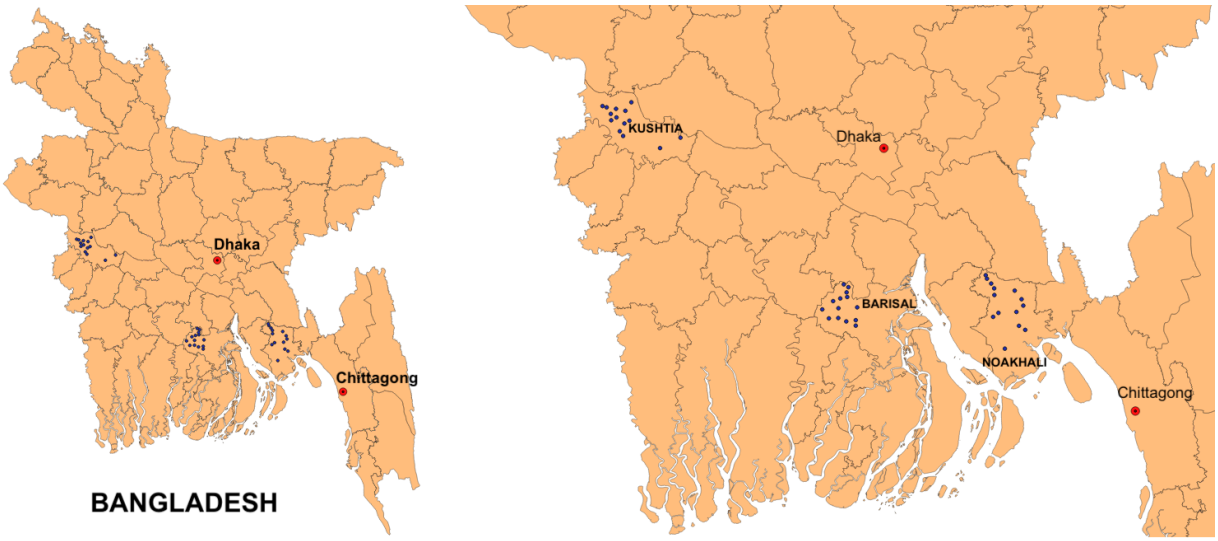
With a GDP per capita of 6,493 USD, a rural population above 60 percent and a society with a strong preference for males, Bangladesh represents an ideal setting to study how gender stereotypes affect intra-household dynamics. Despite substantial progress in narrowing the gender gap in the past two decades, women in Bangladesh continue to experience limited agency (World-Economic-Forum, 2022, Solotaroff et al., 2019). Women are usually married before turning 18 and are more likely to discontinue their education after marriage while being excluded from income-generating activities. Especially in rural areas, restrictive patriarchal norms pose significant barriers to women's labor force participation and access to resources. According to the latest World Values Survey, the majority of the population agrees with the notion that women who earn more than men create problems and that men should have more rights to a job than women. In the most recent DHS report, only one-third of women with cash earnings make independent decisions about their earnings (WVS, 2020, DHS, 2020). Additionally, women comprise the majority of the unbanked population: although mobile financial services are spreading quickly throughout the country, the gender gap in mobile ownership amounts to 23 percentage points (Demirgüç-Kunt et al., 2022, Breza et al., 2020, Lee et al., 2021, Shanahan, 2022). According to BRAC, the largest NGO in the country, for the majority of the women who received digital COVID-19 relief transfers it is their husbands or sons who cashed out the money (Azad et al., 2020).
Women seem to be more able to manage resources if they are devoted to businesses in women-dominated industries or to activities that can be carried out at home (such as poultry rearing or paddy husking). As they move to less traditional or male-dominated sectors, husbands and other

male relatives take control (Goetz and Gupta, 1996, Hashemi et al., 1996, DHS, 2020).

2.2 Sample selection

The experiment took place between July and August 2022 in 42 rural villages across three districts of Bangladesh: Kushtia, Noakhali and Barisal (the village locations are shown in Figure 1). The final sample consists of 525 married couples.

Figure 1: District level map of the villages under study.



Timeline

Participants were recruited in two steps. First, trained enumerators went door-to-door and randomly invited one woman per household to take part in a short survey. To be eligible for the study, a woman had to fulfil four criteria: (i) being aged 18 to 45; (ii) being married and currently residing with her husband; (iii) having no cognitive disability (iv) having a co-residing husband with no cognitive disability.

Three to four weeks after the baseline survey, the field team invited the selected women to participate along with their husbands, in the lab experiment.On the day of the experiment, two enumerators of the opposite gender visited the couples at their home. Both spouses had to be present and independently consent to participate.

Two weeks after the game, the survey team conducted a short phone follow-up survey with the women to collect information post-treatment.

Selection procedure

The last section of the baseline survey included 8 coloured progressive matrices from the Raven test, a standardized psychological test that measures cognitive abilities. Each matrix of the test is composed of a picture with a geometric pattern where one piece is cut out. Respondents chose

from among six options which one completes the pattern of the picture. Only one answer is correct and the matrices become progressively harder (Domino and Domino, 2006).[1] I used the number of correctly answered matrices to select the couples for the lab experiment. On average women answered four Raven matrices correctly. I invited the couples in which the wife's number of correct answers was above average to participate in the game to ensure a selection of women who possess greater cognitive skills. The key feature of the game is that it leverages the wife's cognitive abilities. While some individuals might already be aware of the true skill level of the women, some did not have this information yet. The treatment aimed to fill the information gap whenever women's skills are not observable. In addition, restricting the treatment to the provision of positive feedback avoids the potential for retaliation after the experiment.

Table A1 provides the balance table of demographic characteristics of women who scored just above and just below the district average. The two groups are balanced for most of the characteristics and the difference in means between selected and not-selected women is statistically significant (at the 5 percent level) only for those characteristics that are correlated with cognitive skills: selected women tend to be younger and have a higher level of education. Nevertheless, the two groups are equally involved in household decision-making, they are equally likely to be homemakers and spend the same amount of time on childcare and housework.

2.3 Incentivized decision-making (IDM) game

The game took place at the couple's home three to four weeks after the baseline survey. Spouses played simultaneously: husband and wife each sat in a separate room with an enumerator of the same gender.[2] To ensure the game proceeded simultaneously, the enumerators would exchange via SMS information on allocation decisions, points earned and whether they could move on to the next step.[3]

The game design is illustrated in Figure 2. It is composed of three rounds. The first round is for practice and each spouse answers eight Raven matrices. Matrices have no payoffs and players do not make any allocation decisions. This allows me to rule out that beliefs and behaviors in subsequent rounds are influenced by not knowing the tasks.

In the second and third rounds, each spouse divided the high (20 points) and low-payoff matrices (10 points) between themselves and their partner. From a pool of 16 payoffs (5 high and 11 low), they would choose 8 for themselves and eight for their spouse. They had to allocate exactly eight payoffs; each payoff they kept for themselves could not be given to their spouse and vice versa.[4]

[1] Figure A7 in the Appendix shows an example of an easy matrix.

[2] If the house was composed only of one room, the husband would sit outside while the wife would sit inside.

[3] Enumerators were instructed not to read aloud any of the messages they received or sent. Each SMS was pre-filled by the survey platform. Enumerators only had to hit the *SEND* button on their tablets. Enumerators were instructed to follow an *emergency* protocol in case of a poor network that could obstruct the SMS exchange. Regardless, no such issues were experienced during the lab experiment.

[4] Players made allocation decisions by placing red and yellow tokens on a laminated board. Yellow tokens represented high-payoffs and red tokens represented low-payoffs. Each token had written on both sides either 20 or 10

Spouses were informed that in each round, either the husband's or the wife's allocation would be randomly implemented, but they would not know whose. They were also informed that payoffs appeared in a random order, meaning they did not know if a question was worth 20 or 10 points when they answered it. This setup was designed to incentivize players to put the same effort into each answer.

Once allocation decisions were made, players answered the 8 Raven matrices. At the end of this task, they indicated how many matrices they thought they answered correctly and how many they thought their spouse answered correctly. For each correct guess, they earned 2 additional points. Between the second and last rounds, couples in the treatment arm received an information update. The enumerators read separately to each spouse the following script:

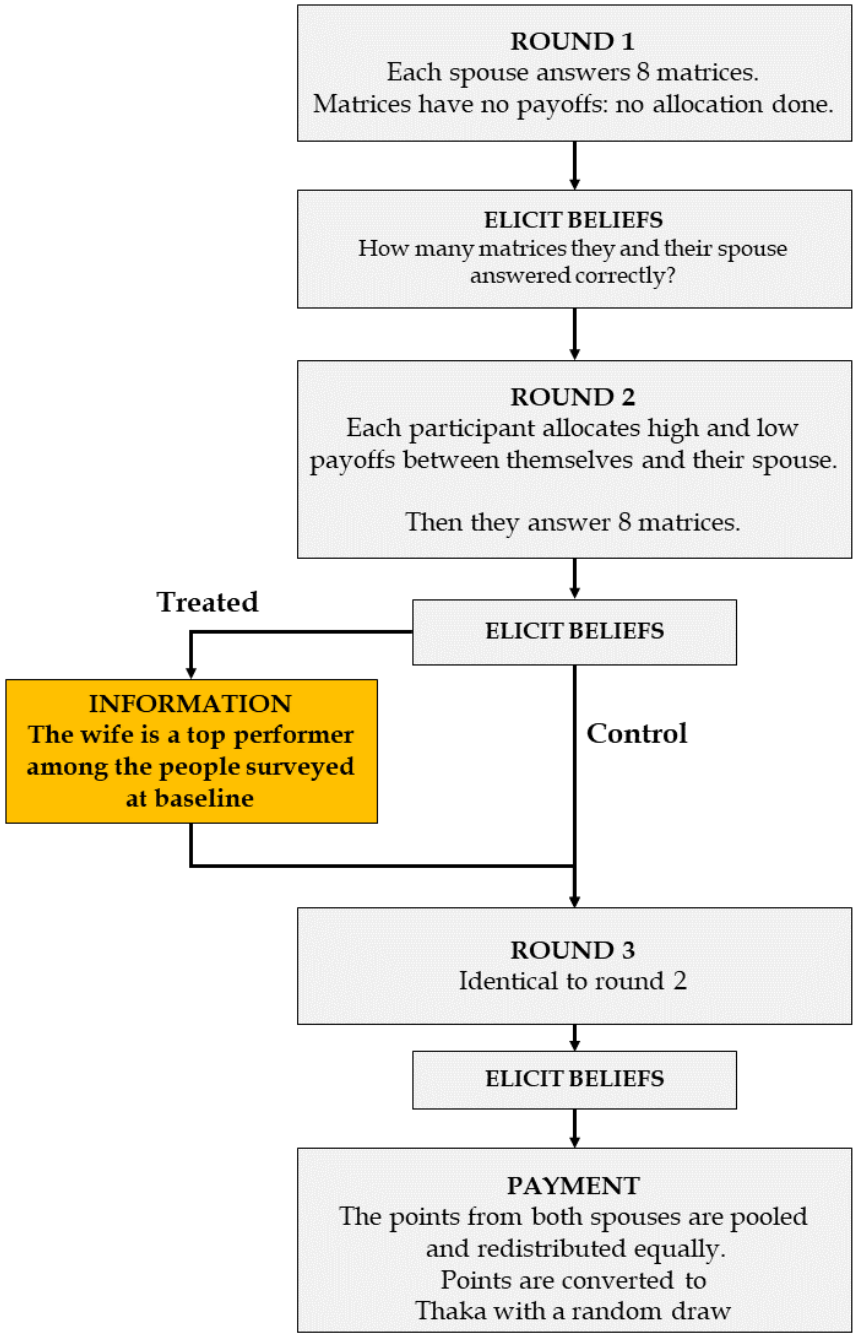
> *"I would like to give you some information on [YOUR/YOUR WIFE'S] abilities in this activity. Some days ago we conducted a survey with people in this area. During that day, we asked your wife to answer a set of questions like the ones you are doing today. We counted how many questions she answered correctly and we compared it with the number of correct answers of other 1000 people that we surveyed on that day. With the number of correct answers she gave, she is part of the best performers among the people who answered the same questions: this means she answered correctly to a number of questions that is higher than what the majority of the people did."*

Couples in the control arm received no information: the enumerators did not read any statement and they proceeded directly to the third, and final, round.

To preserve privacy, the points earned by the spouses were pooled together and distributed equally between them. Players could not infer how many points they or their spouse scored on their own. They were only informed of the total number of points they received after the redistribution. To allow earnings to be concealed, points were converted into Bangladeshi Thaka (BDT) with a random draw. Each player could randomly draw 0.5 with a 20 percent probability; 1 with a probability of 60 percent and 2 with a probability of 20 percent. If they drew 0.5, the amount of points was halved. If they drew 1 it remained the same, if they drew 2 it doubled. To incentivize allocations that maximize the household income, spouses were informed at the beginning of the game about the payment procedure. The maximum amount a couple could earn was 936 BDT (9 USD), which is roughly twice the daily income of a rural household (HIES, 2019). On top of the amount earned, each player received a 50 BDT show-up fee.

(Figure A9 shows the allocation board and tokens).

Figure 2: Incentivized decision-making game design



The experiment was designed to assess whether and how gender stereotypes shape decisions. Specifically, whether there is a *statistical bias* that depends on information gaps about women's skills. The underlying assumption is that allocation decisions and beliefs in the game are shaped by how spouses perceive their own abilities relative to each other. When abilities are not observable, individuals may rely on gender stereotypes to fill those gaps (Bertrand and Duflo, 2017). In this

context, irrespective of their actual cognitive skill level, women may be perceived as less capable than men and receive fewer high-payoffs. If the bias is statistical, filling those gaps should change beliefs and allocation.

After the information treatment, some individuals might assume that if the wife is "good", the husband must be "better". In this case, the treatment could have no effect or make allocations even more biased towards men: this would imply that the gender bias is *taste-based* and is due to an intrinsic disdain for women that does not disappear despite the information provision.

3 Data and Baseline Characteristics

3.1 Data

To test my hypothesis, I used primary data collected through in-person interviews. The baseline door-to-door survey included demographic characteristics of women, household characteristics, involvement in decision-making, control of resources, numeracy skills and financial literacy. The last section of the survey consisted of 8 matrices of the Raven test used to select participants for the game. The second source of information is the survey answered separately by the husband and the wife before the game.[5] It gathered information on gender attitudes, social pressure and norms. In addition, respondents were asked to rate their and their spouse's skills in problem-solving, money management, housework and childcare. Finally, the main outcomes of interest came from the data collected during the game: how many high-payoffs were assigned to women and how many matrices players thought they and their spouses answered correctly.

