Misperceived Social Norms:
Female Labor Force Participation in Saudi Arabia*

Leonardo Bursztyn[†]   Alessandra L. González[‡]   David Yanagizawa-Drott[§]

June 2018

Abstract

Through the custom of guardianship, husbands typically have the final word on their wives' labor supply decisions in Saudi Arabia, a country with very low female labor force participation (FLFP). We provide incentivized evidence (both from an experimental sample in Riyadh and from a national sample) that the vast majority of young married men in Saudi Arabia privately support FLFP outside of home from a normative perspective, while they substantially underestimate the level of support for FLFP by other similar men – even men from their same social setting, such as their neighbors. We then show that randomly correcting these beliefs about others increases married men's willingness to let their wives join the labor force (as measured by their costly sign-up for a job-matching service for their wives). Finally, we find that this decision maps onto real outcomes: four months after the main intervention, the wives of men in our original sample whose beliefs about acceptability of FLFP were corrected are more likely to have applied and interviewed for a job outside of home. Together, our evidence indicates a potentially important source of labor market frictions, where job search is underprovided due to misperceived social norms.

Keywords: social norms, female labor force participation, Saudi Arabia

1 Introduction

Female labor force participation (henceforth, FLFP) remains low in several parts of the world, particularly in the Middle East and North Africa, where thirteen of the fifteen nations with the lowest rates of female participation in the labor market (Global Gender Gap Report 2015) are located. A salient example of a country with low rates of female labor force participation is Saudi Arabia, where less than 15% of the Saudi female population aged fifteen and above were employed in 2017 (a labor force participation rate of around 18%).[1] Anecdotal evidence suggests that social norms are a key constraint on local FLFP; for example, women are expected to work in spaces segregated from men. Men also play a key role in the labor supply decision of women: in order to work, the existing norm (but not law) implies that essentially all women are required to receive approval from their male "guardian," typically the husband or father. In recent years, changes in Saudi law might have created a more friendly environment for FLFP; for example, the ban on women's right to drive was lifted in 2018. Do male guardians in Saudi have correct perceptions of the opinions held by other men regarding FLFP?

In principle, a historically conservative country, in which laws are fast changing is a likely setting for "pluralistic ignorance" (Katz and Allport 1931), a situation where most people privately reject a norm, but they incorrectly believe that most other people accept it, and therefore end up following the norm as well. When individuals believe a behavior or attitude is stigmatized, they might be reluctant to reveal their private views to others for fear of social sanction. If most individuals act this way, they might all end up believing their private views are only shared by a small minority at most.[2] However, factors correcting beliefs about what others think might generate fast changes in the perceived acceptability of certain behaviors, and also in actual behavior (as examined in Bursztyn, Egorov and Fiorin 2017).[3]

In this paper, we combine experiments and surveys to first provide incentivized evidence that the majority of married men in Saudi Arabia in fact support FLFP outside of home, while they substantially underestimate the level of support for FLFP by other men—even those from their same social setting, such as neighbors. We then show that randomly correcting these beliefs about

---

[1]Source: General Authority For Statistics, Kingdom of Saudi Arabia, 2017.

[2]Historic examples of pluralistic ignorance include the late Soviet regime (Kuran 1991), where many individuals opposed the regime but believed others supported it. In 1968, most white Americans substantially overestimated the support for racial segregation among other whites (O'Gorman 1975). Work in psychology has also documented pluralistic ignorance regarding alcohol use on college campuses (Prentice and Miller 1973). A related concept is "preference falsification" (Kuran, 1995): people's stated, public preferences are influenced by social acceptability, and might be different from their true, private preferences. More closely related to our study, González (2014) documents that the majority of male Kuwaiti college students in her sample believed women should work outside the home, while they thought that the majority of their religious community would not approve of it. For a recent overview in social psychology, see Tankard and Paluck (2016). For the related concept of third order inference in sociology, see Correll et al. (2017),

[3]Following Bénabou and Tirole, 2011, we think of social norms as the set of 'social sanctions or rewards' that incentivize a certain behavior. We examine *injunctive*, but not *descriptive* norms.

others increases married men's willingness to let their wives join the labor force (as measured by their costly sign-up for a mobile job-matching service for their wives). Finally, we find that this decision maps onto real outcomes: three to five months after the main intervention, the wives of men in our original sample whose beliefs about acceptability of FLFP were corrected are more likely to have applied and interviewed for a job outside of home.

We first report on an experiment with a sample of 500 Saudi married men aged 18-35, recruited from different neighborhoods in Riyadh. These men each attended a 30-participant session composed of individuals from the same geographical area, thus sharing a common a social network.[4] In an anonymous online survey, 87% of the experimental participants agreed with the statement: "In my opinion, women should be allowed to work outside of the home." When incentivized to guess how other session participants responded to the same question, about three-quarters of the experimental subjects *underestimate* the true number. We interpret this as evidence of misperception of social norms, even among people from the same neighborhood who know each other. Those with larger misperceptions report lower confidence in their guesses when making them. Those with fewer social connections in the sample are also less confident and have more incorrect beliefs about the other session participants.

Next, we evaluate whether correcting these misperceptions matters for household labor supply decisions. Half of the participants were randomly given feedback on the true number of agreements with the statement in their session. At the end of the experiment, subjects were asked to make an incentivized choice between receiving an additional bonus payment (an online gift card) and signing their wives up for a job matching mobile application specializing in the Saudi female labor market. In a control group without belief corrections, 23% of participants chose the job matching service. In the treatment group with feedback on the opinions of other participants, the share went up significantly, by 9 percentage points (a 36% increase). The increase is driven by those who underestimated the true share of FLFP participant supporters in their session: sign-up rates go up by 57% (from a baseline rate of 21%) when this group is provided with information, while information doesn't change sign-up rates for those who did not underestimate support by others (that group also has a higher baseline sign-up rate of 31%).[5]

One might be worried that the sign-up outcome is not strongly indicative of "real" labor supply decisions, or that the immediate decision does not imply a more permanent change in perceived social norms and behavior. To deal with these concerns, three to five months after the original intervention, participants were recontacted by phone and a series of additional outcome questions

---

[4]The average participant reported knowing 15 of the 29 other session participants.

