Stifled by Stigma? Experimental Effects of Updating Husbands' Beliefs on Participation in Household Work[*]

Tamara J. McGavock,[a] Nicholas Magnan,[b]
Ellen McCullough.[c] Thomas Woldu Assefa[d]

January 2025

PRELIMINARY DRAFT - Please do not circulate without permission

Abstract

Women disproportionately shoulder unpaid care and domestic work, a disparity rooted in persistent social norms about gender roles across cultures, income levels, and labor force participation patterns. In our rural Ethiopia sample, over 90% of men personally find it acceptable for men to engage in these tasks. However, many misperceive their peers' beliefs, and almost none participate in chores or childcare at baseline. Our hour-long information session corrected these misperceptions by revealing high peer acceptance of men's involvement in tasks like collecting firewood, caring for children, and doing laundry. Importantly, men also learned that their peers received the same information. Following the intervention, men in the treatment group significantly increased participation in childcare and laundry compared to the control group. While childcare participation effects persisted a year later, the effect on laundry tasks was short-lived.

1 Introduction

Women do much more household work and childcare activities than men (Matulevich & Viollaz, 2019), especially in developing countries (Jayachandran, 2015; Duflo, 2012). This results in excessive time burden and dis-empowerment for women (Jayachandran, 2021), as well as economic inefficiency from the mis-allocation of human resources (Hsieh et al., 2019). In developing countries the time burden is especially strong because there are few opportunities to use technology or markets to complete household tasks (?Jayachandran, 2021).

While norms that limit women's role in society date back centuries (Alesina et al., 2013; Hansen et al., 2015) and are passed on from generation to generation (?), they are mutable. Family composition (Washington, 2008; Healy & Malhotra, 2013; Oswald & Powdthavee, 2010); teachers and government officials (Carlana, 2019); exposure to working women in the family, workplace, or media (Jensen & Oster, 2009; Ferrara et al., 2012; Fernández et al., 2004; Dahl et al., 2021); and technological and medical innovation (Greenwood et al., 2005) have all brought changes in gender norms over time. In some cases, private attitudes about the role of women are more progressive than societal norms would suggest. In Saudi Arabia, Bursztyn et al. (2020) finds that men are generally supportive of their wives working outside the home, but believe their peers are not. They test an intervention that corrects these mis-perceptions, and find that when men are informed of what their peers actually privately believe, they become more likely to allow their wives to enter the labor force.

In this study we use an intervention similar to Bursztyn et al. (2020) (BGY, hereafter) to test whether men's own behavior is malleabe in response to correcting misperceptions about gender norms among rural, Ethiopian men about doing chores generally thought of as women's work: gathering firewood, washing clothes, and childcare. Our experiment is designed to answer the following questions. Do men correctly perceive the gender attitudes of their peers? If not, can correcting mis-perceptions lead them to increase their participation in household labor? Importantly, we seek merely to determine whether a mismatch between private beliefs and behaviors stems from incorrect perceptions about the beliefs of others.

Our paper contributes to the literature on interventions that specifically target gender norms and behavior related to gender norms. Pulerwitz et al. (2015) find that community and small

group education sessions held over several months led to more progressive gender norms and less reported inter-partner violence among young, married Ethiopian men. Mckelway et al. (2019) find that showing family members videos about an employment opportunity decreased opposition to women working outside the home in India, whereas Dean & Jayachandran (2019) find that having conversations with family members about women working outside the home had no effect. Field et al. (2021) find that giving women access to savings accounts increased their labor supply and changed gender norms regarding women working outside the home. Especially relevant to this paper is the work of Dhar et al. (2018), who find that classroom discussions with adolescent boys and girls in India over the course of two years caused lasting progressive changes in gender attitudes among both boys and girls. The intervention also changed boys' *perceptions* of other boys' gender norms and increased their likelihood of doing household chores and encouraging their sisters to pursue college.

There is an expansive literature on using information on others' *actions* to encourage behavior change. This includes information on peers' voting participation (Gerber & Rogers, 2009), water and energy use (Allcott, 2011; Ferraro & Price, 2013; Brent et al., 2015; Byrne et al., 2018), saving (Beshears et al., 2015), and charitable giving (Frey & Meier, 2004). This paper contributes to a related and more recent literature that examines the impacts of providing information on others' *perceptions* of what is socially acceptable or desirable behavior. d'Adda et al. (2020) finds that Italian students participating in a dictator game are very responsive to what others believe to be "socially acceptable" gift. Delfino (2021) attempts to alter perceptions about the share of male social workers in the United Kingdom to encourage them to apply for positions in this field.[1] She finds that showing potential applicants a photo of a male social worker increased perceptions of how prevalent male social workers are, but did not lead to an increase in male applicants. Allen et al. (2021) provides information about public preferences for social distancing in Mozambique during the COVID-19 pandemic. They find that this information decreases social distancing where cases are low and increases it where cases are high.

In many of the cases above, information provided in the intervention is intended to induce behavior that is incongruent with what people want to do, either by causing shame or a desirability

---

[1]Delfino (2021) proposes that underestimating the number of men in the profession could deter men from applying because they fear social stigma, or because they may not enjoy working in an environment with so many women.

to conform. For instance, someone may personally prefer to set their air conditioning very high, but decide to turn it down once learning that their neighbors use less electricity. While an individual using lots of electricity may be environmentally damaging, it is not necessarily Pareto inefficient. In other cases, information can allow people to behave in a way personally want to, but do not think is socially acceptable. The collective mis-perception of social norms that leads people to act on incorrect perceptions rather than their own attitudes has been termed the "illusion of universiality" (Allport, 1924), or "pluralistic ignorance" (Katz et al., 1931). The intervention created by Bursztyn et al. (2020) takes on pluralistic ignorance directly by gathering and distributing information about private attitudes in an effort to change behavior rooted in mis-perceived gender norms. Eighty-seven percent of the Saudi men in this study personally think that their wives should be able to work outside the home. However, working outside the home was very uncommon. The experiment's results demonstrate that the mismatch between their beliefs and behaviors stems from misperceiving what other men believe. Correcting these misperceptions can result in a Pareto improvement; the Saudi women are happier being able to work outside the home (as evidenced by their decision to do so) and the Saudi men are at least as happy, if not happier, from having their wives work outside the home. The job seekers in Delfino (2021) may want to become social workers, but not do so because they wrongly believe others will not find this career socially acceptable for a man. The respondents in Allen et al. (2021) may want to engage in social distancing to avoid catching or spreading COVID-19, but not do so because they think others believe it is unnecessary (or the opposite).

This paper is unique among interventions intended to change men's gender attitudes because it focuses on them taking action (doing household labor) rather than granting permission to their wives or other women in their lives. It is unclear how targeting men's attitudes about their own actions, as opposed to about what women should be permitted to do, will affect women. Information that changes men's attitudes about what women can do could have a strong impact if women are eager to act, and are only being held back by their husbands' regressive attitudes. For instance, many of the Saudi women in Bursztyn et al. (2020) were eager to look for a job and took action as soon as their husbands received information about social norms. However, in these cases the information and behavior change are de-coupled in that one person receives the information and the other changes their behavior. An intervention that targets men's attitudes about their own

4

actions may induce a stronger response because the action is taken by the same person who receives the information.

The rest of the paper proceeds as follows. In Section 2 we describe our study setting and sample. In Section 3 we describe our experimental intervention. Section 4 describes our data and Section 5 presents our empirical model. In Section 6 we present our results, heterogeneity, and mechanisms. We discuss the implications of our findings in Section 7.

2   Study setting and sample

Our study takes place in rural Amhara, Ethiopia, where local norms are highly patriarchal and households are quite poor. These characteristics are important to note for context, though millions of people around the world across multiple continents and even in middle- and high-income countries can be characterized similarly. Repressive gender norms and economic under-development are mutually reinforcing — with clear evidence showing both that lifting gender norms can promote economic development and that economic development can reduce gender inequality (Duflo, 2012). In a setting like Ethiopia, where societal norms are especially patriarchal, 60% of women's total labor supplied is for unpaid household work (compared to only 29% for men) (Charmes, 2019). Given this discrepancy, we would expect to see strong economic returns to interventions that enable women to participate more in economic activities.

Deeply entrenched beliefs about gender roles and relations are pervasive in our study region, with profound implications for the opportunities women have to engage in livelihoods activities and the quality of childcare provided to children. For example, traditionally, husbands walk in front of wives (not side-by-side), and husbands and wives do not call each other by their names. With respect to childcare, during focus group sessions and scoping interviews we were told that according to traditional norms it is better for a father to leave a hungry child crying than for a man to reach into the area where leftover food is kept to feed the child a snack. Similarly, women and men would tell us that husbands would participate in firewood or water collection only if they could use a donkey to do the carrying—otherwise these physically demanding tasks were the job of a wife.

Our study's households are all chronically poor, and beneficiaries of Ethiopia's Productive Safety

Net Program (PSNP) as well as livelihoods training via a graduation program (implemented either by World Vision or CARE Ethiopia). The programs provide not only short-term food security support but also implement training programs aiming to equip them to "graduate" from extreme poverty and, eventually, government safety net programs. The programs include gender-sensitive training for both men and women, aimed at empowering women in their households and community.