3.1.1 Randomization and Balance

The randomization is performed at the couple level. After the baseline survey the selected couples were randomly assigned to treatment and control. To increase power and ensure the two groups were balanced on key variables that might be correlated with outcomes, I stratified the randomization on the village of residence (couples from the same village might share similar norms and values, and be subject to the same social pressure), the wife's completed education (education is correlated with cognitive abilities) and marriage length (couples who have been together longer should be more capable of observing each others' abilities). The final sample is composed of 281 control and 244 treated couples.

Table 1 presents the balance table for baseline characteristics. Overall the two groups are balanced, indicating that the randomization was performed correctly, and the treatment is orthogonal to couples' characteristics.[6] The F-test for joint significance rejects the null hypothesis that baseline characteristics predict treatment status.

[5] Treated and control couples answered the same survey. Therefore, the risk of priming effects affecting couples' outcomes differently based on treatment status should be mitigated.

[6] Only the difference in mean for household size and husband's education is statistically significant.

On average, couples in the sample live in households composed of 5 members. Most of these households (80 percent) have a monthly income below 20,000 BDT (roughly 190 USD), which is in line with the figures reported by the 2016 *Household Income and Expenditure Surveys* (HIES, 2019). Over the previous 12 months, 40 percent of households had gone without eating at least once.
On average, the age gap between spouses is 8 years (women are 32 years old, men 40) and they have been married for almost 16 years. Women are slightly more educated than men; nevertheless, 90 percent of them are not employed and spend 5.5 hours per day on housework and a little over one hour per day on childcare, while the large majority of men are employed (97 percent). These sample averages are in line with those of Abbink et al. (2020).
In the previous three months, women have been to two out of five possible locations outside of their homes and are fully involved in less than two out of 14 household decisions[7]

3.2 Baseline characteristics

The baseline characteristics of men and women are presented in Table 2. The share of men and women reporting that they have a modern lifestyle is the same (30 percent). Nevertheless, men appear more conservative than women. On average, 60 percent of women and 76 percent of men prefer to deal with a man in different types of professions.[8] Using likert scales for 12 statements on gender attitudes, I create a standardized weighted index following the procedure by Anderson (2008). Higher values of the index indicate more conservative gender attitudes: 60 percent of men have a gender attitudes index above the mean while the share drops to 42 percent for women.[9]
To measure gender stereotypes in a more objective way and to reduce social desirability bias, I pilot a non-verbal Implicit Association Test: respondents associate images representing items such as a broom, a veil, a long beard or a business, to images representing the categories *Male-Money; Female-Home* in the first repetition, and *Male-Home; Female-Money* in the second one (details on the test design and the score calculation are provided in Appendix A.3). A higher score of the non-verbal IAT indicates having more gender stereotypes. The results from the IAT are in line with the self-reported measures previously described and confirm that men have stronger gender stereotypes (Table A5 shows that the index is correlated at the conventional levels with self-reported measures of gender stereotypes).
To measure control preferences, I follow Ashraf (2009) and I rely on spouses' perceptions. I create a weighted index of the information obtained by asking respondents who in their household is mainly in control of savings, loan revenues, money for groceries and money for clothes. For 27 percent of

[7]The five places the wife went to are: local bazar; meet a friend; parents' home; NGO/MFI; local health facility; public place in a different union.I quantify involvement in decision-making by asking women to indicate on a scale how involved they are in 14 different household decisions: children education; children marriage; children health; food expenditure; whether the husband should start a new activity; whether the wife should work outside for pay; wife's earnings; large household expenditures; take-up loans; use of loan money; livestock and their products; land; fertility; leisure. Women's involvement broken down by household decision is presented in Figure A1.

[8]The five professions include doctor, elementary school teacher, mobile money agent, shop assistant and MFI officer. Figure A2 presents preferences for each profession separately for men and women.

[9]The summary statistics for the 12 statements are presented in Table A2.

men and 74 percent of women, the husband is mainly in control of resources.

Similarly to Bursztyn et al. (2020) and Bernhardt et al. (2018), men value their reputation within the community and misperceive social pressure more than women do. To capture social pressure, I use two measures. First, following Bernhardt et al. (2018), I asked respondents to rank from one to five the most important values in their life. Based on the ranking given to community respect I calculated a reputation index. If individuals ranked it first, I assigned five points. If it was ranked second, four points and so on until the fifth position which received only one point. I normalized the variable using the methodology by Anderson (2008): individuals with a higher index value their reputation more. The second measure was obtained by asking respondents how much a situation that describes social desirability applies to their life. I categorized someone as valuing village approval if the situation applied to them.[10] Men are more concerned than women about their reputation within their community: 45 percent of men have a reputation index above the mean (38 percent for women) and 79 percent find that the situation described applied to them (73 percent for women). When I calculate beliefs wedges as in Bursztyn et al. (2020), 84 percent of men and 77 percent of women underestimate their community support for women's labor force participation while 85 percent of participants overestimate how much their community disagrees with men consulting their wives on financial matters. Finally, contrary to what is reported by Abbink et al. (2020), women are less risk-averse than men (14 percent against 43 percent).

3.2.1 Perceived skills and decision-making

To measure perceptions, I asked respondents to grade from 0 to 10 their and their spouses' skills at problem-solving, money management, housework and childcare. I compute the gender gap in perceived skills as the difference between the grade given to the husband's and to the wife's skills. A positive gap indicates that respondents rated the husband's skills higher than the wife's. The perceived skills gender gaps are in line with the task specialization of spouses (90 percent of women are homemakers and 60 percent of men are self-employed). In housework and childcare, the share of men and women with a positive perceived skills gap (they rate the husband's skills higher than the wife's) is below 10 percent and it is not statistically different between spouses. However, the share of respondents who rated the husband higher at problem-solving and money management is larger for women than for men (50 and 35 percent respectively). Figure A3 presents the average ratings of husband and wife's skills for each skill type, by gender: more women than men perceive

[10] Men were presented with the following scenario: *Ashraf has a tailor shop in a village. His shop is open every day from Sunday to Saturday, from 9 am to 9 pm. The other people in his village have high respect for hard workers. He keeps the shop open all the time because he wants everyone to know he is a hard worker. How much are you like Ashraf? 1 = Completely; 2 = Somewhat; 3 = Not at all.*
Women were presented with the following scenario: *Sumaiya is married and has three children. Her husband has a tailor shop in the village. Every evening before going to bed she washes and dries his work clothes so that they look neat the next morning. The other people in their village have high respect for good and caring wives. She washes her husband's work clothes every evening before sleeping because she wants everyone to know she is a good wife. How much are you like Sumaiya? 1 = Completely; 2 = Somewhat; 3 = Not at all.*

the husband to be better at problem-solving and money management, and the average woman rates her own skills lower than men do. The gender gap for problem-solving skills is 0.54 for men and 0.97 for women while the gap size for money management is 0.46 for men and 1.17 for women (both differences are statistically significant). For housework and childcare, the perceived skills gaps are negative and not statistically different by gender.

Figure 3 shows the correlations between the gender gap in perceived skills and women's involvement in decision-making. When the husband is perceived to be more capable than the wife in analytical tasks (problem-solving and money management), women exhibit lower levels of involvement in decisions, including those where they are usually more likely to have a say, such as children's outcomes.[11] This pattern persists also for perceived skills in care tasks (housework and childcare), although coefficients are not statistically different from zero.

4 Empirical Analysis

I begin the empirical analysis by examining how the outcomes of the IDM game differ between treatment groups.

I examin four main outcomes, two at the extensive margin and two at the intensive margin. The extensive margin outcomes correspond to a binary variable that equals one if an individual allocates a minimum of 3 high-payoff matrices to the wife, and a binary variable that equals one if the belief gender gap (the difference between the guessed number of correct answers of the husband and of the wife) is greater than zero. The intensive margin outcomes are composed of the number of high-payoffs allocated to the wife and the size of the belief gender gap.

I focus on belief gender gaps, rather than absolute beliefs about women's, because my underlying assumption is that beliefs and allocations are based on individuals' perceptions of their own abilities relative to their spouse and that women are perceived as less capable than men.In the pre-treatment round, women answer correctly a larger number of matrices than their husbands, but men perceive themselves to be more capable (Table A3). Since baseline outcomes are balanced, I can compare post-treatment outcomes for treated and control couples. After the information provision, individuals have a smaller belief gap and the share of individuals with a positive belief gap decreases. However, the differences between treated and controls are significant only within the sample of men (Table 3).

One of the concerns with the experimental design is that individuals who receive the information about the wife "being good" might reason that consequently the husband "should be even better". However, the smaller belief gaps in the treated group are driven by higher beliefs about the women's scores. This suggests that the message delivered by the information treatment was correctly received and understood. Nevertheless, the number of high-payoff matrices allocated to

[11] The breakdown of women's involvement in household decisions by item is presented in Figure A1.

women in the post-treatment round does not change across arms. On average, men allocate two out of five high-payoffs to their wives while women keep three for themselves. Although the share of individuals who allocate three or more high-payoffs to the wife is higher in the treated group (35 percent of treated men and 68 percent of treated women against 31 percent of control men and 64 percent of control women), the differences are not statistically significant.[12]

These averages show weak effects of the treatment on the game outcomes but they mask an important heterogeneity that depends on baseline perceptions of abilities. If some individuals are already aware of women's true skill levels they should react less strongly to the provision of information (they did not receive any *new* information). In the next section, I evaluate the impact of the information treatment depending on spouses' perceptions of each other's abilities.

4.1 Main specification

To test for the presence of statistical bias and allow heterogeneous responses to treatment by gender, I run the following regression, separately for men and women:

$$Y_{iv} = \alpha + \beta_1 \, Treatment_i + \beta_2 \, P_i + \beta_3 \, Treatment_i \times P_i + \beta_4 \, X_i + \delta_v + \varepsilon_{iv} \tag{1}$$

where Y_i corresponds to the outcomes of interest for individual i in the post-treatment round of the game. X_i is a vector of baseline characteristics, including the stratification variables (the age gap between spouses, wife education, husband education, marriage length, marriage length squared); $delta_v$ is the village fixed effect and ε_i is the error term (standard errors are clustered at the couple level, which is the level of the randomization). $Treatment_i$ is a binary variable that equals one if the participant received the information treatment. P_i measures perceived skills gaps in two ways. The first one involves a binary variable that equals one if the respondent thinks the husband is better than the wife at analytical tasks (problem-solving and money management). The second variable equals one if the belief gap of the respondent in the pre-treatment round of the game is positive (the respondent thinks the husband is better than the wife at answering matrices). If individuals' choices are driven by a statistical bias, the treatment should fill information gaps for those with lower baseline perceptions about women: β_3 should be positive for allocations (more high-payoffs are assigned to the wife) and negative for the belief gaps (the difference between the guessed score of the husband and wife shrinks).

Results are reported in Figure 4. The four panels show the impact of the treatment on the belief gap being positive, the size of the belief gap, the number of high-payoffs and whether at least three high-payoffs are assigned to the wife. Within each panel, the left graph shows the interaction of the treatment with perceived abilities in analytical tasks. The right graph shows the interaction

[12] Figure A4 provides further details on the distribution of the high-payoff matrices allocated to women in both rounds, by gender and treatment arm.

with perceived abilities in the game.
Men in the treatment group, who perceived their wives as equally or more skilled, are not receiving new information: their beliefs are further confirmed and their behavior does not change. They are 9 percentage points less likely to have a positive belief gap and the treatment has no significant impact on their allocations. Men who perceive themselves as better than their wives (both in analytical tasks and in the game) and do not receive the information treatment are more likely to have a positive belief gap (the coefficient equals 16 percentage points and 42 percentage points, respectively), have a larger belief gap (the coefficient is 0.28 and 1.05, respectively) and are less likely to give more high-payoffs to their wife (the coefficients equals -18 and -11 percentage points, significant at the one percent level).
On the other hand, the information reverses the outcomes of treated men who perceive themselves as more capable in analytical tasks: they are 25 percentage points more likely to allocate at least three high-payoffs to their wives (the coefficient is significant at the one percent level) while for perceived skills in the game, the treatment has a slightly positive but not significant impact on allocations.

The information has no impact on beliefs or allocations of treated women who do not consider their husbands to be better. On the other hand, women in the control group who consider the husband more skilled are 35 percentage points more likely to have a positive belief gap and 21 percentage points less likely to keep three or more high-payoffs (coefficients are significant at the one percent level). After the information treatment, women with higher perceptions of their husband's skills in the game keep 0.6 payoffs more and are 20 percentage points more likely to keep at least three high-payoffs for themselves. Their belief gap decreases, but the coefficients are not significant. The effects for perceived skills in analytical tasks are weaker but point in the same direction.
These results support the initial hypothesis of statistical bias. Those who perceive the husband to be more skilled than the wife assign fewer high-payoff matrices to her and have larger belief gaps, unless they receive the information treatment which reverses both allocations and beliefs.

Spouses who have been together longer should be more able to observe abilities. On the other hand, spouses in younger couples have been less exposed to each other's skills, and therefore they might use gender stereotypes to fill the information gaps. Table A4 shows that spouses' cognitive skills are correlated. To further investigate whether the unobservability of skills makes individuals turn to gender stereotypes, I interact the treatment with a variable that equals one if the couple has been married for less than 5 years. The results are presented in Figure 5. The treatment reduces the belief gap of men who have been married for a shorter time and improves the allocations of high-payoffs to their wife. For women, the treatment has the opposite effect, although coefficients are not significant: the belief gap increases and they keep fewer high-payoffs. This suggests that spouses in recent unions are less able to observe each other's abilities. The information provision fills the information gaps and reverses the outcomes of men. For women, the information reinforces

the stereotype: one possible explanation is that since they have been less able to observe their husband's skills, when they receive the information treatment women might think that if they are good, then the husband should be better. To further investigate this hypothesis, in Figure A5 I restrict the sample to women, and I look at the impact of treatment on beliefs about the husband's and the wife's score. I run these regressions separately for women in recent (five years or less) and longer (more than 5 years) unions. Although in more recent unions coefficients are not significant, the impact of the information is larger on beliefs about their husband than about themselves. For women in longer unions, both coefficients are close to zero. This seems to support the idea that it is harder to act on gender stereotypes of women in more recent unions who have been exposed to their and their husband's skills for a shorter amount of time.
Further evidence that misperceived abilities drive the results is presented in Figure A6, where I interact the treatment with a variable that equals one if the husband is perceived to be better than the wife at housework and childcare. Indeed, the treatment has no impact: these skills are observed more easily within the household, individuals do not receive new information and consequently, their beliefs and allocations do not change.