[5]This first outcome is a decision made by husbands/guardians, and not by wives themselves. We think husbands' decisions are crucial, since we are examining men's potential reluctance to let their wives join the labor force due to perceived social norms as an obstacle for FLFP. Moreover, due to the custom of male guardianship (until 2011, guardians' permission was legally required, and since then their permission is generally asked by hiring firms), husbands typically have the final word on their wives' labor supply decisions. Also, note that since a participant's wife's eventual employment status is observable, the observability of the sign-up choice itself does not matter independently.

were collected.[6] We document a longer-term impact on real labor supply outcomes. Wives of treated participants were significantly more likely to have applied for a job outside of home (up by 10 percentage points from a baseline level of 6%) and to have interviewed for a job outside of home (up by 5 percentage points from a baseline level of 1%). We are not powered to detect a significant change in the likelihood of the wife being employed outside of home, though we directionally observe an increase. We also document that the change in perceived social norms is persistent: treated participants believe a significantly higher share of their neighbors in general support women working outside of home. Finally, we observe that the persistent change in perceived social norms might spill over to other behaviors: treated participants are significantly more likely to report that they would sign up their wives for driving lessons. These findings are robust to adjustments for multiple testing.[7]

We also conduct a similar-looking, anonymous online survey with a larger, national sample of about 1,500 married Saudi men aged 18-35. The goal of this additional survey is two-fold. First, we assess the external validity of the finding that most Saudi men privately support FLFP while failing to understand that others do as well. In this broader, more representative sample, 82% of men agree with the same statement on FLFP used in the main experiment. When incentivized to guess the responses of other survey respondents, 92% of them underestimate the true share. These are stronger misperceptions, perhaps because they are no longer asked about their own neighbors' opinions.

Second, we examine whether social desirability bias/experimenter demand effects could have been a driver of the misperception finding in the main experiment. Indeed, although the experiment was anonymous, it is possible that some participants may have felt like they had to answer the question about their own views in a certain way. Half of the online survey participants were

---

[6]To preserve anonymity in the original experiment, while still being able to contact participants in the future, phone numbers were first collected at the session level without matching them to specific respondents, before the experiment started. In the follow up phone survey, participants were asked for the last three digits of their phone numbers. We were able to match 95% of the phone numbers to the combination of last three digits and session number.

[7]Since the vast majority of men in our sample privately support FLFP outside of home, we believe updates in perceived social acceptability are the main mechanism driving our findings (in Appendix A, we present a simple model of labor supply and stigma based on this mechanism). However, it is possible that the information provided leads some participants from having a privately negative opinion about FLFP to a positive one. We did not collect updated opinions after the information was provided to verify this possibility. Still, we can check the treatment effects for those originally opposed to FLFP outside of home. We find a large point estimate (10.9 p.p. increase from a baseline sign-up rate of 10.7% in the control group), but this estimate is not significant, perhaps due to the small sample size for that group (N=65). This might be interpreted as suggestive evidence of a persuasive effect of the information treatment in that subsample. However, it is also consistent with these participants not changing their opinions, but changing their behavior because they care strongly about their social image and their perceptions of norms have been substantially updated. Indeed, the average wedge among those originally negative about FLFP is substantially and significantly larger than among those originally positive: -11p.p. vs. -6.7 p.p (p=0.000). Furthermore, the point estimate of the treatment effect is unchanged when we restrict the analysis to those who originally reported positive own views about FLFP outside of home. Finally, the treatment effects are stronger for those who experience a larger update in beliefs about the opinions of others.

assigned an elicitation procedure that provided a "cover" for their opinion on FLFP. In particular, we implemented a "list experiment" (also called the "unmatched count" and the "item count technique", originally formalized by Raghavarao and Federer 1979).[8] Using a method that provides respondents with a higher degree of plausible deniability, we find a very similar level of agreement with the statement regarding whether women should be allowed to work outside of home: 80%. Finally, we examine whether individuals may incorrectly expect others to strategically respond to the FLFP agreement question, which would distort guesses about others since the question asked about how others *answered* the question. We find that beliefs about other participants' *true opinions* were extremely similar to the guesses about others' *answers*.

As a last check on the external validity of the fact we document, we show that the share of Saudi men supportive of FLFP outside of home is also very similar when using the nationally representative sample from the wave of the *Arab Barometer* containing that question for Saudi Arabia (2010-2011). Out of approximately 1,400 male respondents, 75% agree with the statement "a married woman can work outside the home if she wishes." Among male respondents aged 18-35 (the age bracket in our study), the share is 79%. The *Arab Barometer* survey also allows us to establish that older men are also supportive of FLFP outside of home: among those over 35, the share agreeing with that statement is 71%.[9] Moreover, the numbers from the *Arab Barometer* in 2019-2011 suggest that the misperception in social norms might have not have been a short-lived phenomenon, at least in the sense that support for FLFP has been high and remained relatively constant until today.

We contribute to a growing literature on social norms in economics. This literature has focused mostly on the long-term *persistence* of cultural traits and norms (Fernandez, 2007, Voigtlaender and Voth 2012, Giuliano 2007, Alesina et al. 2013). We study how long-standing social norms can potentially change with the provision of information.[10] We also contribute to a large literature on gender and labor markets (e.g., Goldin, 2014, Bertrand, 2011) by studying how social norms impact FLFP, and our work is thus related to a growing literature on the effects of gender identity

---

[8]The list experiment works as follows: first, respondents are (randomly) assigned either into a control group or to one or to a treatment group. Subjects in all conditions are asked to indicate the *number* of policy positions they support from a list of positions on several issues. Support for any particular policy position is never indicated, only the total number of positions articulated on the list that a subject supports. In the control condition, the list includes a set of contentious, but *not* stigmatized, policy positions. In the treatment condition, the list includes the contentious policy positions from the control list, but also adds the policy position of interest (support for FLFP), which *is* potentially stigmatized. The degree of support for the stigmatized position *at the sample level* is determined by comparing the average number of issues supported in the treatment and control conditions. For another recent application of list experiments, see Enikolopov et al. (2017).