Participating households in the same village join the savings groups called Village Economic and Savings Associations (VESAs), and our experiment took place at the regular place where a VESA met on a bi-weekly or monthly basis to discuss and receive training on a variety of subjects. We randomly sampled 60 VESAs for our study with an intended sample size of 900 households after recruiting all households represented in the sampled VESAs (most are husband-wife pairs). We were ultimately able to complete only 75% of the VESAs before ceasing our field operations due to the spread of conflict from Tigray into our study region, so our final sample is 46 VESAs and 646 households. At a baseline interview with the male and female head of household we invited all men to participate. Eventual participants ranged in age from 16-80, with 90% of participants under the age of 60 and 50% under the age of 43.

3   Intervention

Our experimental intervention is an activity, adapting the laboratory experiment protocol of BGY for our context. We deviate in a few noteworthy ways from their protocol, which we highlight after explaining the intervention and in discussing our results.

3.1   Activity Protocol

The activity proceeds in two stages, a beliefs gathering stage and an information treatment stage, separated only by a few moments during which the group is split up according to treatment assignment. The full protocol is described in the Online Appendix. Participants gather at their usual VESA group's meeting place for the activity, enumerators take attendance, and they find a seat while maintaining some meters of distance from others. At the start of the activity, an informed consent script is read and each participant receives an opaque bag of supplies.

In the first stage with all participants together, each man privately reports his own beliefs

(agreement / disagreement with a statement) and makes a guess about the beliefs of the other men who are present. Three of the statements were selected specifically for our treatment: a man reports whether he agrees with the statement "It is acceptable for a man to perform [TASK]" where the tasks are collecting firewood, caring for a young child,[2] and washing his own clothes. To diffuse men's focus from the treatment topics, we alternate these with placebo-like statements about other topics: whether eating eggs helps children in school, whether going to school will make children's futures better, and whether eating orange vegetables will make children's futures better.

To indicate whether or not he agrees with a statement, a man places a black or white chip from his bag into the bag an enumerator brings around corresponding to that question. Each token from his bag bears the participant's ID, and he maintains secrecy by holding the chip in his closed palm as he moves it from his personal supply to the answers bag. After all votes are cast for the yes/no question, enumerators instruct respondents to report what they believe is the number of men in the session who answered in the affirmative. That is, the enumerator asks "How many of the men here do you think answered 'yes' with their black token *to that question*?" Each man indicates his response by circling the corresponding number of male icons on a paper bearing their ID number, and include one for himself if he votes yes. Thus, for the treatment questions about tasks, the respondents records their personal private belief (yes or no) and their perception of the strength of the social norm against men performing the task among their peer group (the share of men they believe answered 'no'). After these responses are collected, the enumerator continues with this procedure for each of the remaining questions, alternating through treatment and placebo questions. Finally, names of men in the control group are called and they split off to find a place to sit far enough away that the groups cannot hear or see each other's activities.

During the second stage, the actual responses of men in the entire group are shared, but only with men pre-assigned to the treatment group. That is, in the treatment group an enumerator reveals the entire group's perceptions about the acceptability of men doing the three tasks by counting out the number of black and white tokens in the corresponding bag. The bag bears a picture of the task and has been in full view for the entire activity since men cast their votes, so that the men in the treatment group receive information in a manner they find credible that either confirms their perception of their peers' acceptance or corrects their mis-perception. To ensure

---

[2]Enumerators are explicit about the childcare.

that the true acceptance rate among their own peers is salient and that men connect it to their guess, the enumerator circles the man icons on a large version of the same response card the men use for reporting their second order beliefs, so that men can see how their card would look if they had guessed correctly. To occupy men in the control group while the treatment group is presented with the information, enumerators ask unrelated questions about the community's infrastructure and local wages until the treatment group has departed.

In both groups, the activity concludes with a third round of voting, each man privately reporting their intention to participate in each task during the 7 days after the activity (or, for the placebo questions, to carry out the corresponding behavior, e.g., feeding eggs to a child). Importantly, we collect the intentions from all men, so that we can compare the average responses in the treatment group to those from the control group to identify a treatment effect on stated intentions. As we discuss in Section ??, this can show a social desirability effect but will not show any spillover effects, since participants have no opportunity to communicate with each other about the activity before submitting these responses. The next section explains the data sources we use to measure treatment effects beyond tallying the men's votes during the activity.

3.2 Differences with BGY Design

Our experimental design differs from the BGY setup in a few ways. First, our experiment seeks to connect information revealed during it directly to behavior that men can take action on themselves. That is, we aim to correct mis-perceptions held by men about the acceptability of *men* performing household work typically done by women, with the hypothesis that we will observe a change in their own behavior. By contrast, BGY corrected a participant's mis-perceptions about peers' approval of women's behavior (work outside the home), thereby interpreting a change in women's work as an indication that updated information affected the permission granting of husbands. Second, participants in our context are well known to one another because they reside in the same village, are from the same socioeconomic group (all are eligible for the safety net program), and are already participating regularly in programming related to savings, livelihoods activities, and even gender norms discussions. In this setting, we might expect men to have good information on each other's beliefs, making the measured misperceptions small. At the same time, to the extent that these misperceptions exist in our context, the treated men are informed not about how other

8

men whom they do not know think, but about their very peer group. Lastly, and relatedly, we note that our intervention not only provides men with information updating their second order beliefs (what other men think) but the delivery of this information is known to all men in the treatment group collectively, and is thus an intervention about third order beliefs as well. Men who receive the updated information not only know the truth but they also know that the other men (in the treatment group) also know the truth. Given the social nature of these gender norms, this is an important channel that some designs might try to shut down, as did BGY by delivering the information treatment privately. But this may also be an important consideration of impact.

4 Data

Our primary outcome is whether a male participant (MP) performs the tasks from the treatment (firewood collection, childcare, and laundry). To limit social desirability biases, the relevant information is elicited during interviews with a linked female respondent (hereafter, FR), not the MP himself, and indirectly, not with a leading question. Specifically, for eight tasks including the three tasks from the experimental activity, the FR responds to a sequence of questions (see the Online Appendix for the instrument, which is also included in the baseline interviews) in the context of a longer interview. First the FR indicates whether the task was performed for the household in the previous 24 hours and, if so, she lists the names of and minutes spent by the top three contributors to each task. Using these data, we construct the primary outcome variables we pre-specified for this study, an indicator for each task indicating whether the MP performed the task during that 24-hour period. We also use variables indicating whether anyone performed the task, whether the FR herself performed the task, and the relative shares (in minutes or in person-tasks) contributed by the FR, by the MP, and by others. In total, we collect these data three times: first, at baseline, prior to the activity with the men; then during the in-person interview between 1 and 7 days after the experimental activity;[3] and finally during a phone-based interview that occurred between 9 and 21 months after the activity. The one-week follow up occurred on a day randomly assigned within

---

[3]During the seven days following the intervention, a linked female respondent (hereafter, FR) from each male participant's (MP's) household was engaged in other survey activities with enumerators. On a pre-programmed but randomly assigned day, an enumerator visited the FR and completed a questionnaire that included the chores module. These days were randomly assigned and stratified on treatment group. For more details, see the Online Appendices for McGavock et al. (2024).

VESA at the household level, stratified on treatment group. The one-year follow-up occurred with a gap that was randomly assigned by village.

Online Appendix Table 4 contains descriptive statistics, with separate means by treatment and control group. Treatment was balanced across all of these variables, with a minimum p-value of 0.155. The men in our study are 43 years old on average, consider themselves to be religiously observant,[4] attend 1.1 VESA meetings in the past week on average, and save an average of 19 Br at those meetings.

Table 1 reports men's participation at baseline in household chores as well as their responses from the activity on the acceptability of men performing these chores and their perception of others' acceptance. Notably, at baseline participation in these tasks is very low, consistent with social norms that they are traditionally the responsibility of women. Less than one-quarter of one percent of respondents had collected firewood in the week preceding the baseline survey, only 12 percent had cared for a child, and essentially none had washed his own clothes.

The table also demonstrates the concept of pluralistic ignorance: despite not doing these tasks, during the activity men reported high rates of acceptance of men doing these tasks but also substantial misperceptions. Approximately 91% of men were accepting of men collecting firewood, 95% were accepting of men caring for children, and 89% were accepting of men washing clothes. On average men's perceptions of others' acceptance of these chores were slightly lower—the average guesses were 80% of their peers, 87%, or 85% for collecting firewood, performing active childcare, or washing their own clothes. While these rates are statistically significantly lower than men's acceptance rates, they are perhaps surprisingly close. However, they vary significantly across men, even in the same village.

On average, men's perceptions of others' acceptance were fairly accurate. For each task, at least one-quarter of men correctly guessed the acceptance rate. For childcare and firewood collection, it was much more common for men to under-estimate acceptance than to over-estimate it. For laundry, the opposite was true. Some mis-perceptions, mostly under-perceptions, were quite extreme. In the analysis that follows we test how the intervention interacts with *ex ante* mis-perceptions about social norms to affect behavior.

---

[4]Respondents were asked to rate their attendance at religious ceremonies, with 1 being highest attendance and 5 being the lowest).

5 Empirical approach

The primary outcomes of interest for this study are binary variables indicating a man's (1) stated intention to participate in a given task (firewood collection, childcare, and laundry), collected immediately after the information reveal at the end of the experimental activity; and (2) participation in the chore during a 24 hour period, collected by speaking with the FR at baseline, during the 7 days following the activity, and during the year following the activity.