In addition, I investigate whether the treatment has an effect on women's involvement in actual household decision-making two weeks after the lab experiment. For each couple, I regress the wife's involvement in decisions on the treatment status. Overall, the coefficients are positive, indicating that following the information provision women seem to be more involved (Figure 6). The impact is larger for decisions about children and household investments, but only the latter is significant at the 10 percent level. As in the main regressions, in Figure 7 I look at how results change based on perceived abilities. Even if not statistically significant, the coefficients for treated women, who at baseline were perceived to be less capable than their husbands in analytical tasks, are positive for all the household decisions. The largest impact is on decisions regarding children and household expenditure. Women in the control group, who are perceived to be less capable, are less involved in household decisions. Women in treated couples, who were not perceived to be less capable are more involved in decisions about labor supply and household investments, but coefficients are smaller (and not statistically different from zero). This provides additional evidence that the treatment filled information gaps and translated to real-life outcomes. It further supports the hypothesis that intra-household dynamics are shaped by gender stereotypes that depend on misperceived abilities.

The results presented so far suggest that if couples cannot perfectly observe skills, they turn to gender stereotypes to fill the information gaps. Nevertheless, these stereotypes can be attenuated by providing information on women's true skill levels.[13] To be able to fully identify the presence of statistical bias and rule out confounding factors, in the next section I focus on alternative mechanisms.

[13]Results are robust to a diff-in-diff specification. I do not use an ANCOVA specification because the outcomes of interest are autocorrelated. Results are available upon request.

4.2 Alternative Mechanisms

Based on the existing literature on intra-household decisions, I investigate whether alternative mechanisms such as gender beliefs, social pressure, control preferences and risk aversion. I explore these heterogeneous effects by interacting the treatment variable with baseline characteristics. The results are presented for the sample of men and women in Table 4 and Table 5, respectively. In both tables, Panel A reports the results when the dependent variable indicates the belief gender gap is positive. In Panel B the dependent variable is the size of the belief gender gap. In Panel C the dependent variable equals one if the respondent allocated three or more high-payoffs to the wife, while in panel D it corresponds to the number of high-payoffs assigned to the wife.

Control Preferences

Couples' behaviors and beliefs in the game could reflect intra-household arrangements concerning the control of resources. Irrespective of abilities, individuals might want to be in control even if it could lead to inefficient outcomes (Ashraf, 2009, Mani, 2020, Schaner, 2017). By design, the game attenuates the influence of control preferences: the final earnings of each player depend on both spouses' scores. While the treatment provides an objective assessment of women's abilities, no information about men is available. Allocating more high-payoff matrices to the husband, after being informed of the wife's skills, could increase the chances of lower final earnings, without increasing the share under the sole control of one player.

The information treatment decreases the belief gap of men with weaker baseline control preferences: they are 12 percentage points less likely to have a positive gap, and their belief gap decreases by 0.36 (Columns 1, Panel A and B). While the treatment does not significantly decrease the likelihood of the belief gap being positive for men with stronger control preferences, it reduces its size by 0.28 (significant at the 5 percent level). In terms of allocations, treated men with stronger preferences for control are 21 percent more likely to give the majority of the high-payoffs to their wife (Column 1, Panel C) and specifically, they give her 0.37 more payoffs (Column 1, Panel D). Both coefficients are significant at the one percent level. Among women, the treatment has a slight impact only on the belief of those with weaker preferences for the husband being in control of resources: they are 8 percentage points less likely to have a positive belief gap (Column 1, Panel A) and their belief gap decreases by 0.28 units (Column 1, Panel B). Coefficients are significant at the 10 and 5 percent level, respectively. The results for women who think the husband should be more in control of resources are the opposite compared to men. The treatment increases their belief gap by 0.21 and decreases the number of payoffs they keep for themselves (although only the coefficient for beliefs is significant, at the 10 percent level). A possible interpretation is that in this case as well, these women think that if *They are good* then the husband *Should be better*.

Risk Aversion

Playing multiple rounds allows players to get familiar with the setting of the game. Nevertheless, risk-averse individuals might prefer to follow gender stereotypes when taking decisions. Women might delegate the task of answering matrices that are worth more to their spouse while men might keep more high-payoffs for themselves. This is consistent with the existing literature stating that the risk of reputation costs *within the household* makes individuals choose less optimal options (Abbink et al., 2020, Buchmann et al., 2023). I measure risk using a likert scale from zero (*Not at all willing to take risk*) to ten (*Very willing to take risk*) and I record someone as risk averse if they select zero on the scale.

The treatment decreases the belief gap of men who are not risk averse (Column 2, Panel A). Men in the control group who are not risk averse give their wife 0.37 fewer high-payoffs and are 16 percent less likely to give her the majority of the high-payoffs (Column 2, Panel C and D). Men with the same characteristics who receive the information treatment are more likely to give more high-payoffs to their wives, although the coefficients are not statistically significant. While risk averse women in the control group have larger belief gaps and keep fewer high-payoff questions, risk averse women in the treated group are 24 percent less likely to have a positive belief gap and they are 22 percent more likely to keep the majority of the high-payoffs. They keep 0.56 more high-payoffs (coefficients are significant only for treated women).

Gender Beliefs and Social Pressure

Individuals bring a set of pre-existing gender attitudes to the game that could shape their outcomes. Moreover, behavior could be influenced by social pressure, which is the reputational cost of not following a social norm (Bursztyn et al., 2020, Bernhardt et al., 2018). The game takes place in a private setting, which should reduce these effects. I capture gender beliefs using the self-reported measures described in Section 3.2 and the non-verbal implicit association test (NV-IAT). The coefficients are reported in Columns 3 and and 4 of each panel. Social pressure is measured as the extent to which the respondent values village approval, as explained in Section 3.2, and the coefficients are shown in Column 5 of each panel.

The treatment decreases by 0.2 the belief gap of men with less conservative gender attitudes and who are less concerned about their reputation within the community (coefficients are significant at the 5 and 10 percent level). By contrast, it increases by 8 percentage points the likelihood of a positive belief gap for men with more conservative gender attitudes and who value their reputation more (although the coefficients are only significant for the former). These results are consistent with the hypothesis that even after being informed about their wife's skills, men with more conservative gender views still believe that they are better than their wife. Although the coefficients are not significant, treated men with more conservative attitudes are more likely to give more high-payoffs to their wife. The coefficients are consistent for both the self-reported measure and the NV-IAT.

Treated women with stronger gender stereotypes (measured with the NV-IAT) are 8 percent more likely to have a positive belief gap and they keep fewer high-payoffs for themselves (but the coeffi-

cients for allocations are not significant).

The analysis of alternative mechanisms yields further evidence of the presence of a statistical bias in intra-household dynamics.[14] Filling information gaps on women's skills reverses the allocations and beliefs of men with stronger control preferences. For the sample of women, the information treatment attenuates the impact of risk aversion on allocations and beliefs by removing the component that is due to misperceived abilities. The treatment does not reverse the outcomes for individuals with stronger gender beliefs or social pressure, which is expected given that these are harder to change and the information provided is not tailored to challenge social norms.

5 Conclusion

In this paper, I study how gender stereotypes that arise from information gaps about women's abilities shape intra-household dynamics. My sample is composed of 525 married couples from 42 villages in rural Bangladesh.

Within households, women are perceived as less capable than men at problem-solving and money management. The greater the gap in their perceived abilities relative to their husband, the less they are involved in household decisions. To investigate whether this gender discrimination has a statistical component (that is, it is based on unobserved women's abilities), I designed an incentivized decision-making game. Spouses played multiple rounds simultaneously but separately: they answered questions from the Raven colored progressive matrices and earned points for each correct answer. Some matrices were worth more, some were worth less and players decided privately how many high and low-payoff matrices they and their spouses would answer. Final earnings depend on both spouses' performances. The main outcomes of interest are the number of high payoffs assigned to women, and the belief gender gap (the difference between the guessed number of correct answers of the husband and of the wife). Couples were randomly assigned to treatment and control arms. Treated couples were informed that the wife possesses higher than average cognitive skills. The control arm received no information.

Despite answering the same number of questions correctly, women were perceived to be less capable than men. After the information provision, allocations and beliefs seemed not to change. However, this masks an important heterogeneity. While some individuals in the sample were already aware of women's abilities, others were less able to observe them. The information provision had a significant and positive impact on the allocations of the latter group: those who thought the husband is more capable than the wife at baseline. After the treatment, they allocated more high payoffs to the wife. The information had a stronger impact on younger couples that have been exposed to each other's abilities for less time. This supports the hypothesis that the unobservability of women's

[14] I run additional regression interacting the treatment with each individual variable used to compute the indices, a measure of the value of village approval, the spouse's level of education, and which gender respondents prefer in five different professions. Results are consistent and available upon request.

skills makes individuals turn to gender stereotypes to fill the gaps. In addition, the treatment reversed the outcomes of men with stronger control preferences and of risk-averse women: the belief gap reduced and the allocations in favor of the wife increased. Providing information did not seem to affect the outcomes of individuals concerned about their reputation within the community or with conservative gender attitudes (measured with self-reported indicators and a non-verbal IAT). Two weeks after the experiment, women in treated couples reported being more involved in household decisions, especially those concerning land and loan take-up.

These results show that the unobservability of women's skills shapes intra-household dynamics by making individuals turn to gender stereotypes. The fact that the outcomes of the individuals in the treated group changed after the information provision indicates that gender discrimination within households has a statistical component that can be corrected. From a policy point of view, these results suggest that interventions that aim at improving women's bargaining power within the household should make their skills more observable. Workshop participation certificates are an intervention in this direction. Future research should focus on exporting this lab experiment to a setting where spouses' decisions are observed by their community.

6 Tables

Table 1: Balance check of baseline characteristics for treated and control couples.

	Mean	(Std.)			T-test
Variables	Control		Treated		Diff.
Household size	5.06	(0.11)	4.80	(0.10)	0.26*
Household monthly income below 20,000 BDT	0.83	(0.02)	0.86	(0.02)	-0.03
Went without eating at least once, past 12 months	0.44	(0.03)	0.39	(0.03)	0.05
Wife age	31.63	(0.43)	32.20	(0.44)	-0.57
Husband age	39.51	(0.55)	39.98	(0.58)	-0.47
Marriage length, in years	15.45	(0.47)	16.44	(0.49)	-0.99
Wife completed education	1.27	(0.04)	1.21	(0.04)	0.06
Husband completed education	1.06	(0.04)	0.96	(0.04)	0.10*
Husband is employed	0.97	(0.01)	0.98	(0.01)	-0.01
Wife is employed	0.08	(0.02)	0.07	(0.02)	0.01
Wife daily hours of housework	5.40	(0.13)	5.54	(0.13)	-0.14
Wife daily hours of childcare	1.58	(0.13)	1.32	(0.12)	0.26
Number of decision wife is fully involved	1.48	(0.16)	1.84	(0.20)	-0.36
Number of places out of 5 wife went to, past 3 months	2.09	(0.07)	2.18	(0.08)	-0.09
F-test of joint significance (p-value)	0.54				
Number of observations	525				

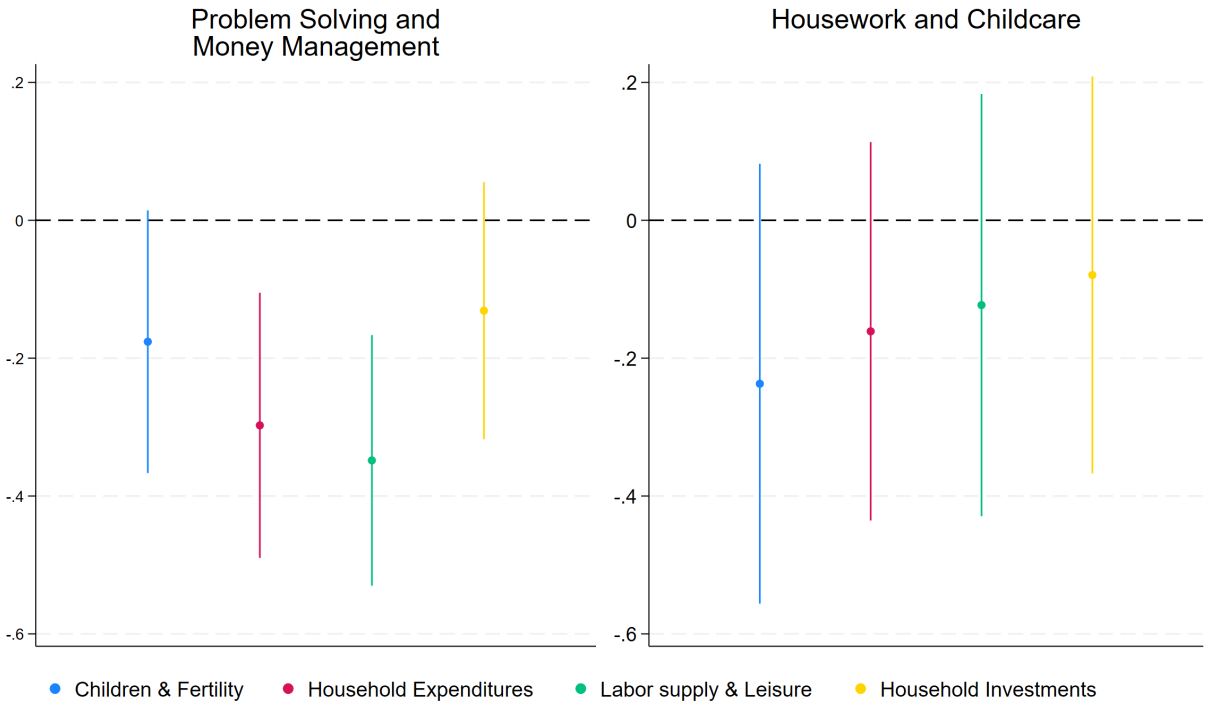
The value displayed for t-tests are the differences in the means between couples randomly assigned to the treatment and control arm. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regressions for the t-test include village fixed effects. Education categories: no schooling; primary; above primary. Household monthly income is reported by the husband. The wife involvement in 14 household decisions is broken down by item in Figure A1. The five places the wife went to include: local bazar; meet a friend; parents' home; NGO or MFI; local health facility; public place in a different union.

Table 2: Baseline characteristics, by gender.