[9]The question is designed to elicit opinions regarding whether women should be allowed to work outside the home (as opposed to views on whether have the right (from a legal sense) to work outside the home. Participants were asked to report their extent of agreement: strongly agree/agree/disagree/strongly disagree. We pool the first two the create the indicator of agreement.

[10]Our paper also speaks to a recent theoretical literature on social norms (e.g., Bénabou and Tirole, 2011, Ali and Bénabou, 2016, and Acemoglu and Jackson, Forthcoming) by documenting how new information may lead to updates in perceptions of norms and fast changes in behavior.

and norms on economic outcomes (see Alesina et al. (2013), Akerlof and Kranton (2000), Baldiga (2014), Bernhardt et al. (2018), Bertrand et al. (2015), Bordalo et al. (2016), Bursztyn et al. al (2017), Coffman (2014), Dohmen et al. (2011), Eckel and Grossman (2008), Fernández and Fogli (2009), Fernández (2004), Fernández (2007) and Field et al. (2016), and Jayachandran (2015) for a discussion of the literature studying the role of social norms in explaining gender inequality in developing countries). Our paper relates to the work by Fernández (2013), which studies the role of cultural changes in explaining the large increases in married women's labor force participation over the last century in the US. Our work adds to a growing literature on social image concerns in economics. Individuals' concerns about how they will be viewed by others has been shown to affect important decisions, from voting (DellaVigna et al. 2017, Perez-Truglia and Cruces 2017) to charitable donations (DellaVigna et al. 2012) to schooling choices (Bursztyn and Jensen 2015). We show that Saudi men's decisions to let their wives work are also affected by perceptions of the likelihood of judgment by others.

On the policy side, our results highlight how simple information provision might change perceptions of a society's opinions on important topics, and how this might eventually lead to changes in behavior. Conducting opinion polls and diffusing information about their findings could potentially be used to change important behaviors in some societies.

The remainder of this paper proceeds as follows. We discuss the experimental design of our main experiment and the underlying conceptual framework in Section 2. In Section 3, we present and interpret the results from that experiment. In Section 4, we present the design and results from the national online survey and discuss evidence from the *Arab Barometer* survey. Section 5 concludes.

2 Main Experiment: Design

2.1 Experimental Design

To organize thoughts and help guide our design, in Appendix A we present a simple model of labor supply and stigma, following the intuition from Bursztyn, Egorov, and Fiorin (2017). If married Saudi men believe that a large share of other men are opposed to FLFP and if they care to a great extent about their social image, they may end up not letting their wives work outside of home, despite the fact that most would prefer to allow their wives to work if the behavior were not observable. However, our model predicts that correcting perceptions about the opinions of others leads to drastic changes in the share of men willing to let their wives work.

Our experiment is comprised of two stages. We first conducted an experiment in the field to establish the effect of correcting participants' perceptions of the beliefs of others on a contemporaneous decision to sign up their wife for a job matching service. We then administered a follow-up survey three to five months later to collect information on longer-term labor supply outcomes.

2.1.1 Sample and Recruitment

We partnered with a local branch of an international survey company and recruited 500 Saudi males between the ages of 18 and 35 living across Riyadh, Saudi Arabia as participants in our main experiment.[11] Participants were required to have at least some college education as well as access to a smartphone.[12] We additionally restricted our recruitment to candidates who were married to ensure that participants would be able to make decisions regarding the labor market participation of their wives.

A recruiter database was used for initial contacts and further recruitment was conducted using a combination of snowball sampling and random street intercept. Participants were recruited from districts representing a range of socio-economic classes. Anticipating that contacts from districts with lower average incomes would be more responsive to the offered incentives, these districts were oversampled.[13]

Participants were organized into 17 experimental sessions of 30 participants each.[14] Importantly, participants for each session were recruited from the same geographical area, so that participants in the same session shared a common social network–the average participant reported knowing 15 of the 29 other session participants. Sessions were held in conference rooms located in a Riyadh hotel, administered over the course of a week starting Oct. 9, 2017 and ending Oct. 13, 2017.

2.1.2 Main experiment

On their scheduled session date, participants arrived at the hotel. Upon arrival, participants were asked to provide their name, phone number and email on a sign-in sheet before entering the designated session room. For each of the sessions, 5 rows of 6 chairs each were set up, with every other chair rotated to face the back of the room so that respondents' survey responses would not be seen by nearby participants. A survey facilitator instructed participants to sit quietly until the start of the survey, at which point all participants were instructed to begin the survey simultaneously. Participants were not allowed to ask further questions after the start of the survey.

The survey itself was administered using the online survey software Qualtrics and was imple-

---

[11] An additional 120 participants were recruited for a pilot study consisting of 4 sessions which took place right before the start of the main experiment. The pilot study provided important logistical experience for survey facilitators and results were used to inform the final experimental design. Data from the pilot sessions are not included in the results presented.

[12] Access to a smartphone was required not only so that participants could take the survey on their own devices, but also so we can ask participants whether they would like to sign up for a job matching service which includes a mobile app.

[13] Participants were offered gift certificates with values ranging from 100-150 SAR (26-40 USD).