We first regress each outcome variable on a single treatment status indicator, controlling for age, village (stratification bin) fixed effects, task fixed effects, and baseline performance of the task, an ANCOVA estimation. We cluster standard errors at the village level because we randomly selected our sample villages from a larger population of villages. We estimate these in a pooled regression with all tasks to ensure that treatment effects across tasks are comparable, and our sample is limited to the set of all households for which all outcomes exist.[5] The empirical specification is:

$$Y_{ivt} = \alpha_1 T_{iv} + \alpha_2 Age_{iv} + \alpha_3 Y^0_{ivt} + \sum_{v=1}^{46} \phi_v + \sum_{v=1}^{3} \lambda_t + \varepsilon_{ivt} \tag{1}$$

In equation 1, i denotes the individual, v denotes the village, and t denotes the task. We also estimate equation 1 with task-specific treatment effects.

Because we expect information to have a different impact if it corrects a mis-perception, we also interact treatment status with whether an individual underestimated his peers' acceptance of a task. The empirical specification becomes:

$$Y_{ivt} = \beta_1 T_{iv} + \beta_2 Under_{ivt} + \beta_3 T_{iv} \times Under_{ivt} + \beta_4 Age_{iv} + \beta_5 Y^0_{ivt} + \sum_{v=1}^{46} \phi_v + \sum_{v=2}^{3} \lambda_t + \varepsilon_{ivt} \tag{2}$$

Here, the estimate of interest is $\beta_1 + \beta_3$, which is the treatment effect on those who underestimate their peers' acceptance of a task. Again, we also estimate equation 2 with task-specific treatment estimates. Finally, we also use a local polynomial analysis to estimate non-parametric treatment effects conditional on misperceptions to unpack the mechanisms, since men's misperceptions will be correlated with their latent probability of doing the task conditional on being treated.

---

[5] For an explanation of the sample and why it is necessary to drop some observations, see Section 4.

6 Results

Tables 5 and 6 report the treatment effects. We find that updating men's information on their peers' beliefs increased the stated intention to complete any task by 4.4 percentage points and the probability of actually completing a task by 3.5 percentage points relative to the control group. This effect comes nearly entirely from firewood collection. When we condition treatment effect on having under-estimated peers' acceptance of a task, we find an impact of 7.7 percentage points on the stated intention to do any task and an impact of 3.9 percentage points on the probability of actually competing a task. Again, these effects are driven entirely by firewood collection. Among those who underestimated peers' acceptance of men collecting firewood, the impact on stated intention to participate in, and actual participation in, firewood collection is nearly 11 percentage points.

6.1 Childcare

Figure 2 shows performance measures by treatment (solid blue line, with 95% confidence intervals) and control (green dashed line) for different levels of the participant's misperception. Values to the left of the vertical line are for men who underestimate the true acceptance in their group, which is at zero, while men who overestimate appear to the right. The figures indicate the following: (a) At baseline, the groups are balanced. However, men whose guess is within 20% of the truth are more likely to perform the task (15 percent of them) compared to men who substantially underestimate (not distinguishable from zero); (b) during the activity's proceedings, the average man is confident that they will perform the task in the coming week, regardless of how their guess compares to the truth. However, the estimate is noisier for men with significnat underestimates of the truth (further to the left in the graph) than for those who guess more accurately; (c) during the follow-up interview with respondents' wives within one week of the activity, on average all men are more likely to perform the task than they were at baseline, regardless of treatment status (because the lines are between 0.1 and 0.2 on the y-axis even for men with misperceptions below -0.2). However, the treatment group men, who received updates about the true extent of acceptance by their peers, are significantly more likely than the control group men to have performed the task if they were reasonably close to the true acceptance level when they guessed. While one might be

surprised about this finding, that men who were already fairly accurate are the men who will be affected, it speaks to an important feature of our experiment that differentiates it from the BGY and other similar models: the men in our treatment group were not only updated with information, but they knew that (some of) their peers *also* received this update. of beliefs about the first- and second-order beliefs about others; (d) shows that these effects last one year later, but the nudge's effect on the control group and on men in the lower half of the misperception range fades. Note that the effect is smaller, and that endline performance rates are lower than at baseline, but this is likely because some men have children who aged out of the age range.

7   Discussion

Women perform a disproportionate share of unpaid household work worldwide, with implications for inequality and for economic development. Women in developing countries have a much larger total time burden then men (when combining economic activities and unpaid household work), and womens' total unpaid work burden has been shown to constrain participation in economic activities that could improve the household's living standard.

We implement a randomized control trial in rural Ethiopia, experimentally assigning men to a treatment updating their beliefs about their peers' views regarding gender norms. Specifically, we measure the effect of informing men about their peers' true beliefs about whether it is acceptable for men to help with unpaid chores traditionally performed by women.

We find that men do misperceive their peers' beliefs, with men more likely to under-estimate the share of their peers who believe it is acceptable for men to perform chores. The men whose perceptions are corrected are more likely to actually participate in chores one week after the intervention *and* one year after the intervention.

References

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics*, *128*, 469–530. From Bursztyn (2020):¡br/¿quot;This literature has focused mostly on the long-term persistence of cultural traits and norms (FernÃ¡ndez 2007, Giuliano 2007, Alesina, Giuliano, and Nunn 2013, VoigtlÃnder and Voth 2015)quot;.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9-10), 1082–1095.

Allen, J., Mahumane, A., IV, J. R., Rosenblat, T., Yang, D., & Yu, H. (2021). Correcting perceived social distancing norms to combat covid-19.

Allport, F. H. (1924). *Social Psychology*. Houghton-Mifflin.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., & Milkman, K. L. (2015). The effect of providing peer information on retirement savings decisions. *The Journal of finance*, *70*(3), 1161–1201.

Brent, D. A., Cook, J. H., & Olsen, S. (2015). Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials. *Journal of the Association of Environmental and Resource Economists*, *2*(4), 597–627.

Bursztyn, L., GonzÃ¡lez, A. L., & Yanagizawa-Drott, D. (2020). Misperceived social norms: women working outside the home in saudi arabia. *American Economic Review*, *110*, 2997–3029. From Bursztyn 2020: ¡br/¿¡br/¿The index is (1/6) â k=1 6 (k â q k)/ Ï k , where k indexes an outcome,¡br/¿and q k and Ï k are the mean and standard deviation of that outcome for participants originally in the control group. In addition, we display p-values adjusting for mul- tiple hypothesis testing following the procedure developed by List, Shaikh, and Xu (2019). Finally, we provide Lee (2009) attrition bounds for our treatment effects.

Byrne, D. P., Nauze, A. L., & Martin, L. A. (2018). Tell me something i donât already know: Informedness and the impact of information programs. *Review of Economics and Statistics*, *100*(3), 510–527.

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

Charmes, J. (2019). *The Unpaid Care Work and the Labour Market: An analysis of time use data based on the latest World Compilation of Time-use Surveys*. ILO Geneva.

d'Adda, G., Dufwenberg, M., Passarelli, F., & Tabellini, G. (2020). Social norms with private values: Theory and experiments. *Games and Economic Behavior*, *124*, 288–304. Cites Bursztyn 2020: ¡br/¿¡br/¿quot;Our study also contributes to the empirical literature on social norms. A large number of experiments, mostly random-¡br/¿ized control trials in the field, show how social information can influence behavior in a variety of settings.3 An interesting recent literature discusses the presence of pluralistic ignorance, defined as incorrect normative or empirical expectations, and shows that correcting it has large effects on behavior (Bursztyn et al., 2020a; Byrne et al., 2018).
URL `https://doi.org/10.1016/j.geb.2020.08.012`

Dahl, G. B., Kotsadam, A., & Rooth, D.-O. (2021). Does integration change gender attitudes? the effect of randomly assigning women to traditionally male teams. *The Quarterly Journal of Economics*, *136*, 987–1030.

Dean, J. T., & Jayachandran, S. (2019). Changing family attitudes to promote female employment. vol. 109, (pp. 138–142).

Delfino, A. (2021). Breaking gender barriers: Experimental evidence on men in pink-collar jobs. *Ssrn*.

Dhar, D., Jain, T., & Jayachandran, S. (2018). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india.

Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic Literature*, *50*(4), 1051â79.
URL `https://www.aeaweb.org/articles?id=10.1257/jel.50.4.1051`

FernÃ¡ndez, R., Fogli, A., & Olivetti, C. (2004). Mothers and sons: Preference formation and female labor force dynamics. *Quarterly Journal of Economics*, *119*, 1249–1299. From Bursztyn

2020: our work is thus related to a growing literature on the effects of gender identity and norms on economic outcomes (see Akerlof and Kranton 2000; Fernández, Fogli, and Olivetti 2004; Fernández 2007; Eckel and Grossman 2008; Fernández and Fogli 2009; Jensen and Oster 2009; Dohmen et al. 2011; Alesina, Giuliano, and Nunn 2013; Baldiga 2014; Coffman 2014; Bertrand, Kamenica, and Pan 2015; Jayachandran 2015; Field et al. 2016; Bursztyn, Fujiwara, and Pallais 2017; Bernhardt et al. 2018; Bordalo et al. 2019; and McKelway 2019.

Ferrara, E. L., Chong, A., & Duryea, S. (2012). Soap operas and fertility: Evidence from brazil. *American Economic Journal: Applied Economics*, *4*, 1–31.

Ferraro, P. J., & Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, *95*(1), 64–73.