	Mean	(Std)			T-test
	Men		Women		Difference
Gender attitudes and norms					
Modern lifestyle	0.29	(0.46)	0.28	(0.45)	0.01
Mother ever worked for pay	0.14	(0.35)	0.13	(0.34)	0.01
Mostly prefers men for different professions	0.76	(0.43)	0.61	(0.49)	0.15***
Gender attitudes index above mean	0.60	(0.49)	0.42	(0.49)	0.18***
Non-verbal IAT score above mean	0.46	(0.50)	0.38	(0.49)	0.08***
Husband mostly in control of resources	0.27	(0.44)	0.74	(0.44)	-0.47***
Social pressure					
Reputation index above mean	0.45	(0.50)	0.38	(0.49)	0.07**
Values village approval	0.80	(0.40)	0.74	(0.44)	0.06**
Overestimates community disapproval					
A man should consult his wife for financial decisions	0.85	(0.36)	0.83	(0.37)	0.02
A woman should work for pay outside of the house	0.84	(0.37)	0.77	(0.42)	0.07***
Risk averse	0.43	(0.50)	0.14	(0.34)	0.29***
Skills					
Positive skills rating gap (husband's - wife's)					
Finding solutions	0.38	(0.49)	0.47	(0.50)	-0.09***
Money management	0.34	(0.48)	0.50	(0.50)	-0.16***
Housework	0.07	(0.25)	0.08	(0.27)	-0.01
Childcare	0.09	(0.29)	0.10	(0.29)	-0.01
Observations	525				

The *T-test difference* column reports the difference in means for men and women: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The professions for which they prefer a man or a woman are: doctor, elementary school teacher, mobile money agent, shop assistant, MFI officer (A breakdown by profession is provided in Figure A2). An individual mostly prefers men for the majority of professions if they selected "men" in at least three out of five options. The gender attitudes index, the husband control of resources and the reputation index are computed using the procedure by Schwab et al. (2020); for higher values of the control index, the respondent thinks the husband should be the one mostly in control of resources; higher values of the gender attitudes index indicate more conservative gender attitudes (The statements on attitudes towards women and their summary statistics are presented in Table A2). Higher values of the reputation index indicate more concern about one's reputation within the community. Higher values of the non-verbal IAT score indicate stronger gender stereotypes. Details on the calculation of the non-verbal IAT score are provided in Section A.3. Overestimation of community disapproval is measured as in Bernhardt et al. (2018) and Bursztyn et al. (2020). Risk preferences are self reported on a scale from 0 to 10. The skills rating gap is calculated as the difference between the rating given to husband's and the wife's skills on a scale from 0 to 10. Figure A1 shows the average rating given to spouses by gender and skill type.

Figure 3: Correlations between the difference in perceived abilities of women relative to their husbands' and involvement in decision making.



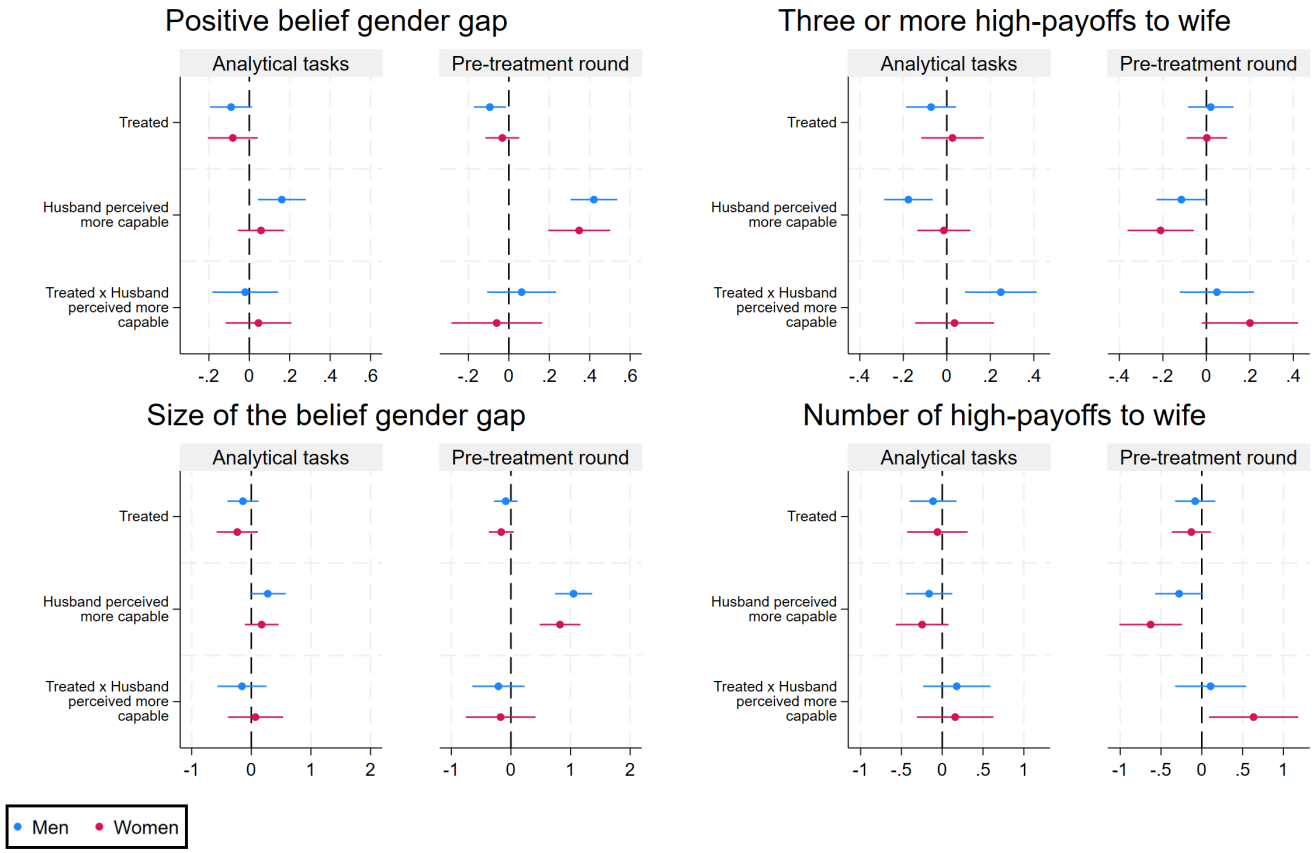
Note: OLS regressions. The dependent variable is the wife's involvement in decision-making. Decisions are grouped into four categories using the standardized weighted index by Schwab et al. (2020). Higher values indicate more involvement. *Children and fertility* includes children's health, children's education, children's marriage and fertility decisions. *Household expenditures* includes decisions on food expenditures, livestock and large expenditures. *Labor supply and leisure* includes decisions on the wife labor force participation, on the husband's starting a new activity and on leisure. *Household investments* includes decisions on land and loans. The independent variable of interest equals one for positive perceived skill gaps (the difference between the rating given to the husbands' and the wives' skills) in analytical tasks (Problem-solving and money management, left panel) and care tasks (Childcare and housework, right panel). A positive gap indicates that the husband is considered more skilled than the wife (Results are consistent, but not presented, when I use a continuous variable to measure the size of the skill gap instead of a binary one). All regressions include stratification and control variables. Stratification variables: wife completed schooling, marriage length, village of residence. Control variables: spouses' age gap, husband completed schooling, marriage length squared. Variables are calculated at the couple level. Standard errors are clustered at the couple level.

Table 3: Comparison of post-treatment outcomes. OLS, by gender.

	Size of belief gender gap (1)	Positive belief gender gap (2)	Belief on wife's score (3)	Belief on husband's score (4)	Nr high-payoffs to wife (5)	More high-payoffs to wife (6)
Panel A: Men						
Treated	-0.21**	-0.09**	0.23*	0.02	-0.03	0.04
	(0.10)	(0.04)	(0.13)	(0.11)	(0.10)	(0.04)
Control Mean	0.36	0.34	6.70	7.06	2.09	0.31
Observations	525	525	525	525	525	525
Panel B: Women						
Treated	-0.17	-0.04	0.09	-0.09	-0.01	0.05
	(0.11)	(0.04)	(0.10)	(0.11)	(0.10)	(0.04)
Control Mean	0.14	0.28	7.15	7.30	2.98	0.64
Observations	499	499	521	500	525	525

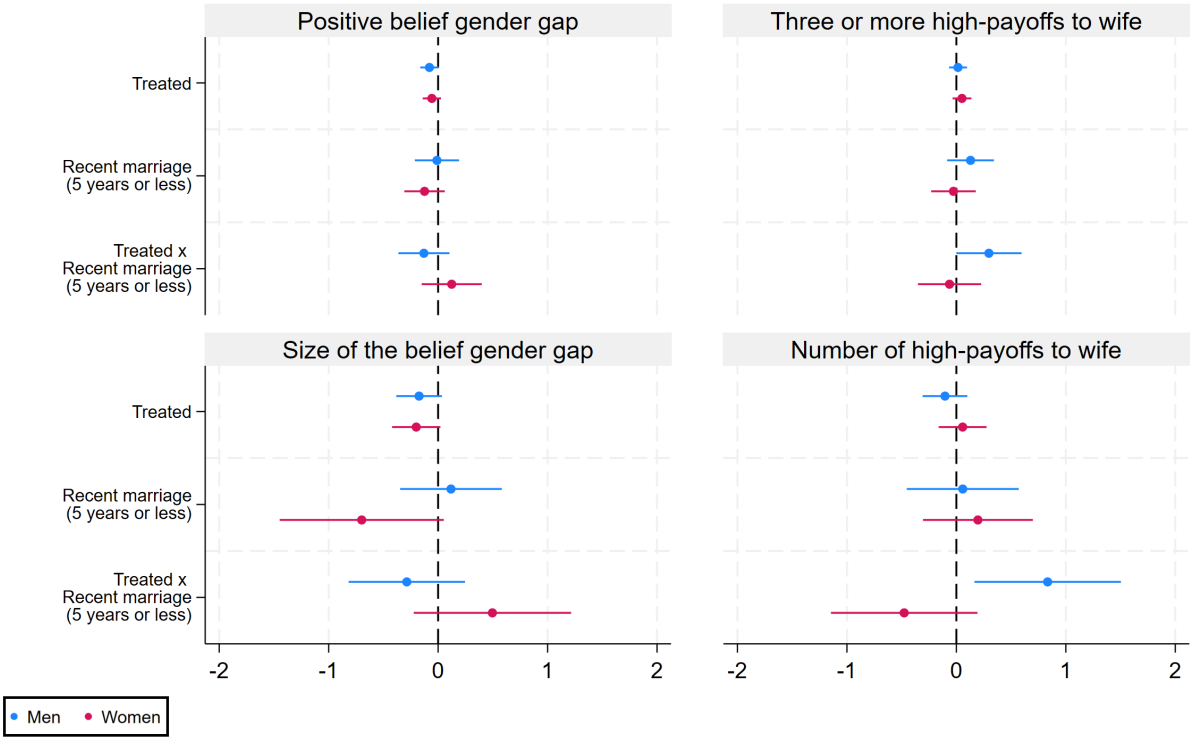
Panel A reports the results for the sample of men. Panel B for the sample of women. The dependent variable in columns (1) is the size of the belief gap. In column (2), whether the belief gap is positive (The husband is perceived better than the wife in the game). In column (3) is the belief on the wife's score. In column (4) is the belief on the husband's score. In column (5) is the number of high-payoffs allocated to the wife. In column (6), whether the respondent allocated 3 ore more high-payoffs to the wife. Standard errors clustered at the couple level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4: Treatment effects based on perceived abilities in analytical tasks and in the game. OLS, by gender.



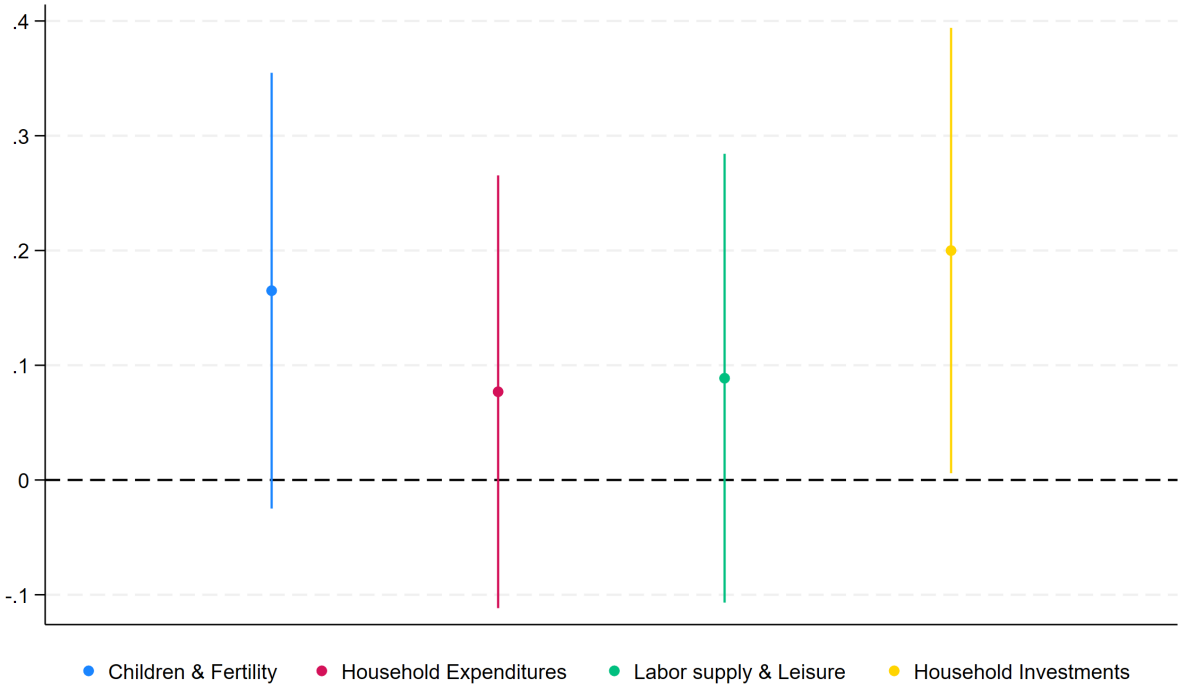
Note: Regressions are run separately for men and women. In each panel, the model on the left includes an interaction term with a binary variable that equals one if the respondent perceives the husband as more capable than the wife at analytical tasks (Problem-solving and money management). The model on the right includes an interaction term with a binary variable that equals one if the respondent perceived the husband as more capable than the wife at answering the Raven matrices in the previous round of the game. All regressions include stratification and control variables. Stratification variables: wife completed education, marriage length, village of residence. Control variables: spouses' age gap, husband completed education, marriage length squared. Standard errors are clustered at the couple level.

Figure 5: Treatment effects based on marriage length. OLS, by gender.