[14] This means 510 total participants were expected, but only 500 completed the survey. Each session included two sequential survey links for each participant: the first part of the survey (corresponding to the first link) simply contained questions; the second part contained the informational conditions and the outcomes. The 10-subject attrition in the experiment is driven by participants who failed to activate the second link.

mented in two parts.[15] In the first part of the survey, we collected demographic information and elicited participants' own opinions about a range of topics as well as their incentivized perceptions of others' beliefs. In the second part of the survey, we randomized participants to our information provision treatment and measure outcomes. At the start of the survey period, a survey link to the first part of the survey was provided on a board at the front of the room. Participants were instructed to navigate to the link and take the survey on their personal smartphones.[16]

Since FLFP may be a sensitive issue for respondents, maintaining anonymity of responses is an important focus of our experimental design. While the names and phone numbers of all participants were collected on the sign-in sheets, this identifying information was *not* collected in the survey itself. As a result, participants' names and full phone numbers are not associated with their individual survey responses at any point in the data collection or analysis process.[17] Instead, participants were asked only to provide the last three digits of their phone number. We then used these trailing digits to randomize participants to treatment conditions, allowing us to recover a given participant's assigned treatment status using only their last three phone number digits.[18] In addition, this also allows us to link participants' follow-up responses to their responses in the main experiment using only the combination of session number and these trailing digits.[19] In sum, only the local surveyors collected a list with names and full phone names for each session. However, these surveyors were not able to link names to answers since they never had access to the data we collected through the online platform. Meanwhile, the researchers have access to the data, but do not have access to the participants' contact list—just the last three digits of their phone numbers. This anonymous design was chosen to facilitate the elicitation of honest opinions regarding FLFP.

In addition to using an anonymous online survey, we attempted to additionally reduce the scope for social desirability bias (SBS), by avoiding priming effects. The study was framed as a general labor market survey, with filler questions asking opinions on the employment insurance system, privileging Saudi nationals over foreigners into job vacancies, and the minimum wage level. In addition, no Western non-Arab was present during the intervention.[20] In section 4 we present

---

[15] All the scripts from the different interventions are included in the Appendix.

[16] While the majority of participants were able to take the survey successfully on their own devices, tablets were provided to those who encountered technical difficulties.

[17] Participants were asked to provide an email address in order to receive a gift card reward for correctly guessing others' beliefs as well as the contact information of their wife if they choose to sign her up for a job matching service. All requests for identifying information occurred after the elicitation of private opinions and perceptions about the beliefs of others. Providing this contact information was also not required to complete the survey.

[18] In particular, we electronically randomized treatment values to all possible combinations of three digits (000-999) before the start of the experiment. Participants were then assigned to the treatment condition corresponding to the treatment value pre-assigned to the last three digits of their phone number.

[19] Provided that trailing digits identify participants one-to-one. Phone numbers are recorded on session specific sign-in sheets so that we only need to worry about matching digits within sessions; in practice we find that 12 within-session pairs have matching trailing digits. Since we have no way to match the follow-up responses of these respondents to their survey responses, we drop these respondents in the analysis of the follow-up calls, and these participants were not recontacted.

[20] SBS would be an important issue if it leads to an upward bias in reported pro-FLFP opinions. We note that it

results from an alternative approach to eliciting opinions regarding FLFP using a method that provides a higher degree of plausible deniability: the "list experiment." The very similar findings there indicate that SBS is not likely to be an issue in our setting.

After collecting basic demographic information, participants were presented with a series of statements. For each statement, we asked the respondent whether they agreed or disagreed with the statement. We began with three statements regarding the labor market in general. These statements were:

- In my opinion, Saudi nationals should receive privileged access to job vacancies before expatriate workers.

- In my opinion, the current unemployment insurance system (Haafez) is good for the economy.

- In my opinion, the minimum wage for Saudis (SAR 3000) should be kept at its current level.

We then presented two statements regarding the participation of women in the labor force:

- In my opinion, women should be allowed to work outside of the home.

- In my opinion, a woman should have the right to work in semi-segregated environments.

For each of the statements regarding FLFP as well as the statement about the minimum wage, participants were asked to estimate how many of the other 29 participants they expected to agree with the statement. To incentivize participants to estimate the beliefs of others accurately, participants were told that the respondent with the closest guess would receive a $20 Amazon gift card code (distributed by email after the end of the session).

This concluded the first part of the survey. Upon completing this part of the survey, participants were instructed to wait for the facilitator's permission before continuing. The session facilitator was instructed to closely monitor the progress of the survey responses through the online platform. After it had been determined that all participants had completed the first part of the survey, participants were instructed to continue on to the second part of the survey by following a link provided on the end page of the first survey.

In the second part of the survey, half the participants were randomly assigned to a treatment condition. This randomization was conducted at the individual level, based on the last three digits of the respondents' phone number (provided in the first part of the survey). In the treatment condition, participants were given feedback about the responses of the other session participants to the two statements about female labor participation. In particular, we provided treatment participants

---

is unclear in what direction the bias would go: if participants are afraid of stigma associated with being pro-FLFP in the case that their neighbors discover their true opinion, we would be *underestimating* the true level of support.

with charts embedded in the survey interface showing the proportion of respondents who reported agreeing and disagreeing with each statement in the first part of the survey. Participants assigned to the control group received no information.

2.1.3 Outcomes

All participants were then asked to read a short passage about a Saudi startup which provides an online platform aiming to connect job-seeking Saudi women with employers. The startup's platform lists thousands of job opportunities for Saudi women, matching workers to jobs based on skills and interests. The informational passage contained basic information about the online platform as well as the company's outreach and mentorship initiatives.

We then asked respondents to make an incentivized choice between receiving a $5 Amazon gift card and the opportunity to sign up their wives for access to the company's platform and services. An important service in the case of sign-up is a weekly email to be received listing a number of links to descriptions and application forms for postings of jobs outside of home for women in the participants' areas. Participants who chose to sign up for the service were subsequently asked to provide the contact information of their wife.[21] Note that since a participant's wife's eventual employment status is observable to peers, the observability of the sign-up choice itself does not matter independently, so we chose not to randomize the observability of the sign-up decision.

If participants who privately support FLFP underestimate the level of support for FLFP among others, they may incorrectly expect that signing up their wives for the job matching service is stigmatized. If this is indeed the case, participants who receive information correcting their beliefs about others' support for FLFP (treatment) should exhibit a higher sign-up rate than participants who receive no information (control).