Field, E., Pande, R., Rigol, N., Schaner, S., & Troyer Moore, C. (2021). On her own account: How strengthening women's financial control impacts labor supply and gender norms. *American Economic Review*, *111*(7), 2342–75.

Frey, B. S., & Meier, S. (2004). Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment. *American Economic Review*, *94*(5), 1717–1722.

Gerber, A. S., & Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, *71*(1), 178–191.

Greenwood, J., Seshadri, A., & Yorukoglu, M. (2005). Engines of liberation. *The Review of Economic Studies*, *72*, 109–133.

Hansen, C. W., Jensen, P. S., & Skovsgaard, C. V. (2015). Modern gender roles and agricultural history: the neolithic inheritance. *Journal of Economic Growth*, *20*, 365–404.

Healy, A., & Malhotra, N. (2013). Childhood socialization and political attitudes: Evidence from a natural experiment. *The Journal of Politics*, *75*, 1023–1037.

Hsieh, C., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, *87*, 1439–1474.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics*, *7*, 63–88. From Bursztyn: our work is thus related to a growing literature on the effects of gender identity and norms on economic outcomes (see Akerlof and Kranton 2000; FernÃ¡ndez, Fogli, and Olivetti 2004; FernÃ¡ndez 2007; Eckel and Grossman 2008; FernÃ¡ndez and Fogli 2009; Jensen and Oster 2009; Dohmen et al. 2011; Alesina, Giuliano, and Nunn 2013; Baldiga 2014; Coffman 2014; Bertrand, Kamenica, and Pan 2015; Jayachandran 2015; Field et al. 2016; Bursztyn, Fujiwara, and Pallais 2017; Bernhardt et al. 2018; Bordalo et al. 2019; and McKelway 2019.

Jayachandran, S. (2021). Social norms as a barrier to womenâs employment in developing countries. *IMF Economic Review*, *69*, 576–595. Cites Bursztyn 2018: ¡br/¿¡br/¿quot;Several other studies analyze attempts to change norms about womenâs work. Bursztyn¡br/¿et al. (2018) show compelling evidence of the importance of norms in Saudi Arabia. They elicit menâs beliefs about womenâs work, and their âsecond-order beliefs,â meaning their beliefs about othersâ beliefs. They find that men systematically overestimate peersâ disap- proval of womenâs work. When the researchers provide accurate information to men about their peersâ views, menâs wives are more likely to begin seeking employment.quot;.

Jensen, R., & Oster, E. (2009). The power of tv: Cable television and women's status in india. *Quarterly Journal of Economics*, *124*, 1057–1094. From Bursztyn: our work is thus related to a growing literature on the effects of gender identity and norms on economic outcomes (see Akerlof and Kranton 2000; FernÃ¡ndez, Fogli, and Olivetti 2004; FernÃ¡ndez 2007; Eckel and Grossman 2008; FernÃ¡ndez and Fogli 2009; Jensen and Oster 2009; Dohmen et al. 2011; Alesina, Giuliano, and Nunn 2013; Baldiga 2014; Coffman 2014; Bertrand, Kamenica, and Pan 2015; Jayachandran 2015; Field et al. 2016; Bursztyn, Fujiwara, and Pallais 2017; Bernhardt et al. 2018; Bordalo et al. 2019; and McKelway 2019.

Katz, D., Allport, F. H., & Jenness, M. (1931). *Students' Attitudes: A Report of the Syracuse University Reaction Study*. Craftsman Press.

Matulevich, E. C. R., & Viollaz, M. (2019). Gender differences in time use: Allocating time between the market and the household. *World Bank Policy Research Working Paper*.

McGavock, T., McCullough, E., & Assefa, T. (2024). Calling for time: Examining bias in time-use measurement using high-frequency phone surveys. *Working Paper*.

Mckelway, M., Grela, E., Nath, P., Samaddar, B., Sharma, S., Srinivasan, K. P., & Srivastava, A. (2019). Vicious and virtuous cycles: Self-efficacy and employment of women in india *. From Burszytn: our work is thus related to a growing literature on the effects of gender identity and norms on economic outcomes (see Akerlof and Kranton 2000; FernÃ¡ndez, Fogli, and Olivetti 2004; FernÃ¡ndez 2007; Eckel and Grossman 2008; FernÃ¡ndez and Fogli 2009; Jensen and Oster 2009; Dohmen et al. 2011; Alesina, Giuliano, and Nunn 2013; Baldiga 2014; Coffman 2014; Bertrand, Kamenica, and Pan 2015; Jayachandran 2015; Field et al. 2016; Bursztyn, Fujiwara, and Pallais 2017; Bernhardt et al. 2018; Bordalo et al. 2019; and McKelway 2019.

Oswald, A. J., & Powdthavee, N. (2010). Daughters and left-wing voting. *The Review of Economics and Statistics*, *92*, 213–227.

Pulerwitz, J., Hughes, L., Mehta, M., Kidanu, A., Verani, F., & Tewolde, S. (2015). Changing gender norms and reducing intimate partner violence: results from a quasi-experimental intervention study with young men in ethiopia. *American Journal of Public Health*, *105*, 132–137.

Washington, E. L. (2008). Female socialization: how daughters affect their legislator fathers. *American Economic Review*, *98*, 311–332.

Table 1: Baseline summary statistics by task

	(1) Firewood	(2) Childcare	(3) Laundry
Performed: by someone	0.235	0.524	0.171
	(0.424)	(0.500)	(0.377)
by her	0.143	0.514	0.094
	(0.350)	(0.500)	(0.293)
by him	0.014	0.122	0.004
	(0.118)	(0.328)	(0.063)
Minutes spent: by anyone	13.614	123.671	11.757
	(33.637)	(177.194)	(32.375)
by her	6.978	94.574	6.265
	(23.055)	(147.080)	(25.160)
by him	0.572	6.476	0.161
	(5.505)	(21.395)	(2.832)
Acceptability: he says yes	0.914	0.954	0.888
	(0.281)	(0.210)	(0.316)
average guess	0.804	0.870	0.849
	(0.249)	(0.200)	(0.220)
he underestimates	0.430	0.363	0.335
	(0.496)	(0.481)	(0.473)
he overestimates	0.293	0.285	0.398
	(0.456)	(0.452)	(0.490)
he guessed 100%	0.496	0.592	0.576
	(0.500)	(0.492)	(0.495)
Observations	498	498	498

Table 2: Attrition analysis by assigned treatment status

	Sample Size			P-value
	Treatment	Control	Total	on Treat coef
Recruited households (Her Time study)	325	321	646	
Households eligible for GNA experiment	281	285	566	0.29
GNA male completed experiment	247	249	496	0.77
Outcomes recorded after the experiment				
One-week post experiment	243	248	491	0.94
One-year post experiment	238	240	478	0.84

Notes: The table reports the number of observations for which each outcome is measured for households assigned to the treatment group (column 1), the control group (column 2), and total (column 3). For each outcome, column 4 reports the p-value for a t-test that the coefficient on treatment equals zero, when the outcome is whether the household completed the step referred to in the row and the sample includes all recruited households (for the eligible households test) and all households eligible to participate in the GNA experiment (for all other tests).

Figure 1: Treatment Effects

(a) Effects on Participant
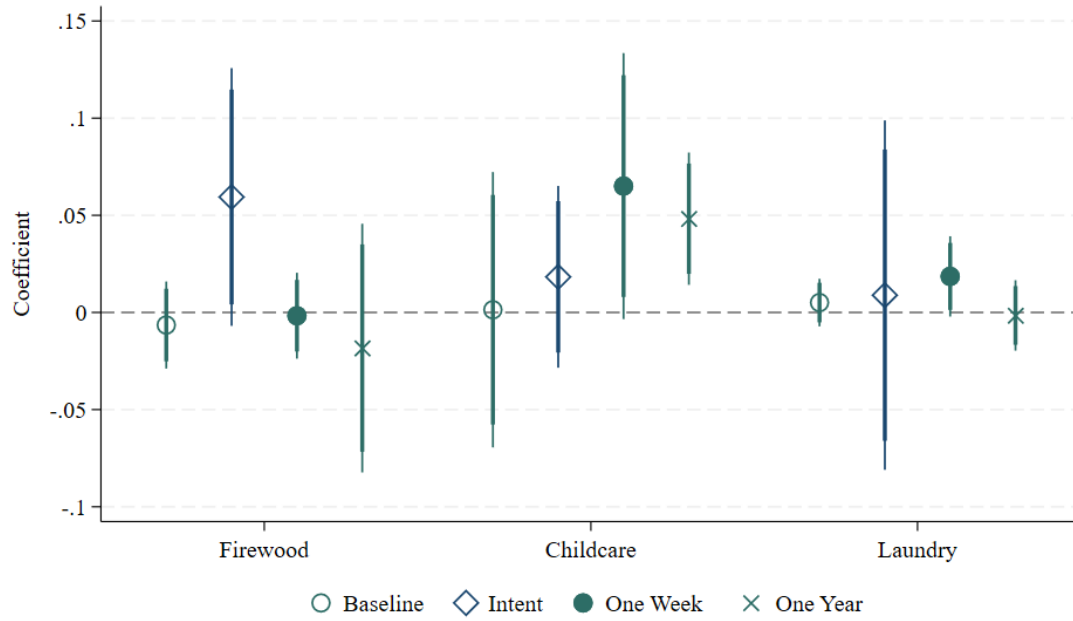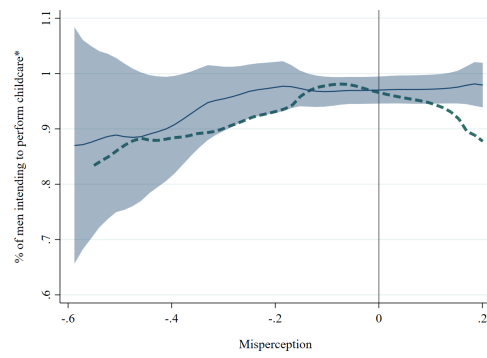