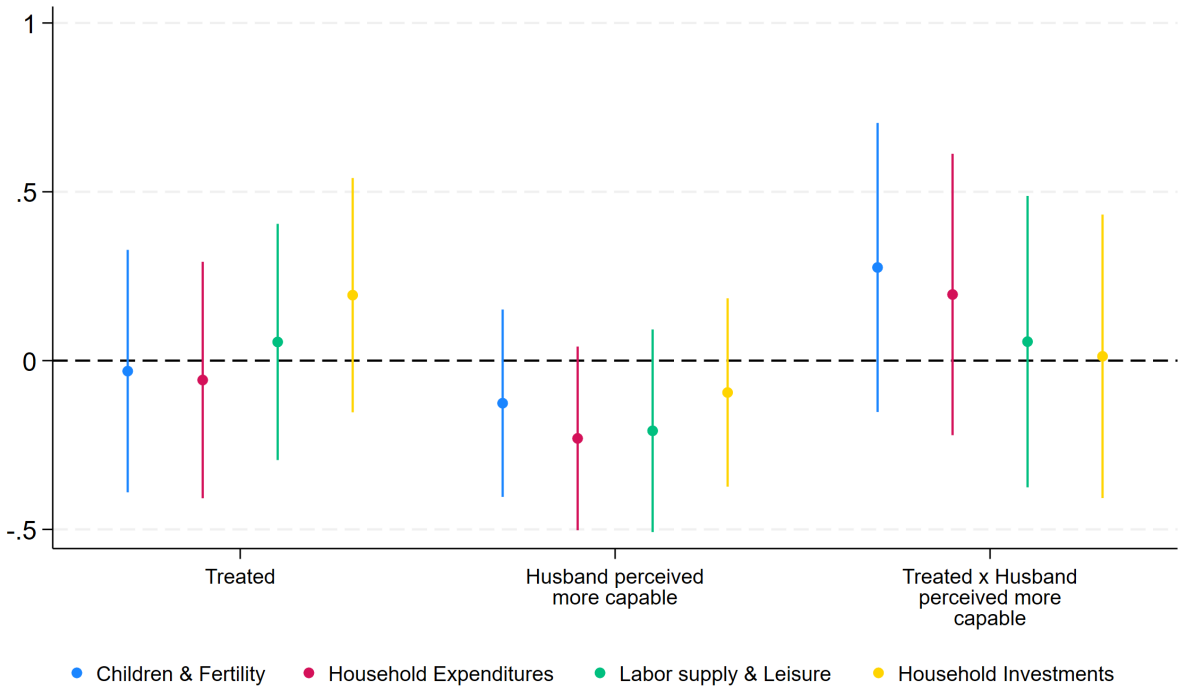
Note: Regressions are run separately for men and women. Each regression includes an interaction term with a binary variable that equals one if the spouses have been married for 5 years or less. All regressions include stratification and control variables. Stratification variables: wife completed education, marriage length, village of residence. Control variables: spouses' age gap, husband completed education, marriage length sqaured. Standard errors are clustered at the couple level.

Figure 6: Treatment effects on women's involvement in household decisions two weeks after the lab experiment



Note: OLS regressions. The dependent variable is measured during the follow-up survey, two weeks after the experiment (conducted with 455 women). Decisions are grouped into four categories using the standardized weighted index by Schwab et al. (2020). Higher values indicate more involvement. *Children and fertility* includes children's health, children's education, children's marriage and fertility decisions. *Household expenditures* includes decisions on food expenditures, livestock and large expenditures. *Labor supply and leisure* includes decisions on the wife labor force participation, on the husband's starting a new activity and on leisure. *Household investments* includes decisions on land and loans. The coefficients reported are those for treatment status (=1 if treated). All regressions include stratification and control variables. Stratification variables: wife completed schooling, marriage length, village of residence. Control variables: spouses' age gap, husband completed schooling, marriage length squared. Variables are calculated at the couple level. Standard errors are clustered at the couple level.

Figure 7: Treatment effects on women's involvement in household decisions 2 weeks after the lab experiment, based on perceived abilities in analytical tasks.



Note: OLS regressions. The dependent variable is measured during the follow-up survey, two weeks after the experiment (conducted with 455 women). Decisions are grouped into four categories using the standardized weighted index by Schwab et al. (2020). Higher values indicate more involvement. *Children and fertility* includes children's health, children's education, children's marriage and fertility decisions. *Household expenditures* includes decisions on food expenditures, livestock and large expenditures. *Labor supply and leisure* includes decisions on the wife labor force participation, on the husband's starting a new activity and on leisure. *Household investments* includes decisions on land and loans. Each regression includes an interaction term with a binary variable that equals one if the respondent perceives the husband as more capable than the wife at analytical tasks (Problem-solving and money management), stratification and control variables. Stratification variables: wife completed schooling, marriage length, village of residence. Control variables: spouses' age gap, husband completed schooling, marriage length squared. Variables are calculated at the couple level. Standard errors are clustered at the couple level.

Table 4: Alternative mechanisms. OLS, men sample.

Baseline characteristics:	Husband control of resources (1)	Risk Aversion (2)	Gender attitudes (self-reported) (3)	NV-IAT score (4)	Values reputation (5)	Husband control of resources (1)	Risk Aversion (2)	Gender attitudes (self-reported) (3)	NV-IAT score (4)	Values reputation (5)
Dependent variable	*Panel A: Positive belief gender gap*					*Panel B: Size of the belief gender gap*				
Treated	-0.12**	-0.10*	-0.12***	-0.10**	-0.09**	-0.36***	-0.18	-0.25**	-0.21**	-0.21**
	(0.05)	(0.06)	(0.04)	(0.04)	(0.04)	(0.13)	(0.14)	(0.10)	(0.10)	(0.10)
Baseline characteristics	0.02	-0.01	0.02	0.01	0.01	0.15	0.00	0.06	0.03	0.01
	(0.04)	(0.06)	(0.03)	(0.03)	(0.03)	(0.10)	(0.16)	(0.08)	(0.07)	(0.07)
Treated × Baseline characteristics	-0.05	-0.00	0.08*	0.00	0.04	-0.28**	-0.10	0.11	-0.06	0.05
	(0.06)	(0.08)	(0.04)	(0.04)	(0.04)	(0.15)	(0.21)	(0.11)	(0.09)	(0.10)
Dependent variable	*Panel C: Three or more high-payoffs to the wife*					*Panel D: Nr of high-payoffs to the wife*				
Treated	0.16***	-0.00	0.04	0.04	0.04	0.18	-0.11	-0.04	-0.05	-0.03
	(0.05)	(0.06)	(0.04)	(0.04)	(0.04)	(0.13)	(0.13)	(0.11)	(0.10)	(0.10)
Baseline characteristics	-0.03	-0.16***	-0.05**	-0.04	-0.02	-0.01	-0.37**	-0.12*	-0.01	0.05
	(0.04)	(0.06)	(0.03)	(0.03)	(0.03)	(0.09)	(0.14)	(0.07)	(0.07)	(0.07)
Treated × Baseline characteristics	0.21***	0.12	0.04	0.02	0.03	0.37***	0.17	0.08	0.03	-0.03
	(0.06)	(0.09)	(0.04)	(0.04)	(0.04)	(0.15)	(0.21)	(0.10)	(0.09)	(0.11)
Observations	524	512	524	520	524	524	512	524	520	524
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

OLS regressions for the sample of men. *Dependent variables.* Panel A: equals 1 if the belief gender gap is positive. Panel B: size of the belief gender gap. Panel C: equals 1 if the respondent allocated 3 or more high-payoffs to the wife. Panel D: number of high-payoffs allocated to the wife. In columns (1) the baseline characteristic used is the index of the husband's control of resources (higher values indicate the husband is more in control of resources). In columns (2) is risk aversion (=1 if risk averse, 0 otherwise). In columns (3) is the self-reported measure of gender attitudes (higher values indicate more conservative attitudes). In columns (4) is the score of the non-verbal IAT (higher values indicate stronger gender stereotypes). In columns (5) is the reputation index (higher values indicate the respondents attach more value to their reputation in the community). Risk aversion equals one if the respondent indicated they are *Not at all willing to take risk*. The gender attitudes index, the husband control of resources and the reputation index are computed using the procedure by Schwab et al. (2020); for higher values of the control index, the respondent thinks the husband should be the one mostly in control of resources; higher values of the gender attitudes index indicate more conservative gender attitudes (The statements on attitudes towards women and their summary statistics are presented in Table A2). Higher values of the reputation index indicate more concern about one's reputation within the community. Higher values of the non-verbal IAT score indicate stronger gender stereotypes. All regressions include stratification and control variables. Stratification variables: wife completed schooling, marriage length, village of residence. Control variables: spouses' age gap, husband completed schooling, marriage length squared. Standard errors clustered at the couple level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Alternative mechanisms. OLS, women sample.

Baseline characteristics:	Husband control of resources (1)	Risk Aversion (2)	Gender attitudes (self-reported) (3)	NV-IAT score (4)	Values reputation (5)	Husband control of resources (1)	Risk Aversion (2)	Gender attitudes (self-reported) (3)	NV-IAT score (4)	Values reputation (5)
Dependent variable	*Panel A: Positive belief gender gap*					*Panel B: Size of the belief gender gap*				
Treated	-0.08*	-0.01	-0.04	-0.05	-0.04	-0.28**	-0.08	-0.17	-0.16	-0.16*
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.13)	(0.12)	(0.12)	(0.12)	(0.11)
Baseline characteristics	-0.01	0.09	0.02	-0.05	0.00	0.01	0.28	0.14*	-0.08	0.10
	(0.03)	(0.09)	(0.03)	(0.03)	(0.03)	(0.07)	(0.21)	(0.07)	(0.11)	(0.07)
Treated × Baseline characteristics	0.05	-0.24**	0.03	0.08*	0.02	0.21*	-0.52	-0.04	0.13	-0.00
	(0.04)	(0.12)	(0.05)	(0.05)	(0.04)	(0.12)	(0.34)	(0.11)	(0.14)	(0.12)
Observations	498	497	498	483	497	498	497	498	483	497
Dependent variable	*Panel C: Three or more high-payoffs to the wife*					*Panel D: Nr of high-payoffs to the wife*				
Treated	0.06	0.02	0.04	0.04	0.05	0.08	-0.04	-0.00	-0.00	0.02
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.13)	(0.12)	(0.11)	(0.11)	(0.11)
Baseline characteristics	0.03	-0.10	-0.02	0.01	-0.01	0.05	-0.13	-0.04	0.00	0.02
	(0.03)	(0.10)	(0.03)	(0.04)	(0.03)	(0.08)	(0.23)	(0.08)	(0.10)	(0.08)
Treated × Baseline characteristics	-0.02	0.22*	-0.04	-0.05	0.03	-0.11	0.54*	-0.11	-0.10	-0.01
	(0.04)	(0.13)	(0.05)	(0.05)	(0.04)	(0.11)	(0.30)	(0.13)	(0.13)	(0.12)
Observations	524	523	524	505	522	524	523	524	505	522
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

OLS regressions for the sample of women. *Dependent variables.* Panel A: equals 1 if the belief gender gap is positive. Panel B: size of the belief gender gap. Panel C: equals 1 if the respondent allocated 3 or more high-payoffs to the wife. Panel D: number of high-payoffs allocated to the wife. In columns (1) the baseline characteristic used is the index of the husband's control of resources (higher values indicate the husband is more in control of resources). In columns (2) is risk aversion (=1 if risk averse, 0 otherwise). In columns (3) is the self-reported measure of gender attitudes (higher values indicate more conservative attitudes). In columns (4) is the score of the non-verbal IAT (higher values indicate stronger gender stereotypes). In columns (5) is the reputation index (higher values indicate the respondents attach more value to their reputation in the community). Risk aversion equals one if the respondent indicated they are *Not at all willing to take risk.* The gender attitudes index, the husband control of resources and the reputation index are computed using the procedure by Schwab et al. (2020); for higher values of the control index, the respondent thinks the husband should be the one mostly in control of resources; higher values of the gender attitudes index indicate more conservative gender attitudes (The statements on attitudes towards women and their summary statistics are presented in Table A2). Higher values of the reputation index indicate more concern about one's reputation within the community. Higher values of the non-verbal IAT score indicate stronger gender stereotypes. All regressions include stratification and control variables. Stratification variables: wife completed schooling, marriage length, village of residence. Control variables: spouses' age gap, husband completed schooling, marriage length squared. Standard errors clustered at the couple level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Abbink, K., A. Islam, and C. Nguyen (2020). Whose voice matters? an experimental examination of gender bias in intra-household decision-making. Journal of Economic Behavior & Organization 176, 337–352.

Afzal, U., G. d'Adda, M. Fafchamps, and F. Said (2022). Intrahousehold consumption allocation and demand for agency: A triple experimental investigation. American Economic Journal: Applied Economics 14(3), 400–444.

Alesina, A., M. Carlana, E. La Ferrara, and P. Pinotti (2018). Revealing stereotypes: Evidence from immigrants in schools. National Bureau of Economic Research No.w25333.

Almås, I., A. Armand, O. Attanasio, and P. Carneiro (2018). Measuring and changing control: Women's empowerment and targeted transfers. The Economic Journal 128(612), F609–F639.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. Journal of the American statistical Association 103(484), 1481–1495.

Ashraf, N. (2009). Spousal control and intra-household decision making: An experimental study in the philippines. American Economic Review 99(4), 1245–1277.

Azad, S. S., M. Shehrin, M. M. Mishu, F. Andalib, and O. Nawal (2020). Report on emergency cash transfer through digital wallets during covid-19 pandemic. brac microfinance. Technical report, BRAC, Dhaka, Bangladesh.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: does exposure reduce bias? The Quarterly journal of economics 124(4), 1497–1540.

Bernhardt, A., E. Field, R. Pande, and N. Rigol (2019). Household matters: Revisiting the returns to capital among female microentrepreneurs. American Economic Review: Insights 1(2), 141–60.

Bernhardt, A., E. Field, R. Pande, N. Rigol, S. Schaner, and C. Troyer-Moore (2018). Male social status and women's work. In AEA Papers and Proceedings, Volume 108, pp. 363–67.

Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. Handbook of economic field experiments 1, 309–393.

Breza, E., M. Kanz, and L. F. Klapper (2020). Learning to navigate a new financial technology: Evidence from payroll accounts. National Bureau of Economic Research w28249.

Buchmann, N., P. Dupas, and R. Ziparo (2023). The good wife? reputation dynamics and financial decision-making inside the household. Technical report, National Bureau of Economic Research.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived social norms: Women working outside the home in saudi arabia. American economic review 110(10), 2997–3029.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. The Quarterly Journal of Economics 134(3), 1163–1224.

Cassidy, R., A. Dam, W. Janssens, U. Kiani, K. Morsink, et al. (2022). Father of the bride, or steel magnolias?: Targeting men, women or both to reduce child marriage. Working paper 22/50.

Castilla, C. (2015). Trust and reciprocity between spouses in india. American Economic Review 105(5), 621–624.

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

Chang, W., L. Diaz-Martin, A. Gopalan, E. Guarnieri, S. Jayachandran, and C. Walsh (2020). What works to enhance women's agency: Cross-cutting lessons from experimental and quasi-experimental studies. J-PAL Working Paper.

De Mel, S., D. McKenzie, and C. Woodruff (2009a). Are women more credit constrained? experimental evidence on gender and microenterprise returns. American Economic Journal: Applied Economics 1(3), 1–32.