At the end of our survey, we ask participants to estimate the percent of private sector firms in Saudi Arabia that have semi-segregated work environments as a way to proxy for perceived market demand for female labor.[22] Observing that our information provision has a positive effect on the expected percentage of firms with semi-segregated environments would provide evidence consistent with perceived female labor demand being a function of perceived social norms regarding female labor force participation. We use this information to assess the extent to which the decision to sign up is driven by changes in perceptions about labor demand—a different mechanism that is potentially relevant.

Figure 1 summarizes the design of the main experiment.

---

[21]In the results we present, as pre-registered, those who do not provide their wife's name and number (even if initially choosing the sign-up option) are treated as having chosen not to sign up. Results are almost identical if we include those as effectively signing up.

[22]Semi-segregated workplaces have a separate section for women.

2.1.4 Follow-up Calls: Longer-Term Outcomes

We conducted follow-up calls three to five months after the conclusion of the main experiment in order to collect longer-term labor market outcomes.[23] We again partnered with the same survey company to re-contact participants and conduct the follow-up survey.

Participants were contacted using the phone numbers collected on the session sign-in sheets during the main experiment. Participants who were not reached initially were called several times. At the beginning of each call, the surveyor reminded the respondent of his participation in the main experiment and provided further information about the follow-up call. Participants who consented to participate in the follow-up were offered a payment of up to 35 SAR in phone credit for successful completion of the survey.

As noted above, we link the data collected in the follow-up call to individual survey data in the main experiment using the last three digits of the respondent's phone number. Twelve pairs of respondents in the same experimental session provided identical trailing digits. We drop these participants from the subsequent follow-up analysis since we cannot uniquely match them to their survey responses. Surveyors were able to recontact 82% of the original experimental subjects with unique trailing digits, and 98% of recontacted participants completed the survey. We find no evidence of selective attrition overall, or within initial conditions (see Appendix Tables A1 and A2), both based on observables and on the sign-up decision in the original experiment. We note that it is possible that some respondents changed their cell phone numbers in the four months between the experiment and the follow-up survey, which could have contributed to the attrition rate (a number of numbers called were disconnected, which could have resulted from changes in telephone service or, alternatively, to participants providing incorrect numbers in during the experiment). In Section 3.1.3, we provide Lee (2009) attrition bounds for our treatment effects.

In each follow-up call, we collected a variety of outcomes related to the wife's labor supply outcomes. We asked whether the wife was currently employed and if so, whether this job was outside of home. Jobs outside of home are our outcome of interest since the information intervention focuses on norms about working outside of home, and the job-matching service focuses on access to jobs of this kind. We also asked whether the wife was employed 3-4 months ago (depending on the follow-up survey data), i.e., at the time of the main experiment. We did this for two reasons. First, since we had already asked whether the wife was employed during the main experiment itself, asking the question again in the follow-up allowed us to compare responses to the two questions as a sanity check for internal consistency. Reassuringly, 94% of respondents respond consistently, giving the same answer to the question in both instances. Second, we were initially surprised that almost two-thirds (65%) of respondents in the main experiment reported that their wife was employed. This high proportion led us to question whether some of the respondents' wives were in fact working

---

[23]Follow-up calls began on Jan. 10, 2018 and ended Mar. 6, 2018.

from home. We used the follow-up survey as an opportunity to ask whether respondents' wives were employed *outside* of home. We find that while 61% of the respondents in the follow-up survey reported that their wives were working three months before (i.e., around the time of the original experiment), only 8% of them reported that their wives were then working *outside of home*.[24] Participants were also given the option to answer an open-ended question asking for their wives' exact occupation. However, only a small share of participants provided an answer to this question, so we interpret the numbers with caution. The most common occupation categories for wives working from home were customer service – via phone calls (51% of them) and data entry (20% of them).

We then asked whether their wife participated in any job search activities in the last 3-4 months. In particular, we collected information on whether their wife applied for a job during this period, interviewed for a job during this period and whether she had any interviews currently scheduled. For all questions, affirmative answers were followed by a question asking whether the job of interest was outside of home. Next, we considered whether the information intervention in the original experiment may have had spillover effects on related behavioral outcomes. Taking advantage of the recent announcements lifting the ban on women's right to drive, we posed a hypothetical question asking participants if they would be willing to sign up their wife for driving lessons if given the chance. Finally, we checked for persistence in the correction of perceptions of others' beliefs induced by the information provision treatment. In particular, we asked participants to guess how many people out of 30 randomly selected residents of their neighborhood would agree with each of the statements presented in the main experiment. Finding systematic differences in perceptions about norms between the treatment group and the control group would suggest that the information provision treatment had a persistent effect on participants' beliefs about others. Since participants were recruited from the same neighborhoods within sessions, information spillovers from treatment to control participants could have washed out differential persistence in perceived norms.

Due to the number of outcomes we collect, we address the issue of multiple hypothesis testing by constructing an index pooling all of our collected outcomes following Kling et al. (2007). We discuss the construction of this index in Section 3.1.3. We also adjust p-values for multiple testing following the procedure developed by List, Shaikh and Xu (2016).

---

[24]In the larger, online national survey we conducted later, 43% of participants reported that their wife was currently working, while only 10% reported that their wife was working outside of home.

3 Main Experiment: Results

3.1 Main Results

3.1.1 Misperceived Social Norms

We start by describing the measured misperception about social norms relating to FLFP. Figure 2 displays the CDF of incentivized guesses about the responses of other session participants across all experimental sessions, the CDF of true shares by session, and the true share of respondents across all sessions who agreed with the main statement regarding FLFP ("work outside of home"). The average (median) guess is that 63% (67%) of other session participants agree with the pro-FLFP statement. The average level of agreement across all sessions is 87%, a number larger than the guesses of close to 70% of participants. Of course, the proper comparison is between a participant's guess and the share agreeing *in his session*. In Figure 3, we calculate the wedge between the participant's belief about the opinions of other session subjects and the actual opinions of all session participants, and then plot the distribution of these wedges across all sessions. We find that 72% of participants strictly underestimate the support for FLFP among session participants (noting that these are individuals recruited from the same neighborhoods).[25]

To further examine what is underlying these guesses, we also asked people to assess on a 1-5 scale how confident they were with their guesses. We also asked how many other people they knew in the room. In Appendix Figure A3, we show that more accurate guesses are significantly correlated with higher levels of confidence in the guess and that knowing more people in the room is significantly correlated with both higher levels of confidence in the guess and with more accurate guesses. These are just suggestive correlations, but they are consistent with our hypothesis that these Saudi men attempt to learn the acceptability of female labor supply behavior from the people around them. We make further use of the information on the confidence level of guesses in section 3.2.