(b) Effects on Relative Shares
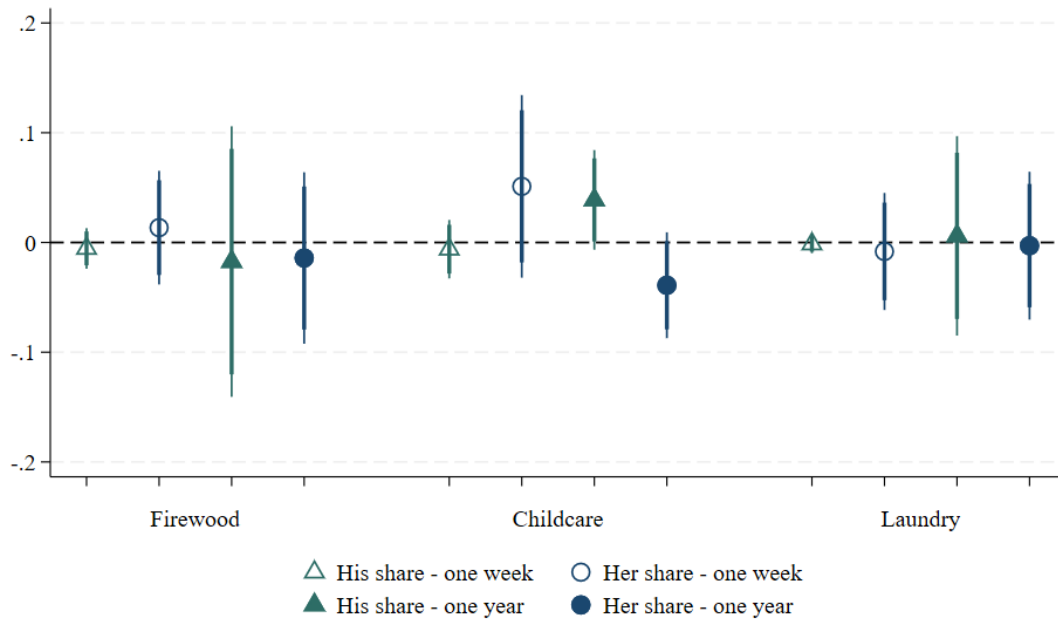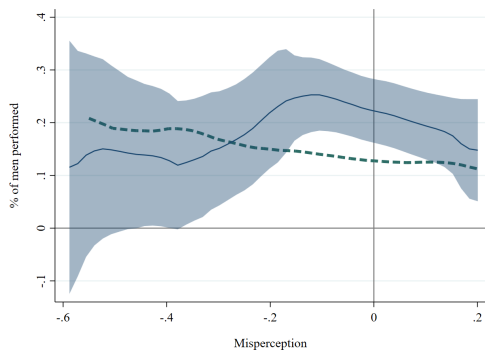
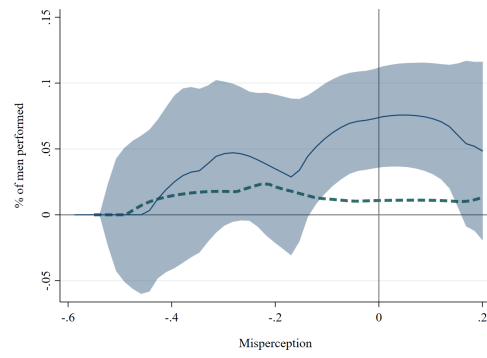Figure 2: Childcare Outcomes by Misperception

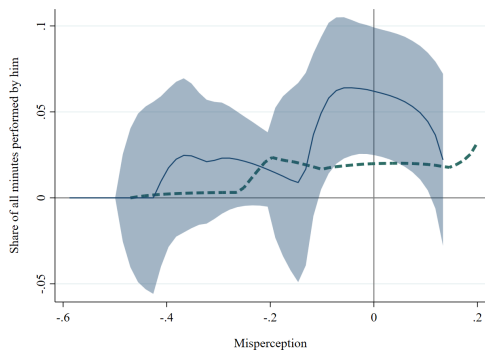(a) Baseline
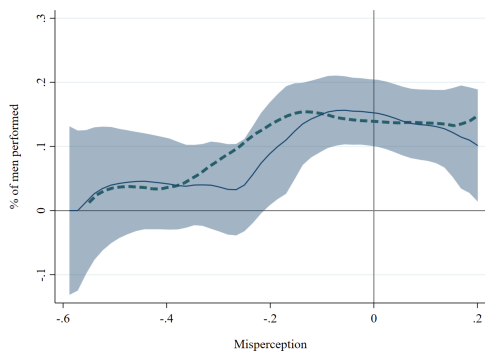
(b) Intends to Perform
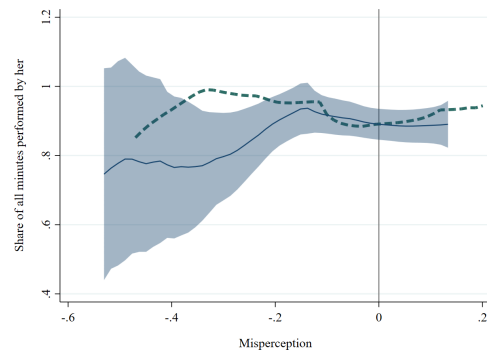

(c) Childcare: One Week
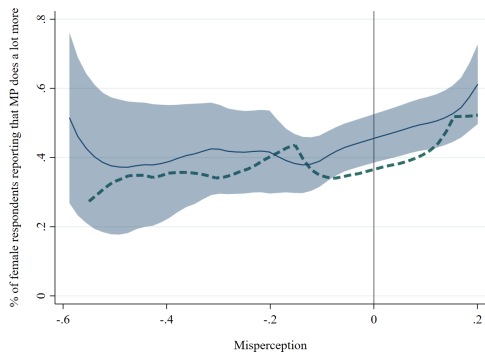

(d) Childcare: One Year

(e) His Share

(f) Her Share

(g) She says 1 year later that he does more

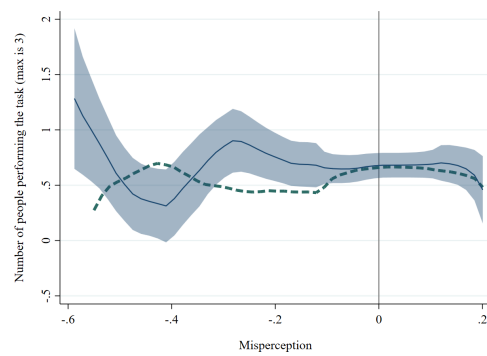(h) Number of people in household performing the task (1 year)

Table 3: Descriptive statistics and balance

	(1) Full Sample	(2) Treatment	(3) Control
Household demographics			
Age: male participant	43.421	43.757	43.082
	(11.885)	(12.131)	(11.646)
female respondent	36.937	36.680	37.196
	(8.941)	(9.151)	(8.736)
gap (male - female)	6.484	7.077	5.886
	(10.189)	(10.331)	(10.030)
Number of children: under 24 months	0.297	0.308	0.286
	(0.466)	(0.471)	(0.462)
2 to 5 years	0.453	0.466	0.441
	(0.553)	(0.546)	(0.560)
6 to 15 years	1.671	1.632	1.710
	(0.996)	(1.039)	(0.950)
Dependency ratio: children/seniors dependent	1.013	1.047	0.978
	(0.593)	(0.637)	(0.543)
adult males also dependent	3.376	3.479	3.271
	(1.597)	(1.585)	(1.606)
Household economic activities			
Land owned (ha)	0.722	0.695	0.749
	(0.529)	(0.494)	(0.563)
Economic activities of HH: operates farmland	0.984	0.984	0.984
operates sharecropped land	0.388	0.413	0.363
owns livestock	0.974	0.955	0.992
operates any enterprise	0.183	0.186	0.180
any wage work past year	0.337	0.328	0.347
female respondent participates	0.787	0.769	0.804
Household chores—performance of tasks			
Tasks performed for household: Total number of person-tasks	5.238	5.198	5.278
	(1.773)	(1.773)	(1.775)
Total minutes	328.400	324.850	331.980
	(214.969)	(206.724)	(223.340)
Tasks performed by female respondent: Number of tasks	3.313	3.259	3.367
	(1.740)	(1.762)	(1.719)
Her share of all tasks	0.640	0.640	0.639
	(0.299)	(0.312)	(0.286)
Her share of all minutes	0.727	0.708	0.746
	(0.281)	(0.303)	(0.256)
Tasks performed by male participant: Number of tasks	0.144	0.138	0.151
	(0.363)	(0.357)	(0.370)
His share of all tasks	0.038	0.039	0.036
	(0.100)	(0.105)	(0.095)