De Mel, S., D. J. McKenzie, and C. M. Woodruff (2009b). Innovative firms or innovative owners? determinants of innovation in micro, small, and medium enterprises. Determinants of Innovation in Micro, Small, and Medium Enterprises (May 1, 2009). World Bank Policy Research Working Paper (4934).

Demirgüç-Kunt, A., L. Klapper, D. Singer, and S. Ansar (2022). The global findex database 2021: Financial inclusion, digital payments, and resilience in the Age of COVID-19. World Bank Publications.

Dhar, D., T. Jain, and S. Jayachandran (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. American economic review 112(3), 899–927.

DHS (2020). Bangladesh demographic and health survey 2017–18. National Institute of Population Research and Training (NIPORT).

Domino, G. and M. L. Domino (2006). Psychological testing: An introduction. Cambridge University Press.

Fafchamps, M., D. McKenzie, S. Quinn, and C. Woodruff (2014). Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in ghana. Journal of development Economics 106, 211–226.

Field, E., R. Pande, N. Rigol, S. Schaner, and C. Troyer Moore (2021). On her own account: How strengthening women's financial control impacts labor supply and gender norms. American Economic Review 111(7), 2342–2375.

Goetz, A. M. and R. S. Gupta (1996). Who takes the credit? gender, power, and control over loan use in rural credit programs in bangladesh. World development 24(1), 45–63.

Hashemi, S. M., S. R. Schuler, and A. P. Riley (1996). Rural credit programs and women's empowerment in bangladesh. World development 24(4), 635–653.

HIES (2019). Household income and expenditure survey 2016-2017. Technical report, Bangladesh Bureau of Statistics.

Jakiela, P. and O. Ozier (2016). Does africa need a rotten kin theorem? experimental evidence from village economies. The Review of Economic Studies 83(1), 231–268.

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

Jayachandran, S. (2021). Social norms as a barrier to women's employment in developing countries. IMF Economic Review 69(3), 576–595.

Lee, J. N., J. Morduch, S. Ravindran, A. Shonchoy, and H. Zaman (2021). Poverty and migration in the digital age: Experimental evidence on mobile banking in bangladesh. American Economic Journal: Applied Economics 13(1), 38–71.

Mani, A. (2020). Mine, yours or ours? the efficiency of household investment decisions: An experimental approach. The World Bank Economic Review 34(3), 575–596.

McKelway, M. (2021). Women's employment in india: Intra-household and intra-personal constraints. Working Paper.

Montoya, A. M., E. Parrado, A. Solís, and R. Undurraga (2020). Bad taste: gender discrimination in the consumer credit market.

Raven, J. (1965). Coloured progressive matrices : sets A, AB, B. London : H.K. Lewis.

Raven, J. and J. Raven (2008). Uses and abuses of intelligence: Studies advancing Spearman and Raven's quest for non-arbitrary metrics. Royal Fireworks Press.

Riley, E. (2022). Resisting social pressure in the household using mobile money: Experimental evidence on microenterprise investment in uganda.

Roy, S., J. Ara, N. Das, and A. R. Quisumbing (2015). "flypaper effects" in transfers targeted to women: Evidence from brac's "targeting the ultra poor" program in bangladesh. Journal of Development Economics 117, 1–19.

Schaner, S. (2015). Do opposites detract? intrahousehold preference heterogeneity and inefficient strategic savings. American Economic Journal: Applied Economics 2(7), 135–174.

Schaner, S. (2017). The cost of convenience? transaction costs, bargaining power, and savings account use in kenya. Journal of Human Resources 52(4), 919–945.

Schwab, B., S. Janzen, N. P. Magnan, and W. M. Thompson (2020). Constructing a summary index using the standardized inverse-covariance weighted average of indicators. The Stata Journal 20(4), 952–964.

Shanahan, M. (2022). The mobile gender gap report 2022. GSMA Association: London, UK.

Solotaroff, J. L., A. Kotikula, T. Lonnberg, S. Ali, and F. Jahan (2019). Voices to choices: Bangladesh's journey in women's economic empowerment. World Bank Publications.

World-Economic-Forum (2022). Global gender gap report. World Economic Forum, Geneva (Switzerland).

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

WVS (2020). World values survey: Round seven – country pooled datafile. doi.org/10.14281/18241.1. Technical report, JD Systems Institute WVSA Secretariat.

A Appendix

A.1 Additional Tables and Figures

Table A1: Comparison of characteristics for women with a Raven score (from the recruitment survey) just above and just below the district average.

	Just below		Just above		T-test
Variables	Mean	(Std.)	Mean	(Std.)	Diff.
Age	32.18	(0.45)	30.96	(0.36)	1.22**
Age at marriage	16.27	(0.21)	16.26	(0.16)	0.01
Marriage length, in years	16.04	(0.50)	14.84	(0.39)	1.21**
Completed education	1.93	(0.08)	2.13	(0.06)	-0.20**
Answered correctly to financial literacy question	0.56	(0.03)	0.50	(0.03)	0.06
Home maker	0.91	(0.02)	0.93	(0.01)	-0.02
Housework, daily hours	5.51	(0.13)	5.30	(0.10)	0.21
Childcare, daily hours	1.31	(0.13)	1.49	(0.11)	-0.18
Has at least an account (bank, saving, mm)	0.65	(0.03)	0.66	(0.02)	-0.01
Modern lifestyle	0.26	(0.03)	0.29	(0.02)	-0.03
Mother ever worked for pay	0.12	(0.02)	0.12	(0.02)	0.00
Agrees women should be allowed work outside for pay	0.76	(0.02)	0.77	(0.02)	-0.01
Husband controls money for groceries	0.57	(0.03)	0.63	(0.02)	0.06
Husband controls money for clothes	0.41	(0.03)	0.46	(0.03)	-0.05
Husband controls savings	0.32	(0.03)	0.35	(0.02)	-0.03
Number of decisions she is fully involved	1.51	(0.17)	1.50	(0.13)	0.01
Nr of places she went to, past 3 months	2.24	(0.07)	1.98	(0.06)	0.26***
Took public transport at least once, past 3 months	0.86	(0.02)	0.82	(0.02)	0.04
F-test of joint significance (p-value)	0.229				
Number of observations	263		403		

The value displayed for t-tests are the differences in the means between women with a Raven test score just below (3 correct answers) and just above (4 correct answers) the district average. Those above average are selected to take part in the lab-experiment. Regressions for the t-test include village fixed effects. Education categories: no schooling; primary; above primary. Household decisions include: children education; children marriage; children health; food expenditure; whether the husband should start a new activity; whether the wife should work outside for pay; wife's earnings; large household expenditures; take-up loans; use of loan money; livestock and their products; land; fertility; leisure. The five places the wife went to include: local bazar; meet a friend; parents' home; NGO or MFI; local health facility; public place in a different union. *$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A1: Self-reported involvement in household's decisions, by item. The sample ncludes women who took part in the lab experiment.

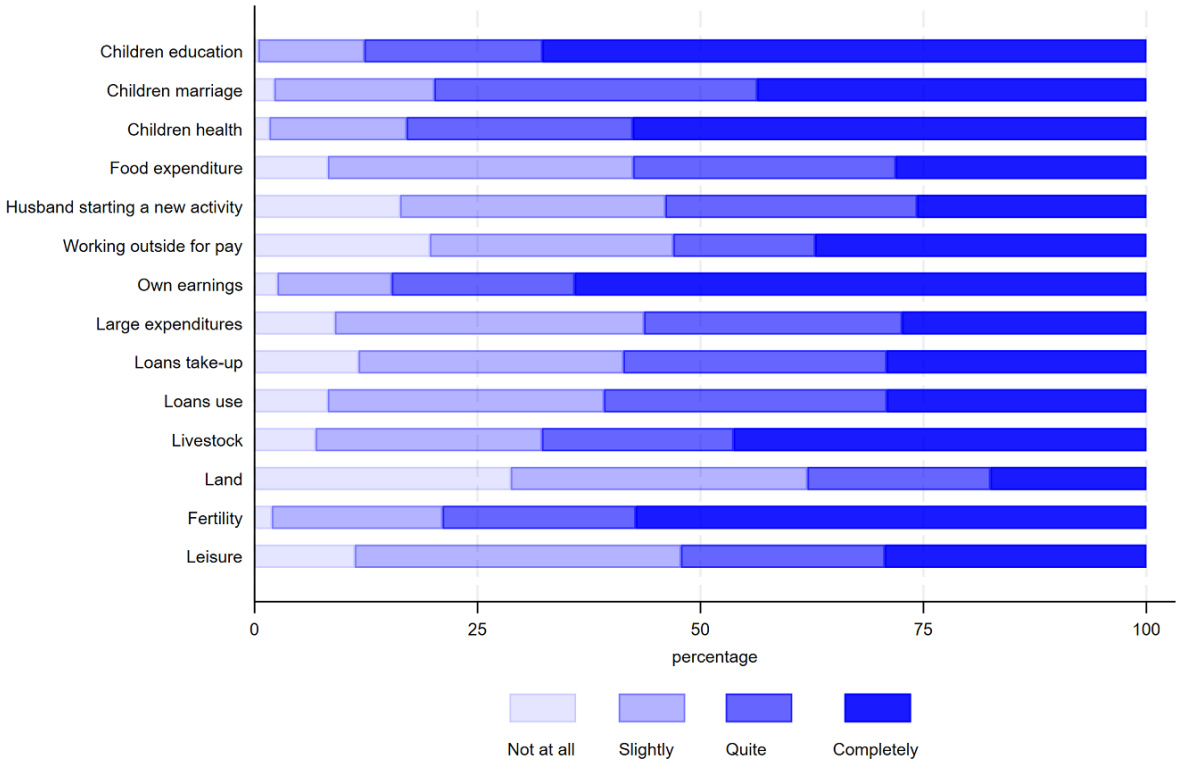


Figure A2: Gender preferences over professions. Percentages by gender

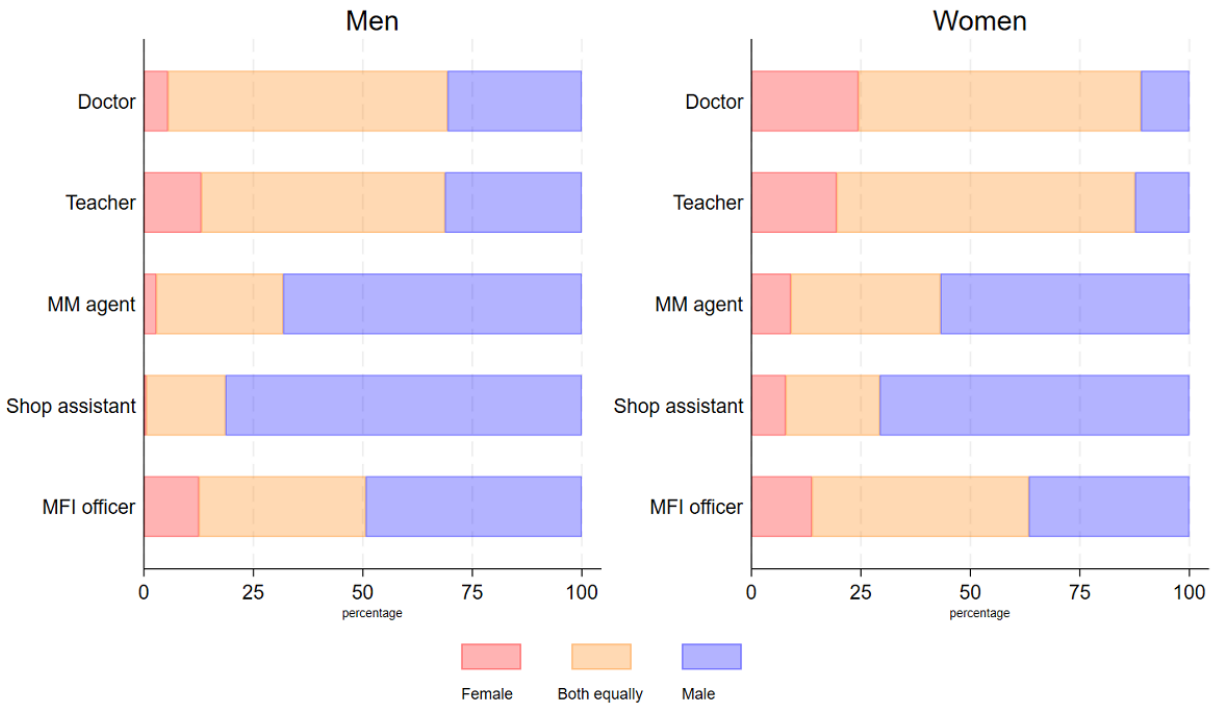


Table A2: Statements on gender attitudes, by gender

	Mean	(Std)			T-test
	Men		Women		Diff.
a) Women are discriminated against in all areas of life, including job opportunities	0.77	(0.02)	0.91	(0.01)	-0.14***
b) Women's responsibilities to family don't leave time for working	0.77	(0.02)	0.84	(0.02)	-0.07***
c) Women make as good employees as men	0.84	(0.02)	0.92	(0.01)	-0.08***
d) It is important for men and women to share household work equally	0.90	(0.01)	0.94	(0.01)	-0.03*
e) Men and women should get equal opportunities in all spheres of life	0.84	(0.02)	0.95	(0.01)	-0.11***
f) Daughters should have a similar right to inherited property as sons	0.70	(0.02)	0.92	(0.01)	-0.23***
g) Wives should be less educated than their husbands	0.58	(0.02)	0.52	(0.02)	0.06*
h) Sons should get more opportunities/resources for education than daughters	0.66	(0.02)	0.50	(0.02)	0.15***
i) Fewer women than men have the skills required to work for pay.	0.85	(0.02)	0.82	(0.02)	0.03
j) A man should be the main provider of their family	0.80	(0.02)	0.73	(0.02)	0.08***
k) When a mother works for pay, the children suffer	0.75	(0.02)	0.67	(0.02)	0.08***
l) If a woman earns more than her husband, it is almost certain to cause problems	0.87	(0.01)	0.89	(0.01)	-0.02
Observations	525				

The table reports the difference in means for men and women. The measures equal 1 if the respondent agrees and 0 if they disagree.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A3: Size of the skill rating gaps, by skill type and gender.