The information collected in the follow-up survey on whether wives were working outside of home at the time of the original experiment allows us to conduct another validity check: 100% of participants whose wife was working outside of home reported supporting FLFP outside of home. We also note that participants supportive of FLFP outside of home (the vast majority of participants in the sample) have a significantly higher perception of support by other session subjects than participants opposed to FLFP outside of home (average guesses of 18.7 vs 13.4, respectively, with p=0.000).

---

[25]Appendix Figures A1 and A2 display very similar findings for the second question relating to FLFP ("semi-segregated environments").

3.1.2 Main Results (1): Job-Matching Service Sign-up Rates

We now turn to our experimental intervention correcting beliefs about the opinions of other participants. Appendix Table A3 provides evidence that individual characteristics are balanced across the two experimental conditions, confirming that the randomization was successful.

Figure 4 displays our main findings. In the control condition, 23% of participants prefer to sign up their wives for the job service as opposed to taking the gift card payment. In the group receiving feedback on the responses of other participants to the questions on support for FLFP, the sign-up rate goes up to 32%, a 39% increase (the p-value of a t test of equality is 0.017). Table 1 displays the findings in regression format: column (1) replicates the findings from Figure 4. In column (2), we add session fixed effects, column (3) controls for baseline beliefs and column (4) includes additional individual covariates. Results are unchanged across specifications. In this table and all tables which follow, in addition to p-values from robust standard errors, we also present p-values from wild cluster bootstrap standard errors (where we cluster at the session level) and permutation tests for relevant coefficients.[26]

3.1.3 Main Results (2): Longer-Term Outcomes

We now evaluate the longer-term effects of the information intervention (three to five months later). Figure 5 displays the results. We first present the impact of the original treatment on three labor-supply outcomes: whether or not the participant's wife applied for a job outside of home since the original intervention, whether or not she interviewed for a job outside of home during this period and whether or not she is currently employed outside of home. Across all outcomes, we see increases stemming from the original treatment, though this difference is significant only for the first two outcomes. The percentage of wives who applied for a job outside of home during the time frame of interest goes up from 5.8% to 16.2%, a 179% increase (the p-value of a t test of equality is 0.001). Regarding interviews for this type of job, the share increases from 1.1% to 5.8% (p-value=0.06). Rates of employment outside of home go up from 7.4% to 9.4% (p-value=0.235). If we restrict the analysis to wives who were not employed outside of home at the time of the intervention (92% of them), the effect is an increase from 1.1% to 2.8% (p-value=0.25).

We also examine whether the information provided might lead to changes in other behaviors related to women's rights in Saudi Arabia. Figure 5 shows that the share of husbands who report in a hypothetical question that they would sign their wives up for driving lessons goes up from 65.2% to 76.4% (p-value=0.008). This suggests that the effects of the information intervention extend beyond behaviors narrowly associated with FLFP outside of home, and may act more generally with respect to norms about women's rights in Saudi Arabia.

---

[26]We use wild cluster bootstrap standard errors because of the small number of clusters (Cameron, Gelbach and Miller 2008). We use permutation tests due to potentially small sample sizes.

We next assess whether the original treatment produced a persistent change in perceived norms (and whether spillovers via subsequent communication with the control group undid the original intervention). To that end, we asked participants to provide a non-incentivized guess on the number of 30 randomly selected people from their neighborhood they think would agree with the same pro-FLFP statements from the original experiment. We find evidence of a persistent change in perceptions by treated relative to control participants: the average guessed share for "working outside of home" increases from 43.9% to 56.6% (p-value=0.000). These findings also suggest that the intervention updated beliefs about neighbor's opinions in general—not just regarding responses by other session participants.

To deal with the potential issue of multiple testing, we follow Kling et al. (2007), and construct an index pooling all six of our used outcomes.[27] The index is $(1/6) \sum_{k=1}^{6} (k - q_k)/\sigma_k$, where k indexes an outcome, and q_k and σ_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 multiple testing following the procedure developed by List, Shaikh and Xu (2016). Finally, we provide Lee (2009) attrition bounds for our treatment effects.

Table 2 presents the results in regression format, with and without controls.

We can further examine the longer-term effects of the information provided on perceptions of social norms. Appendix Figure A4 provides visual evidence that for control group participants, higher beliefs about the share of pro-FLFP session participants are associated with higher beliefs about the share of pro-FLFP neighbors in the follow-up survey (the effect is significant at the 1% level). The figure also shows that the levels of beliefs in the follow-up survey are higher for treated participants, and that the relationship between prior and posterior beliefs is (significantly) flatter for treated participants. These results suggest that the treatment indeed updated priors about neighbors' opinions and that treated participants used the signal from the treatment to override their priors.