23

Table 4: Descriptive statistics and balance

|	Control	Treatment	p-value
Age	42.837	43.780	0.389
	(11.832)	(12.404)	
Religiosity	1.801	1.711	0.191
	(0.759)	(0.757)	
VESA saving last week (Br)	18.736	19.089	0.687
	(10.609)	(8.741)	
VESA meetings attended	1.130	1.138	0.827
	(0.423)	(0.401)	
Collected firewood	0.024	0.008	0.155
	(0.155)	(0.090)	
Cared for child	0.118	0.122	0.890
	(0.323)	(0.328)	
Washed clothing	0.004	0.000	0.318
	(0.064)	(0.000)	
Personally accept (firewood)	0.919	0.907	0.633
	(0.274)	(0.292)	
Personally accept (child care)	0.943	0.963	0.287
	(0.232)	(0.188)	
Personally accept (laundry)	0.886	0.886	1.000
	(0.318)	(0.318)	
Perceived acceptance (firewood)	0.810	0.796	0.546
	(0.251)	(0.250)	
Perceived acceptance (child care)	0.861	0.877	0.369
	(0.205)	(0.195)	
Perceived acceptance (laundry)	0.841	0.857	0.439
	(0.229)	(0.213)	
N	246	246	

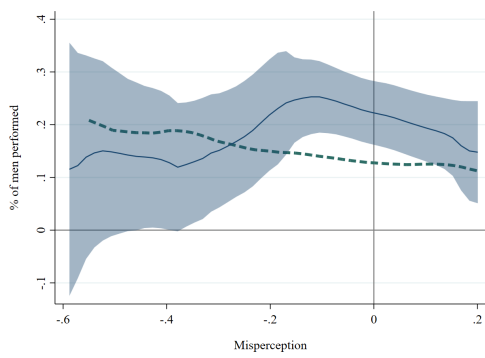Figure 3: Men's Performance of Tasks
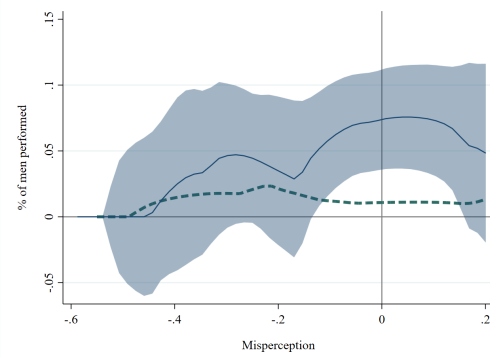
(a) Firewood: One Week
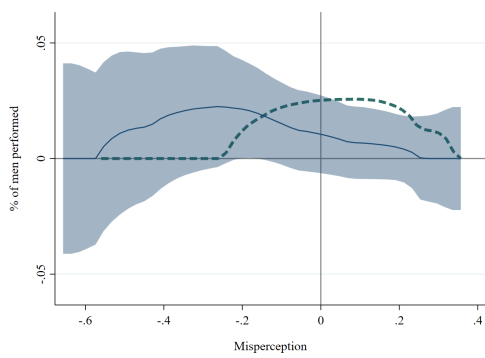
(b) Firewood: One Year
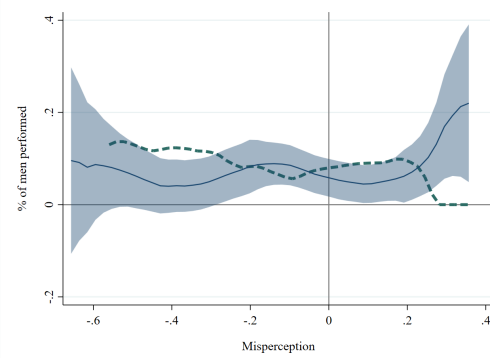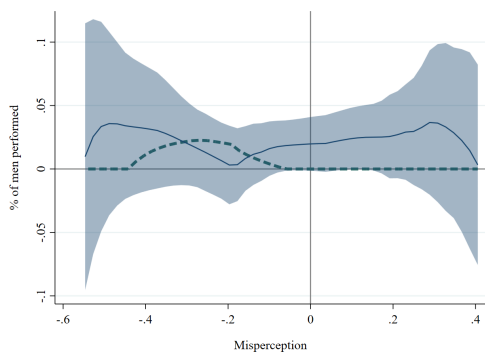
(c) Childcare: One Week

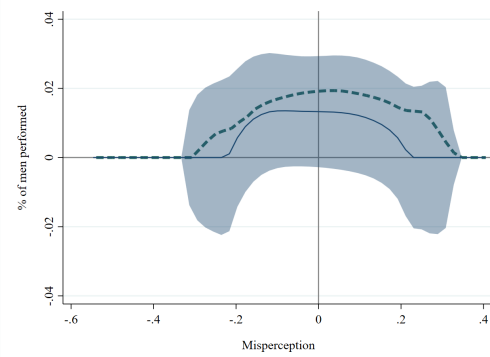(d) Childcare: One Year

(e) Laundry: One Week

(f) Laundry: One Year

Table 5: Treatment effects on stated intent to do a task

	All tasks	Firewood	Child care	Laundry
Treatment (α_1)	0.044	0.056	0.017	0.005
	(0.041)	(0.097)	(0.496)	(0.923)
Treatment (β_1)	0.025	0.020	0.000	-0.023
	(0.304)	(0.636)	(0.993)	(0.643)
Under (β_2)	-0.027	-0.077	-0.056	-0.047
	(0.423)	(0.159)	(0.073)	(0.381)
Treatment $\times$ Under (β_3)	0.051	0.087	0.043	0.081
	(0.129)	(0.254)	(0.289)	(0.216)
Treatment | Under ($\beta_1 + \beta_3$)	0.077	0.107	0.043	0.059
	(0.014)	(0.073)	(0.295)	(0.398)
Control mean	0.888	0.854	0.945	0.866
N	1,494	498	498	498

Notes: α_1 is estimated from equation 1 and $\beta_1, \beta_2, and \beta_3$ are estimated from equation 2 P-values in parenthesis.

Table 6: Treatment effects on actually doing a task

|	All tasks	Firewood	Child care	Laundry	Any task
Treatment (α_1)	0.027	-0.002	0.061	0.016	0.064
	(0.022)	(0.883)	(0.050)	(0.146)	(0.086)
Treatment (β_1)	0.025	-0.017	0.074	0.020	0.091
	(0.076)	(0.385)	(0.040)	(0.097)	(0.084)
Under (β_2)	0.003	-0.016	0.060	-0.001	-0.035
	(0.880)	(0.302)	(0.253)	(0.905)	(0.525)
Treatment $\times$ Under (β_3)	0.005	0.035	-0.031	-0.011	-0.078
	(0.863)	(0.180)	(0.699)	(0.444)	(0.382)
Treatment | Under ($\beta_1 + \beta_3$)	0.030	0.018	0.044	0.009	0.013
	(0.189)	(0.203)	(0.507)	(0.568)	(0.838)
Control mean	0.054	0.016	0.141	0.004	0.161
N	1,491	498	495	498	498

Notes: α_1 is estimated from equation 1 and $\beta_1, \beta_2, and \beta_3$ are estimated from equation 2 P-values in parenthesis.
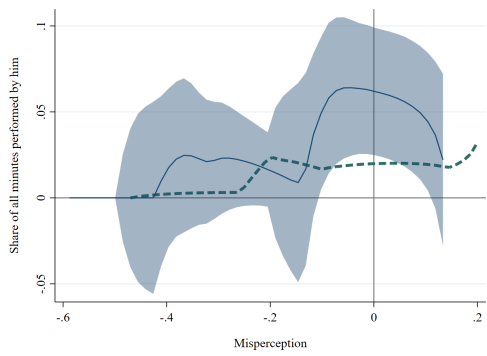
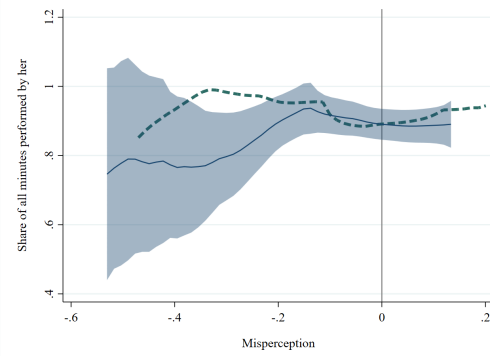Figure 4: Share of Minutes on Tasks