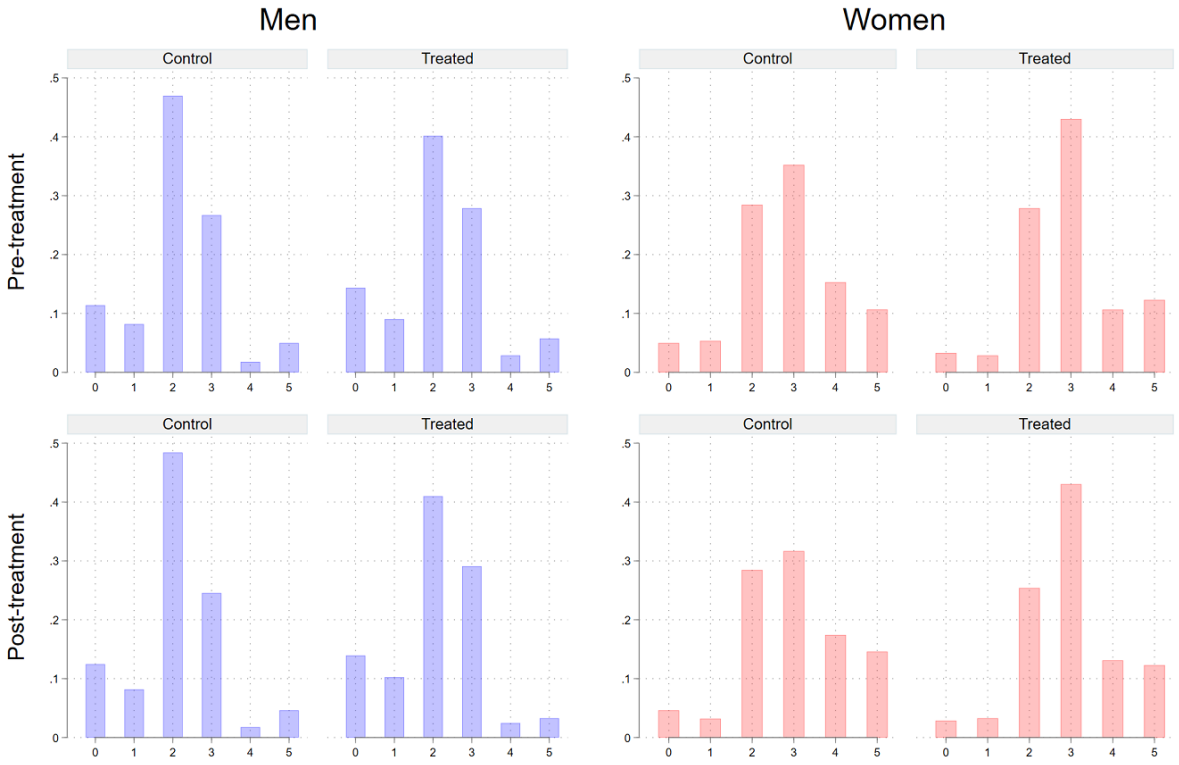
Note: The gaps correspond to the difference between the rating given to the husband's skills and the rating given to the wife's on a scale from 0 to 10. A positive gap indicates that the husband's skills are rated higher. The differences in the average gap sizes are statistically different by gender only for problem-solving and money management.

Table A3: Summary statistics of game outcomes. Pre-treatment round, by gender and treatment status

	Mean / (Std.)					
	Men			**Women**		
	Treated	Control	T-test diff.	Treated	Control	T-test diff.
Belief gender gap (size)	0.35 (1.29)	0.39 (1.23)	-0.04	-0.05 (1.07)	0.06 (1.21)	-0.11
Belief gender gap is positive (%)	0.33 (0.47)	0.37 (0.48)	-0.04	0.22 (0.41)	0.25 (0.43)	0.03
Belief on wife's score	6.71 (1.44)	6.70 (1.43)	0.01	7.35 (1.08)	7.20 (1.14)	0.15
Belief on husband's score	7.07 (1.27)	7.09 (1.18)	-0.02	7.30 (1.12)	7.27 (1.19)	0.03
Nr high-payoffs to the wife	2.13 (1.25)	2.14 (1.14)	-0.01	2.92 (1.13)	2.83 (1.23)	0.09
At least 3 high-payoffs to the wife (%)	0.36 (0.47)	0.33 (0.48)	0.03	0.66 (0.49)	0.61 (0.47)	0.05
	Treated Couples	Control Couples	T-test Diff			
Actual score gap	-0.05 (2.25)	-0.16 (2.37)	0.11			
Observations	244	281				

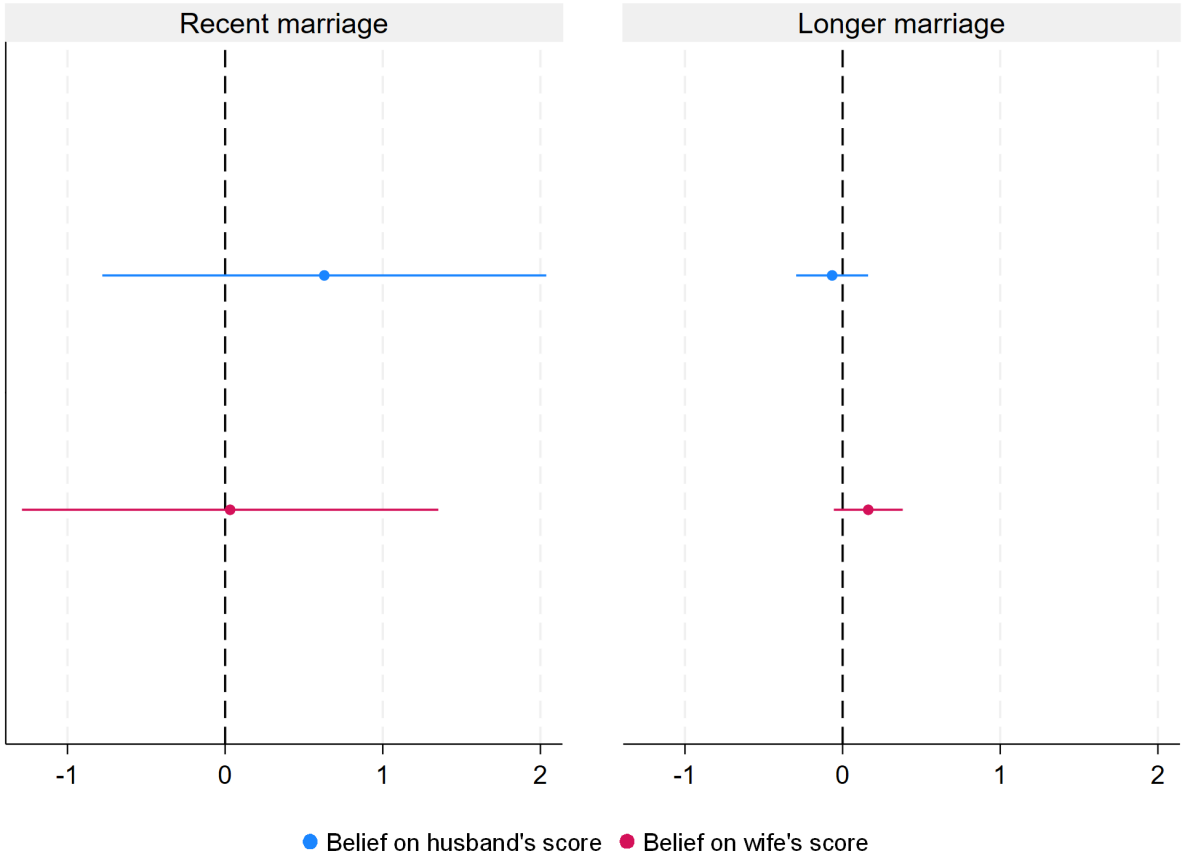
The value displayed for t-tests are the differences in the means between treated and control, calculated separately for men and women. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Beliefs range from 0 to 8. The belief gap ranges from -8 to +8 and is calculated as the difference between the guessed score of the husband and of the wife. A positive belief gap indicates the individuals considers the husband more skilled than the wife in the game. The number of high-payoffs allocated to the wife ranges from 0 to 5. *At least 3 high-payoffs to the wife* equals 1 if the respondent assigned 3 ore more high-payoffs to the wife. The actual gap is computed at the couple level and corresponds to the difference between the number of correct answers of the husband and of the wife.

Figure A4: High-payoffs allocated to the wife before and after the treatment, by gender



Note: The y-axis reports percentages; the x-axis reports the number of high-payoffs allocated.

Figure A5: Effects on beliefs about performances in the game. OLS, women sample, by marriage length.

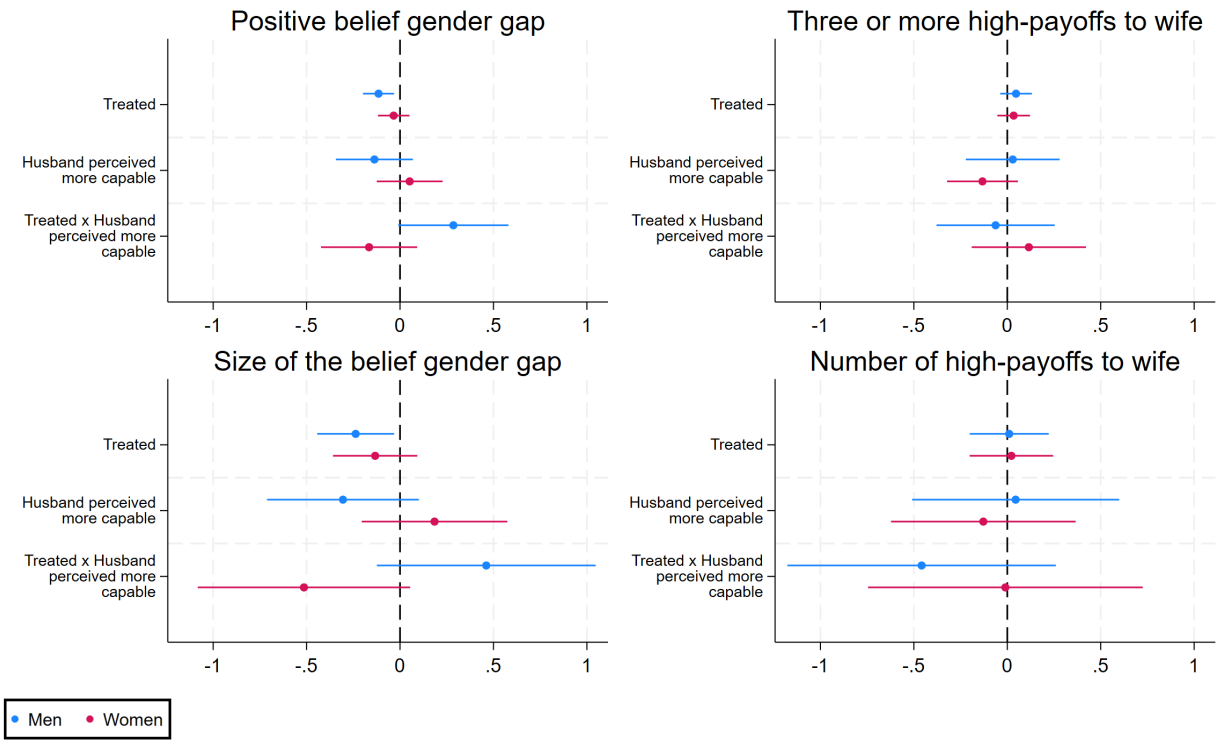


Note: Each regression is run separately for women in shorter (five years or less, left panel) and longer (more than 5 years, right panel) marriages. The dependent variables are the belief on the wife's score in red, and beliefs on the husband's score in blue. 9?The coefficients reported are those for treatment status (=1 if treated). All regressions include stratification and control variables. Stratification variables: wife completed education, marriage length, village of residence. Control variables: spouses' age gap, husband completed education, marriage length squared. Standard errors are clustered at the couple level.

Table A4: Correlation between the number of correct matrices answered by the wife and by the husband in each round of the game

		Husband's score	
		Round 1	Round 2
Wife's score	Round 1	0.11***	
	Round 2		0.10***

Figure A6: Effects based on perceived abilities in care tasks. OLS, by gender.



Note: Each regression is run separately for men and women. All regressions include an interaction term with a binary variable that equals one if the respondent perceives the husband as more capable than the wife at care tasks (housework and childcare), stratification and control variables. Stratification variables: wife completed education, marriage length, village of residence. Control variables: spouses' age gap, husband completed education, marriage length squared. Standard errors are clustered at the couple level.

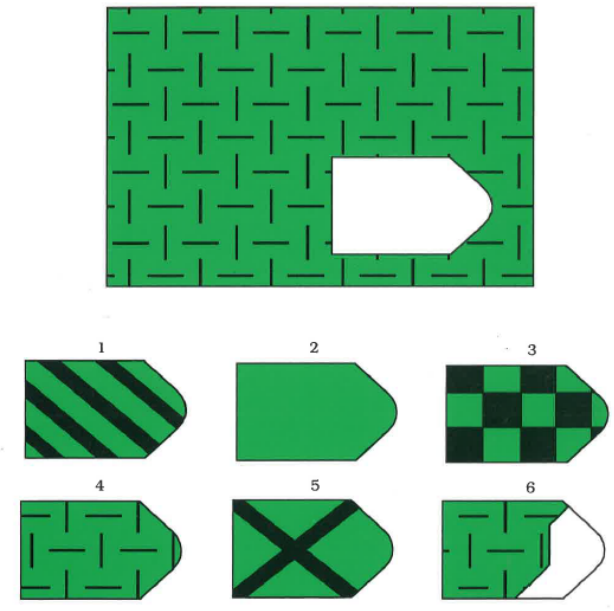
A.2 The Raven progressive matrices

The Raven standardized progressive matrices are a psychological tool used to measure cognitive abilities. It is used in the economics literature to measure abstract logical thinking and entrepreneurial skills (De Mel et al., 2009b). The usual test consists of 5 sets of 12 matrices each. Each matrix contains a geometric pattern with a missing piece. Respondents should choose among six to eight options, which one fills in the design. Only one answer is correct and the difficulty of the patterns increases as the test proceeds (Domino and Domino, 2006, Raven and Raven, 2008).

In the incentivized decision-making game, I use 24 questions (8 per round) from the A, B and AB sets of the Raven's Coloured Progressive Matrices (RCPM).[15] Respondents have one minute to answer each question. If they do not know the answer or do not want to answer, they can let the time pass and the tablet proceeds automatically to the next matrix.

[15]The RCPM is a coloured version of the sets A and B of the standardized matrices, and a third set of easy items interposed between them (set AB).