3.2 Heterogeneity by Direction of Priors' Update

We move beyond the average effects of the information provision and now analyze heterogeneous effects by the direction (and size) of priors' updating. Figure 6 provides the raw averages and Table 3 displays regression results. For those who originally underestimated the true share of session participants agreeing with the statement regarding women's right to work outside of home, we observe a large increase in sign-up rates for the app when information is provided: the share signing up increases from 21.4% in the control group to 33.5% in the treatment group (a 57%

---

[27]The index here includes six outcomes (interviews scheduled is included in the index but not in the tables since only 1 person had an interview scheduled). Results are unchanged if we include as a seventh outcome guesses about opinions regarding the statement on semi-segregated environments, since the answers were very correlated for both types of guesses.

increase, p-value=0.004). For those who originally overestimated the true share of session subjects agreeing with the statement, the baseline sign-up rate is higher in the control group (30.9%), which is by itself consistent with the hypothesis of a role of perceived social norms affecting labor supply decisions for their wives. The effect of the intervention is no longer significant for that group, and is in fact directionally negative (albeit small). We note that the interaction term in the regression including the whole sample (that is, the treatment interacted with a dummy on the sign of the wedge) is not significant at conventional levels (the p-value of the interaction in the regression without additional covariates is 0.115, while it is 0.204 for the full specification).[28] Interestingly, the sign-up rate among control participants with positive wedges is very similar (and not statistically different) from the sign-up rate in the treatment group, which provides correlational evidence from the control group consistent with our hypothesis that perceptions of social norms matter for the sign-up decision. The evidence on heterogeneous updating for the longer-term outcomes is less conclusive, and is shown in Tables A4 (without controls) and A5 (full specification).[29]

If one assumes that treated participants use the numbers given by the information provision as their posterior beliefs about the other session subjects, we can compute a *continuous* measure of belief update during the experiment. For control participants, this update is equal to zero. For treatment participants, it is equal to the negative of their "wedge." We can then evaluate the impact of the *size of the prior updating* on the sign-up decision, as displayed in Table 4. Note that unlike in the other tables, we control for participants' own views on FLFP outside of home, their beliefs about other session subjects, and their confidence in their own guesses, since conceptually all these variables should matter for a given prior updating level. We find that higher levels of prior updating lead to significantly higher sign-up rates. For example, using the specification in column (1), a positive update in prior corresponding to one standard deviation of the wedge in the treatment group causally leads to a 6.9 percentage point increase in the app sign-up rate. In Appendix Table A6, we provide suggestive evidence that higher levels of updating also lead to larger changes in the longer-term outcomes.

---

[28]Also note that, as discussed before, participants with larger negative wedges are more likely to be personally opposed to FLFP outside of home, which in principle would reduce the scope for an effect of the information provision to them. When we restrict the regression to those who reported supporting FLFP, the interaction term yields p-values of 0.099 (no additional controls) and 0.209 (full specification).

[29]We find similar treatment effects on sign-up rates for participants with fewer or more social connections in the original experimental session. We note though that *a priori*, it is unclear whether the effects should be stronger or weaker according to the number of social connections: the treatment might be less informative because socially connected participants might already have better guesses about the opinions of others, but those participants might potentially care more about the opinions of the other subjects, thus strengthening the treatment.

3.3 Interpreting the Results

3.3.1 Understanding the Longer-Term Effects

It is difficult to separate the extent to which the longer-term effects are driven by the higher rate of access to the job service vs. a persistent change in perceptions of the stigma associated with FLFP outside of home. The fact that we observe a persistent treatment effect on beliefs about neighbors measured in the follow-up survey and on the (reported) willingness to sign up one's wife for driving lessons (that is, a different, related outcome) suggests that the longer-term effects on job search behavior might not be entirely mechanically driven by the job matching service alone.

One might wonder if there could have been informational spillovers between treatment and control participants: these participants often reported knowing each other and could have discussed the information given to them after the main experiment. This is particularly related to our hypothesis: if perceived social sanctions associated with FLFP outside of home have been reduced, some participants would be more open to discussing the topic with their neighbors or friends. While we observe persistent differences in beliefs about others in the follow-up survey, we have no way of knowing if the differences between control and treatment groups would have been even larger in the absence of potential spillovers.[30]

Interestingly, we find significant effects for outcomes (applying and interviewing for jobs outside of home) that are not easily observable (when compared to working outside of home itself). It is possible that the difference in beliefs about others between the two conditions would have been smaller if control participants started observing more women working outside of home. Unfortunately, we do not have the data to test this hypothesis.

3.3.2 Signaling Labor Demand

The channel we have in mind to explain our results is an update in the perceived social sanctions associated with having a wife working outside of home. However, updating one's beliefs about what others think regarding FLFP might also turn on a potential additional channel, since this update may also lead to learning about the labor demand side. Subjects may think: "If so many people in fact support FLFP, then there are probably many firms willing to hire women." This is an interesting alternative mechanism and could explain our findings if: i) inducing positive updates in perceptions about other participants' opinions leads to positive updates in perceptions about

---

[30]We find similar treatment effects on longer-term outcomes for participants with fewer or more social connections in the original experimental session, suggesting that informational spillovers alone might not have been enough to induce more job search among control participants' wives. Possible reasons for this pattern include: a) the other, known participants were not very close friends but their opinions are still relevant signals about the relevant peer group in the neighborhood; b) perhaps one has no economic incentives to tell one's friends until one's wife gets a job because communicating to others would increase competition for the same jobs; c) if treated participants now think that FLFP outside of home is socially desirable and a symbol of status, they may want to be the first to have an employed wife (to distinguish themselves), or derive utility from being the "leader" and having others follow them.

the availability of jobs for women *and* ii) this second update matters for the sign-up decision (i.e., participants originally believed labor demand was a constraint). In Appendix Table A7, we find that the treatment leads to a small (and generally not-significant) increase in participants' beliefs about the percentage of private-sector firms in Saudi Arabia with semi-segregated environments. In Appendix Table A8, we show that once we include covariates, there is no correlation in the control group between participants' perception of firms' willingness to hire women and participants' decision to sign up for the job-matching app. We therefore believe that this alternative mechanism is not driving our findings.

4 National Survey

We now examine results from an anonymous, online survey with a larger, national sample of about 1,500 married Saudi men that used the same platform and had the same layout as the original experiment. The goal of this additional survey is two-fold. First, we assess the external validity of the finding that most Saudi men privately support FLFP while failing to understand that others do. Second, we examine whether social desirability bias/experimenter demand effects could have been a driver of the misperception finding in the main experiment.