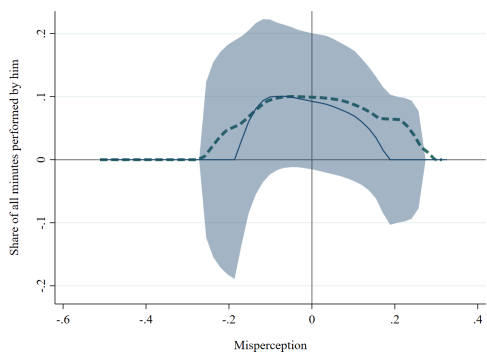(a) Firewood: His Share
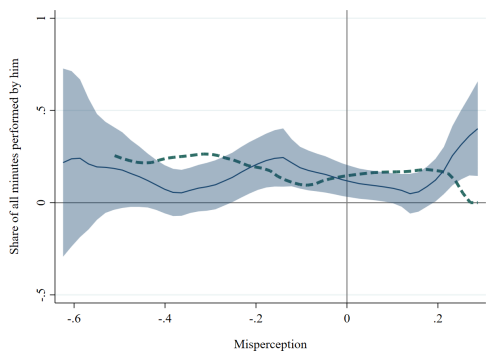
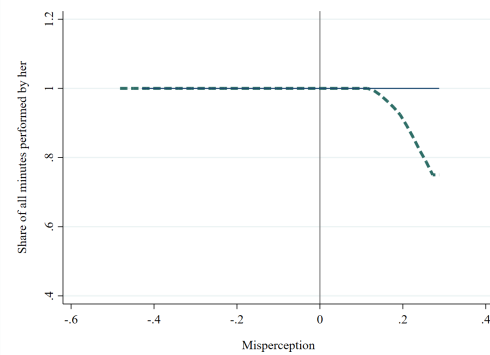(b) Firewood: Her Share
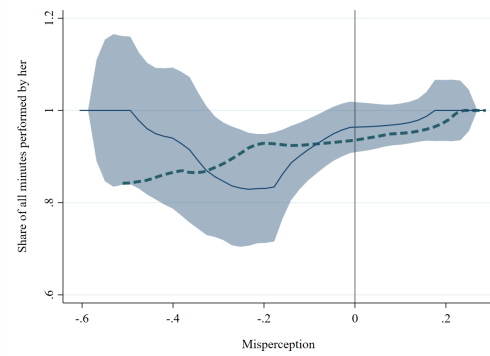
(c) Childcare: His Share

(d) Childcare: Her Share

(e) Laundry: His Share

(f) Laundry: Her Share

created for sharing with discussant Jan 3, 2022

Gender Norms Activity Protocol

For every female respondent (see Sample Design), a male household member will be recruited to participate in a Gender Norms Activity that seeks to measure the male participants' beliefs about local gender norms. The meeting will take place on the morning after the last baseline interview is conducted in the VESA or the evening of the last baseline interview. The meeting will consist of two parts. First in the voting session, all of the men will be gathered together, instructions will be given, and men will provide their answers to the questions about tasks and about their guesses about other men's votes. After the voting session, men will be divided into two groups. Half of the male participants in the Gender Norms Activity will be randomly assigned to the treatment group, and will receive an information intervention designed to address possible misperceptions by participants about average beliefs in the community. The other half of the group will receive an unrelated "placebo" treatment. After the information revealing session is concluded, all male participants will be asked to provide responses to a few additional questions before the meeting concludes, while remaining in their separate groups. Therefore, there should be at least two enumerators present for the GNA, one to lead activities with the treatment group and one to lead activities with the control group.

COVID protocol for GNA Activity

In addition to the COVID protocols associated with in-person interviews in the previous section, additional protocols apply to the GNA activity. All group meetings should occur outdoors, and group size should be strictly limited to 30 including enumerators and respondents. Strict social distancing (2M) should be enforced between all participants, and between enumerators and participants. Enumerators should wear a mask, and they should offer facemasks to participants as well. Everyone (enumerators and participants) should be offered hand sanitizer at the start and end of the activity.

The research team should screen individual participants one at a time for COVID symptoms as they approach the group (see the consent form for the screening language), and shall divert any participants who are experiencing COVID symptoms. Once the group is convened and the enumerator reads the full consent script, the enumerator shall again include the language about COVID risks and ask any participant who meets the COVID screening criteria to leave the group at that time.

The survey team should disinfect all pencils, markers, and tiles that are handled by participants with disinfectant spray at the time they are reorganizing and reassembling packets for participants. Participant packets should not be reused within a 72 hour period (the team should switch to the second set of materials if two GNAs are conducted within a 72 hour period, which is not anticipated to occur except in rare circumstances).

Prior to the meeting

In collaboration with the VF, the survey team will set a date, time and location for the Gender Norms Activity prior to arrival in the VESA. The VF will inform the VESA members in communications prior to the survey team's arrival in the VESA. The survey team should avoid changing the date, time, and location for the GNA once it has been announced within the community.

During the baseline interviews with female respondents, the enumerator will use the household roster to identify the male who should participate in the Gender Norms Activity. Details about how to identify the male participant are listed in the sample design section of this document. At the time of gaining consent from the survey's primary (female) respondent, the enumerator will mention that it is expected also that a male household member (to be identified) will be invited to participate in the Gender Norms Activity. The enumerator will share the time and place for the Gender Norms Activity with the female respondent during the baseline visit to the household, will clearly state which male member is expected to attend the meeting, and will attempt to directly communicate with the targeted male the meeting time and location. The enumerator will record the GNA participant's name and, if possible, the phone number of someone who may be contacted in the event that the participant does not arrive on time (note that for phone treatment and phone diets groups, this can be the phone number provided to the female responding in this study).

Materials

There are two types of required materials for the GNA: participation bags and enumerator kits.

Participant bags: each participant in the GNA will receive a bag containing all of the materials required for him to vote on each of the questions. The bags should be labeled with a Token ID number. The Token ID number assigned to men should be recorded on the GNA participant list along with the vesa and household ID numbers, respondents' names, and the name of the female respondent they are matched with. These bags should contain:

- 13 white tokens and 13 black tokens. Each token should be labeled with the Token ID number that matches the number on the bag so that their responses can be matched back to the female respondent in their household.
- 7 response cards with the same Token ID numbers. These cards will depict a sample number of men in a group. The men will use the card to record their beliefs about how many men in the group agree with each statement. These should be printed on papers with unique colors and stacked in the same order for each participant kit so that the participants can answer using the same color as the group. This will help to keep track of papers during the data entry process.
- A crayon for circling their responses on the response cards.

The Enumerator Kits contain the materials for collecting responses from participants and for displaying the results (to participants in the treatment group). The list below details materials

required before and after the participants are divided into two groups. The enumerator kit contains:

- A copy of the consent script and the GNA script
- Mock participant bag for demonstrating instructions to participants
- Materials for collecting men's beliefs and perceptions: 12 Question Bags (1A, 1B, 2A, 2B, …, 6A, 6B) and 2 practice question bags (PA, PB). These bags will be used for men to share their beliefs and the perceptions about their group's beliefs during the large group session. These bags will be numbered with the question number. There will be two question bags for each question number, labeled PA/PB, 1A/1B, 2A/2B, etc.
  - The A bags will be used to collect the men's tokens in response to the first part of each question: "is it acceptable for men to …?" These bags should also have task picture cards attached to them, with the picture corresponding to the activity asked in that question.
  - The B bags will be used to collect the response cards of the men in response to the second part of each question "of the men in the group here, how many do you think responded to the first question with a 'yes' response?"
- Materials for displaying results: (these stay with the "Treatment" group when the participants are divided into two groups)
  - Banner and banner stand, used to display the number of yes and no responses of the men to the questions about whether it is acceptable for men to perform the different activities (see details below).
  - Picture cards (4 - one for each chores related question plus one for the practice "coffee" question). Each question will have an associated visual depiction (e.g. a bundle of firewood) so that when the group level acceptability votes are displayed to the treatment group, the visual depiction will be displayed next to it. Then the men will be able to associate the acceptability information they are provided with its corresponding question.
  - Large display response cards (4 - one for each chores related question plus one for practice question). There will be one large response card for each question asked during the GNA large group session. These response cards will simply be larger versions of the response cards in the participant bags. Enumerators will use these to display the group's actual responses to the questions about acceptability
  - Plastic folders for large response cards (4 - one for each chores related question plus one for practice question). Each large response card will be placed in a dry erase envelope to allow enumerators to use markers to show the results. The dry erase folders will be reused in subsequent GNA sessions.
  - Dry erase markers. The enumerators should each have three dry erase markers: a red marker, a blue marker, and a black marker. These will be used to mark the large response cards on the banner display.
- Outcome bags to record participants' intentions.
  - The "Treatment" group will have 6 outcome bags (1T, 2T, 3T, …, 6T)
  - The "Control" group will have 6 outcome bags (1C, 2C, 3C, …, 6C)

- CAPI for administering the community questionnaire to the control group.

During the Gender Norms Activity, the field team will go through the following steps. These have been outlined in more detail in the GNA Script and "cheat sheet" handouts for the field team.

- Explain the activities that will take place and read the consent script. Reiterate that this Gender Norms Activity is not connected to any VESA activities or benefits; assure participants that their responses will not be seen by their VF or anyone associated with the program; politely ask the VF to leave the meeting if still present. Gain verbal consent from each respondent and put a tick on the participant list to confirm this.
- Complete a practice round about coffee, explaining the contents of the participant kits and how men can keep their votes private. See checklist and GNA script for more details.
- For each question (both the gender norms questions and the placebo questions), collect each participant's acceptability vote and group belief response card using the question bags. After voting for each question is completed, the question bags should be placed in a location that is completely visible to the men so that they do not become suspicious that tokens have been tampered with.
- Divide men into two groups (the "GNA treatment" group and the "GNA control" group) based on the pre-assigned treatment for each man. The groups should be separated so that the control group cannot hear or see the display banner of the treatment group.
- Treatment group: The display banner should be set up with the picture cards for the questions assigned to that group and the large response cards in the dry erase folders. The picture cards should be oriented right-side-up so the participants will recognize the item. The enumerator will reveal the responses for the treatment questions one by one, by counting the number of tokens out loud and by showing on a large response card the group's true response. The enumerator will color in the responses on the large response card which is inside a plastic folder and mounted on the banner (see figure 2 of Appendix 1 for an example). After the information is presented on the board, the enumerator will return the tokens to the question bag. The token numbers and responses will be recorded later. Participants' response cards will not be displayed or discussed in the group.
- Control group: will respond to questions about community infrastructure.
- Both groups: the enumerator will ask men to vote one additional time per question about their intentions. Using their response tokens, they will vote in the same fashion in response to the question "Do you plan to do [task/activity] in the next 7 days?"
- It is very important that the treatment group men are dismissed before the control group men, and that no control group men be allowed to see the display boards shown to the treatment group.