Figure A7: Easy matrix from the Raven Colored Progressive Matrices (RCPM) by Raven (1965)



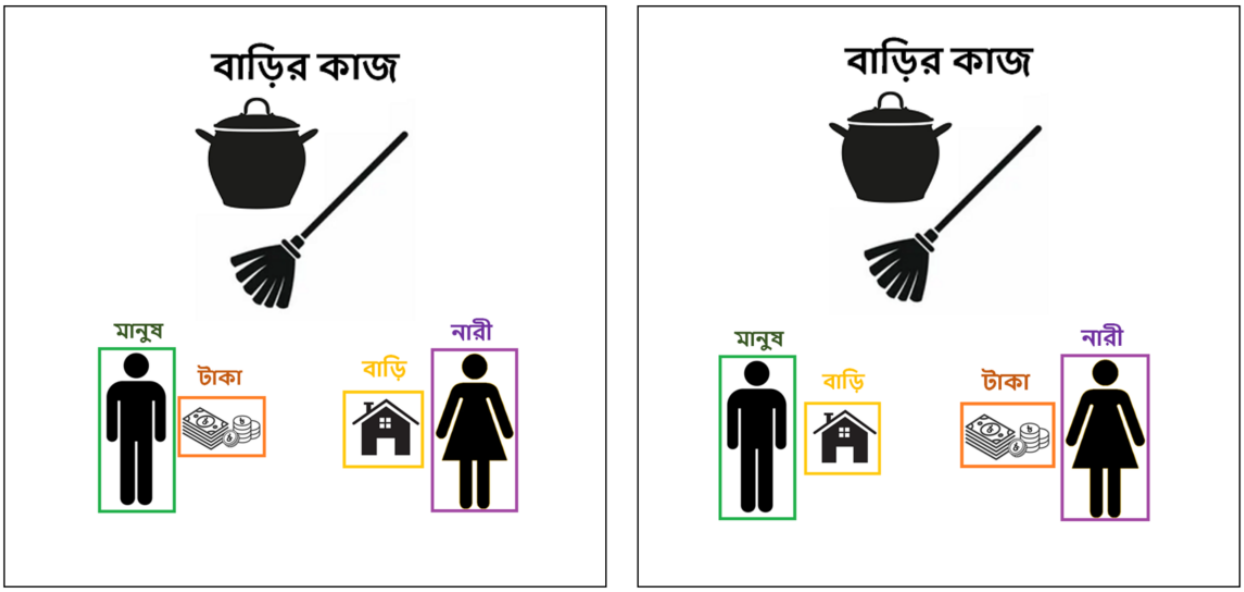
A.3 Non-verbal Implicit Association Test (NV-IAT)

To measure gender stereotypes while taking into account the low digital literacy of respondents living in rural areas, I design a non-verbal implicit association test (IAT).[16]

The IAT is finalized based on the evidence collected during a pilot day in the field and therefore, it differs from the usual IAT in a few features. First, it is composed of two repetitions only. Figure A8 provides an example of how the IAT questions appear on the tablet in the two repetitions. In the first one, respondents have to associate the picture of *Housework* with either *Male-Money* to the left or *Female-Home* to the right. In the second one, the options are *Male-Home* to the left or *Female-Money* to the right.

[16] *"This test is a computer-based tool developed by social psychologists and has recently been used by economists studying discrimination in the context of gender and race bias. The test exploits the reaction time to associations between male or female names and scientific or humanistic fields. The underlying assumption is that responses are faster and more accurate when gender and field subjects are more closely associated by the individual"* Carlana (2019)

Figure A8: Example of non-verbal IAT question



Note: In the panel to the left, the respondent should associate the image on the top with the categories *Male-Money* to the left or *Female-Home* to the right. In the panel to the right, the categories are *Male-Home* to the left or *Female-Money* to the right.

Usually, IAT questions proceed only if the respondent gives the right answer and the final score is calculated using the response time. During the pilot day, respondents appeared confused between the IAT instructions and the Raven test. The fact that questions in the Raven test proceed irrespectively of whether the answer is correct, gave respondents a false sense of their performance that biased their belief elicitation. Therefore, questions in the NV-IAT proceed even if the association is wrong.

This implies that the way the score is calculated changes compared to the usual test. Each association was counted as follows:

- 1 point if the association is wrong and the response time is below the median.
- 2 points if the association is wrong and the response time is above the median.
- 3 points if the association is correct and the response time is above the median.
- 4 points if the association is correct and the response time is below the median.

I use the total points of the respondents to compute a standardized weighted score as in Schwab et al. (2020). A higher score indicates stronger stereotypes. Table A5 shows the correlation table between the NV-IAT score and the self-reported measures of gender attitudes and beliefs. The NV-IAT score is positively correlated with conservative gender attitudes, the number of professions for which the respondent prefers a man (significant at the 1 percent level) and the number of anti-women statements the respondent agrees with. It is negatively correlated with the number of

pro-women statements the respondent agrees with (significant at the 1 percent level). This means that stronger implicit gender stereotypes are associated with conservative gender attitudes. Nevertheless, because of the differences in the test design compared to the usual IAT, this measure of gender stereotypes should be interpreted with caution.

Table A5: Correlation between the non-verbal IAT score and self-reported measures of gender attitudes

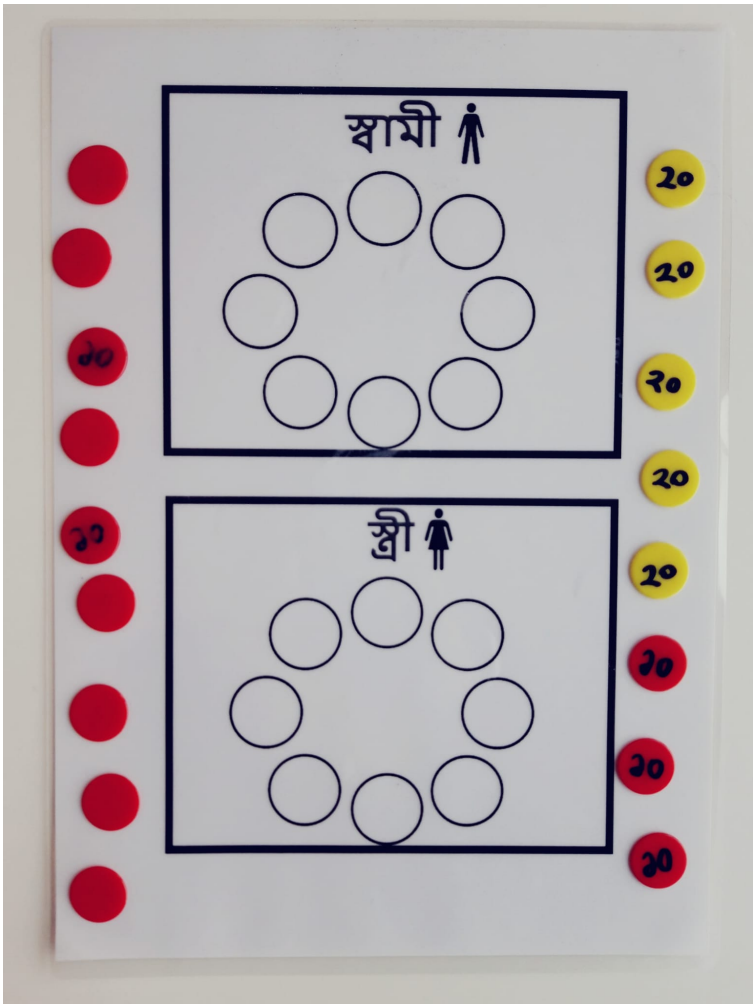
Self-reported measures	Non-verbal IAT score
Gender attitudes index	0.06***
Prefers men in at least 3 professions out of 5	0.11***
Nr pro-women statements agreed out of 6	-0.09***
Nr anti-women statements agreed out of 6	0.02

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The descriptive statistics for each of the statements used is presented in Table A2. The descriptive statistics for the gender attitudes index and the non-verbal IAT are reported, separately by gender, in Table 2

A.4 Game instruments

To help respondents visualize the payoffs, I provide them with a laminated board and 16 coloured tokens. The board is divided into two panels. The top panel contains the payoffs the respondent would like to allocate to the husband. The bottom panel is for the payoffs to be allocated to the wife. The red tokens represent the low-payoff matrices (10 points), and the yellow tokens the high-payoff matrices (20 points). Each token has its value written with Bangladeshi numbers on both sides. I ask respondents to keep all the tokens in front of them and choose one by one by placing them in the circles belonging to the husband or the wife's panels. Figure A9 presents a picture of the board and of the tokens used during the game.

Figure A9: Board and tokens used by players for the allocation decisions



Instructions

In this activity, you and your spouse are partners. The game is going to be repeated twice. In each repetition, you and [SPOUSE NAME] will take some decisions and then answer 8 questions. When you answer correctly, you can earn some BDT. Some questions are worth more (20 BDT) and some are worth less (10 BDT). The amount of BDT does not represent the difficulty of the questions: some easy questions are worth more BDT, and some hard questions are worth fewer BDT. When you answer a question you do not know how many BDT it is worth. Remember that to earn BDT you must answer correctly to the questions and that your final earnings will depend on the correct answers of you and [SPOUSE NAME] together.

In the end, the BDT you obtained and the BDT that [SPOUSE NAME] obtained by answering correctly, are going to be added together and then split equally between the two of you. Please notice that you will not be able to know how many BDT you earned alone or how many BDT

your spouse earned alone. This means she will also not be able to know. This tablet will do the assignment and the calculations of the BDT. You will only know how many BDT you two earned together, and how many you get in the end, after the earnings are redistributed equally.

Before answering the questions, I am going to give you these tokens *[Enumerator show the bag with the tokens]*. In total, there are 16 tokens: 11 red and 5 yellow. The yellow tokens correspond to a higher amount of BDT for each correct answer (they are worth 20 BDT) and the red ones correspond to a lower amount of BDT for each correct answer (they are worth 10 BDT). I am going to ask you to choose 8 tokens for yourself and 8 for [SPOUSE NAME] and put them on these boards in the corresponding spots [Show the laminated board with the circles]. The tokens you put on these circles [the husband board] represent the BDT that you can earn if you answer correctly to your questions, the tokens you put on these other circles *[the spouse board]*, are for [SPOUSE NAME] and they represent the BDT that she can earn if she answers correctly to her questions. You can freely pick any combination that you think is the best, as long as you choose 8 tokens for you and 8 for her. You must choose exactly 8 tokens for each participant, you cannot pick more or less than 8 tokens. Remember that the value of the questions does not correspond to their level of difficulty.

In the meantime, your spouse will be asked to do the same task: my colleague will give her a bag of token and will ask her to pick 8 for herself and 8 for you. Your decisions will remain private to your spouse, you cannot know what she chose and she cannot know what you chose. When both of you have taken your decisions, we will randomly select one to implement in the game. We might pick your choice of tokens or her choice of tokens.
Please, notice that you will not be allowed to communicate spouse [SPOUSE NAME] while you take these decisions.

Is everything clear so far? I know this is a lot of information to remember. Do not worry, I cannot take decisions for you or help take them, but I will guide you step by step throughout the activity.

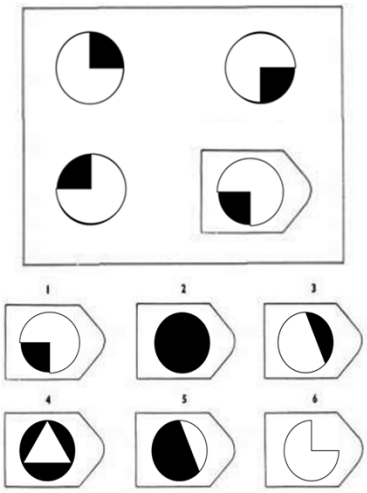
Once all the payoffs have been allocated and the decisions are taken, you will start answering the questions. You can also decide not to answer a question. In this case, you will not earn any BDT for it.

When you answered all the questions, we are going to have a little break, and then you will have to do the exact same activity one more time.

As I said before, the BDT you obtained and the BDT that [SPOUSE NAME] obtained by answering correctly, are going to be added together and then split equally between the two of you.

At this point, you will have the chance to double or halve your share of BDT. In this bag, there are five pieces of paper. On each of them is written either 0.5, 1, 1, 1 or 2. *[Enumerator, show the bag and the papers]* You will have to extract one of them without looking. If you extract the number 2, we will DOUBLE your BDT. If you extract the number 1, they will remain the same, If you extract 0.5, we will HALVE your BDT. Irrespectively of what you drew, you will still get the 100 BDT fee to thank you for taking part in this activity. You will receive your payment today, at the end of the activity, before leaving this room.
Do you have any questions?

This is an example of the questions you are going to answer. *[Enumerator, show the tablet to the participant with the example question. Please do not rush while you explain the example]*: as you can see, there is a box with a picture in it, and with a piece cut out of the picture. The picture follows a pattern from left to right and top to bottom. Look at the picture in the box, and think what the missing piece must be like to complete the pattern correctly both across (left to right) and down (top to bottom). Find the right piece out of the six pieces shown below the picture. Only one of these pieces is perfectly correct.
Please, identify which piece is the one that is missing, and tell me your choice.



[Enumerator, show the tablet to the participant and ask him to select one answer. Then take the tablet back and explain:]
In this case, the correct answer is (1) as you can see from the position of the black quarter.
[Enumerator show how the position of the black quarter completes the pattern. If the participant did not understand the example, please explain it again. Then, hand the tablet to play more practice questions]

I am going to ask you a couple of questions to make sure everything is clear. If you do not

know how to answer let me know and I will explain again.

1. Could you tell me what do the tokens represent? If the respondent does not say it explicitly: how much are the yellow tokens worth? How much are the red tokens worth?
   Answer: the tokens represent how much each question is worth. However, the value of a question does not correspond to its difficulty. The yellow tokens are worth more (20 BDT), the red tokens are worth less (10 BDT).

2. Could you tell me what decision is your spouse taking?
   Answer: she is taking the same decisions as him. In the end only one of the allocations is randomly chosen for implementations, either his or hers. He cannot know what she chose.

3. Can you know how much a question is worth when you answer it?
   Answer: No.

4. At the end, the BDT you and your spouse earned are pooled and then redistributed equally. Is this true or false?
   Answer: True. The BDT earned by the husband are summed to the BDT earned by the spouse and then divided in equal shares between the two.

5. Can you know how much your spouse earned by herself?
   Answer: no. The same applies to her, she cannot know how much the husband earned by himself.

This is the last step before we begin. I am going to ask you to perform two exercises to practice. Feel free to ask me any question you may have. Even if this is just to practice, I would like you to behave as if it was already the real activity.

First, could you allocate 8 questions to you and 8 to [SPOUSE NAME] by choosing the tokens from this box and put them on the boards as I showed you before. Remember, the yellow ones are worth more BDT than the red ones but BDT amounts do not reflect the difficulty of the questions. Enumerator, hand in the boards and the tokens.

Thank you. In this scenario, you picked XX red tokens worth 10 BDT and YY yellow tokens worth 20 BDT for you. You picked ZZ red tokens worth 10 BDT and WW yellow tokens worth 20 BDT for [SPOUSE NAME]. Is this clear? Do you have any question? This is a practice round so please feel free to ask me any question you may have.

Please notice that throughout the activity you will see me using my phone. This is because I have communicate with my colleague that is sitting with [SPOUSE NAME].