Following the GNA, question bags should be emptied carefully and the data should be recorded on the spreadsheet provided in accordance with the data entry instructions (see Appendix 2). Once all data have been recorded, the tokens can be returned to the participant bags in preparation for the next VESA. The Gender Norms Activity measurement data entry sheet should be used to record the individual male responses to the gender norms activity. It will be filled by the field team supervisor and not by respondents. Instructions for the data entry are included in Appendix 2.

Appendix 1: Gender Norms Activity Cheat Sheet

HER TIME: A TIME USE STUDY OF WOMEN PARTICIPATING IN LIVELIHOODS PROGRAMS IN ETHIOPIA

GNA Cheat Sheet for Field Team – 2 pager
For more complete directions, see GNA Script and Instructions, or the Her Time Basic Information Document (BID)

BEFORE GNA
- Confirm day/time with the VF
- Recruit participants, record their names and phone numbers on GNA participant list
- Prepare materials
  - Response cards with enough figures on it for VESA size [see soft copies] – for display and for participants
  - Participant Kits with tokens, cards, and crayons
  - Enumerator Kit for Part 1 with coffee display cards and markers
  - Enumerator Kit for Part 2 Treatment with display cards and markers, outcome bags
  - Enumerator Kit for Part 2 Control with outcome bags & CAPI for questionnaire

DURING GNA – PART I
- Confirm participants list; record token numbers; adjust participant list for new/missing participants
- Ensure privacy: sit far apart, VF is not present
- Read consent form; tick consent column next to name on participant list
- Complete practice round (coffee last evening?)
  - Demonstrate voting with tokens
  - Demonstrate response cards – should put a circle for self if they voted yes
  - Check response cards of each participant as they put them in the bag
- Show tokens results from practice round, display on board, interpret (see Fig. 1)
  - Show the photo on the bag
  - Take the tokens out
  - Count out loud and show the white tokens, then the black
  - Black out extra figures on display response card

- In RED circle number of figures corresponding to number of white tokens
- In BLUE cross out (with an X) number of figures corresponding to black tokens
- Explain under/over estimation
- Complete remaining questions:

1. Do you believe it is acceptable for men to collect firewood for the household?
2. Do you believe that eating vegetables such as orange sweet potatoes, pumpkin, carrots, and squash can make you healthier?
3. Do you believe it is acceptable for a man to take care of children such as bathing, feeding, holding, or supervising while his wife participates in IGA or attends a community meeting?
4. Do you believe that eating eggs will help children to learn more in school?
5. Do you believe it is acceptable for a man to wash his own clothes?
6. Do you believe that going to school will make children's lives better in the future?
- Call out names of Control participants, ask them to take their token bags and follow enumerator

DURING GNA PART II
- In the treatment group ONLY, show responses to the treatment questions (TREATMENT: 1/3/5) – see Fig 2
- Do NOT show responses to the control group
- Do not discuss the results
- In BOTH groups, ask the outcome questions – all 6 questions should be asked in both groups
  1. Do you think that you will collect firewood for your household in the next 7 days?
  2. Do you think that you will eat vegetables such as orange sweet potatoes, pumpkin, carrots, and squash in the next 7 days?
  3. Do you think that you will take care of children such as bathing, feeding, holding, or supervising while your wife participates in IGA or attends a community meeting in the next 7 days?
  4. Do you think that your children will eat eggs in the next 7 days?
  5. Do you think that you will wash your own clothes in the next 7 days?
  6. Do you think that your children will go to school in the next 7 days?
- Dismiss T group
- In C group, ask community questions from the VF questionnaire, make sure T group is finished, then dismiss
- Pack up all question bags carefully

AFTER GNA
- Enter the data – see data entry instructions
- Prepare materials for the next VESA

Figure 2: Correct marking for DISPLAY BOARD if there are 4 white and 4 black

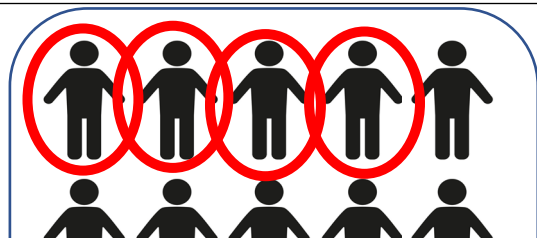Figure 1: Correct marking for PARTICIPANT CARD if they believe that 4 participants (including themselves) answered "yes"

Appendix 2: Data Entry Instructions for Gender Norms Activity

Prior to beginning the data entry, ensure that each of the following items is present:
- o All question and outcome bags from the intervention. There should be four bags per question, each marked with the question number and a category: A, B, CT, and CC. In total, there are 24 bags. marked with question number and category (token or card). Make sure they are ordered by question.
- o Participant list from the activity (paper), completed with household IDs, treatment status, and token ID numbers. The token ID is the number that appears on the token and sheets they used to answer the question.
- o "GNAdataentrytemplate.ods" spreadsheet file into which the data will be entered.

Notes:
- o If there is no token or card in the question bag for a certain Token ID number, use -99.
- o Make sure to enter the data from the practice question under the columns marked PA and PB. The corresponding data will come from the "PA" and "PB" bags.

Steps:
1. In the Her Time Dropbox folder, in the GNAData folder, single click on the GNAdataentrytemplate.ods file. While it is highlighted, press Ctl+C. Then open the folder corresponding to the wave number. Press Ctl+V to paste the template file into the wave folder. Rename this file vesaXX (where XX corresponds to the two-digit numerical VESA ID). Open this file. Do not change the name of the sheet – leave it as Sheet1. Only one VESA's data should be entered in each file, so for each wave there should be 7-8 .ods files, one for each VESA. This means that once a VESA's GNA data have been entered into the file for that VESA, that VESA's file should *never* be opened again by the field team.
2. Verify that the Token IDs in the first column correspond to the numbers that were used during the experiment for this VESA.
3. Enter the *numerical code* for VESA into cell V2. Click the bottom right corner of this cell and drag down to fill it out for all of the rows that will be used (corresponding to the number of participants in the GNA).
4. Now the demographic information should be entered for each participant, one row at a time. The steps for each row are:
   a. Find the Token ID number in Column A of the spreadsheet on the GNA Participant List.
   b. Then enter the participant information that appears in the GNA Participant List for that Token ID into the spreadsheet (Columns Y to AE).
   c. Then move on to the next Token ID number and repeat.
   d. Save the file by pressing Ctl + S.
5. Now we are ready to enter the data from the participants' answers during the activity. Pull out the two bags corresponding to the practice question: PA and PB
6. Carefully empty the tokens from the PA bag onto a flat surface.
   a. Turn them over so that the Token ID can be observed for each one.

b. The tokens should be lined up in order by Token ID order for easy data entry.

c. Verify that there are as many tokens as total respondents. If there are not, make sure to use -99 in that cell for that question for that Token ID number, as mentioned above.

7. Making sure to enter the data from each token into the correct Token ID row, fill the 'PA' column, with *0 for black and 1 for white*. This means that a "yes" answer to the question will have the value 1 and a "no" answer will have the value zero.

   a. Replace the tokens into the bag carefully.

   b. Save the file by pressing Ctl+S.

8. Carefully empty cards from the PB bag on to a flat surface.

   a. The cards should be sorted in order by Token ID for easy data entry.

   b. Verify there are as many cards as total respondents If there are not, make sure to use -99 for that cell, as mentioned above.

9. Populate the 'PB' column with the *number of figures circled in red on the card*, being sure to match the ID number on the card with the Token ID row in the sheet.

   a. Do NOT make any attempt to correct the participants' responses. For example, if they have circled more figures compared to the total number of participants (including themselves), record the actual number of figures circled on their card. However, if it is very clear that they made a circle and then tried to cross it off (to say they had circled one to many), record the number of figures they circled *not* including the one they crossed off.

   b. Replace the cards into the bag carefully.

   c. Save the file by pressing Ctl+S.

10. Repeat steps 3-7 for each of the additional questions (1A/1B, 2A, 2B, etc.).

11. Repeat steps 3-6 for each of the outcome question bags from the treatment group. Note that each bag will have only some of the tokens in them, because only some of the men were in the treatment group.

12. Repeat steps 3-6 for each of the outcome question bags from the treatment group. Note that each bag will have only some of the tokens in them, because only some of the men were in the treatment group.

13. When you are finished, be sure that you have saved all of the data and closed the spreadsheet.

The final product should look like the following:

TokenID	PA	PB	1A, 1B, etc.	woreda	kebele	vesa	HHID	etc.
1	0	1		1	2	1	101	
2	1	3		1	2	1	102	
3	1	4		1	2	1	103	
4	0	0		1	2	1	104	
5	1	3		1	2	1	105	
6	1	-99		1	2	1	106	

The data show that person #1 answered the question with a "no" (has a zero for PA), and thought that only 1 person in the group answered with a yes (has 1 figure circled on his card

for PB). Person #6 answered "yes" to the question (PA = 1) but did not submit a response card. So PB is marked -99 for Person #6 to show that the response is missing.