4.1 Survey Design

The sample of survey takers was recruited through the same survey company as before and the (visually identical) survey was administered online through Qualtrics. As in the main experiment, we restricted the survey to married Saudi males between the ages of 18 and 35. The final sample of participants was designed to be a nationally representative sample of the targeted demographic category, enabling us to provide evidence of the external validity of our main finding.

After signing an online consent form, participants were asked the same set of demographic questions as in the main experiment. Then, we departed from the design of the main experiment by implementing a list experiment to introduce plausible deniability to our elicitation of individual beliefs. In a list experiment, participants are randomly assigned to either a control group or to a treatment group. In the control group, participants are presented with a list of statements or policy positions that are contentious but not stigmatized. In the treatment group, participants are presented with an identical list of items but also an additional, potentially stigmatized item for which the experimenter would like to elicit beliefs. Participants in both conditions are then asked to indicate how many of the statements they agree with. The true degree of support for the item of interest at the sample level can then be inferred by comparing the average number of agreements in the treatment group with the average number of agreements in the control group.

This design allows the experimenter to ask participants only *how many* of the statements or policy positions in a given list they agree with rather than needing to know which items they

support in particular. In our case, as long as participants do not disagree with all the non-FLFP statements, the list experiment provides plausible deniability to those who do not support FLFP and might otherwise be affected by social desirability bias.

We operationalized the list experiment in our online survey as follows. Participants in the control group were presented with a list of three general statements about the labor market (the same statements used in the main experiment).[31] These statements were chosen to be contentious but *not* stigmatized.

- In my opinion, Saudi nationals should receive privileged access to job vacancies before expatriate workers.

- In my opinion, the current unemployment insurance system (Haafez) is good for the economy.

- In my opinion, the minimum wage for Saudis (SAR 3000) should be kept at its current level.

Participants were asked to read all three statements carefully and to indicate the number (from 0 to 3) that they agreed with. Note that we did not ask participants which of the statements they agreed with, only how many.

In the treatment group, participants were presented with the same list of statements with the single addition of the statement of interest regarding FLFP that *is* potentially stigmatized:

- In my opinion, women should be allowed to work outside of the home.

Treatment participants were likewise asked to indicate the number of statements (from 0 to 4) that they agreed with. We recover the true share of participants supporting FLFP by subtracting the average number of stated agreements to the three non-FLFP statements (given by the control group mean) from the average number of stated agreements to the list including the FLFP statement (given by the treatment group mean).

Next, we directly elicited private opinions about FLFP by asking participants in the control group whether they agreed with the statement about the right of women to work outside the home. This enables us to compare, at the sample-level, the *stated* degree of support for FLFP with the degree of support for FLFP *inferred* using the list experiment. Since the latter should be much less susceptible to social desirability bias, this comparison reveals the magnitude of any potential social desirability bias effects.

Finally, we ask participants to gauge the responses of other participants to the statement about FLFP. In the control condition, participants are asked to estimate the percentage of other participants who *reported* agreeing with the same statement. Participants are informed that the study

---

[31]The ordering of the statements in the list was randomized for both the treatment and the control groups.

18

consists of a nationally representative sample of 1,500 married Saudi males aged 18-35. As in the main experiment, this elicitation was conducted in an incentivized manner–participants were informed that the respondent with the most accurate guess would receive a \$50 Amazon gift card.

Since we do not directly elicit private beliefs about the FLFP statement in the treatment condition we cannot elicit participants' beliefs about others in the same way (without revealing the existence of our treatment and control conditions). Instead, we ask participants to estimate the percentage of other participants who would *privately* agree with the FLFP statement.[32] Participants are again informed about the characteristics of the sample. Comparing the distribution of these estimates of the *true opinions* of others to the distribution of estimates of the *answers* of others allows us to examine whether guesses about the answers of others are distorted by expectations that others might answer strategically.

A pilot sample of 100 participants was launched on Feb. 15, 2018 and concluded on Feb. 16, 2018. No changes were made to the experimental design after reviewing the pilot results. The survey was then administered to the rest of the sample (1,496 survey participants, 1,460 of which completed the whole survey) starting on Feb. 17, 2018 and concluding on Mar. 2, 2018. We drop the data from the pilot in our analysis. Appendix Table A10 displays summary statistics for the national survey sample and shows that covariates are balanced across conditions.

4.2 Results

We first assess the external validity of the finding that most Saudi men privately support FLFP while failing to understand that others do. As displayed in Figure 7, in this broader, more representative sample, 82% of men agree with the same statement on FLFP used in the main experiment regarding working outside of home. When incentivized to guess the responses of other survey respondents, 92% of them underestimate the true share. These are stronger misperceptions, perhaps because they are no longer being asked about their own neighbors' opinions.

We next use the survey to examine whether social desirability bias/experimenter demand effects could have been a driver of the misperception finding in the main experiment. These results are also shown in Figure 7. By subtracting the average number of agreements in the control list from the treatment list, we get the share of respondents who agree with the statement of interest, but under response conditions with a higher degree of plausible deniability. We find a very similar share of 80%, indicating that social desirability bias/experimenter demand effects are not a driver of the finding. Finally, we examine whether individuals might be incorrectly expecting others to strategically respond to the FLFP agreement question, which would distort guesses about others, since the question asked about how others *answered* the question. We find that beliefs about other participants' *true opinions* were extremely similar to the guesses about others' *answers*.

---

[32]This was done without incentives since there is no obvious measure of accuracy available.

Finally, we again find that more accurate guesses (for both types of questions regarding others) are correlated with more confidence in the guess, as depicted in Appendix Figure A5.

4.3 Evidence from the *Arab Barometer*

As a last check on the external validity of the fact we document, we find that the share of Saudi men supportive of FLFP outside of home is also extremely similar when using the nationally representative sample from the wave of *Arab Barometer* containing that question for Saudi Arabia (2010-2011). Out of approximately 700 male respondents, 75% agree with the statement "a married woman can work outside the home if she wishes." Among male respondents aged 18-35 (the age bracket in our study), the share is 79%. The *Arab Barometer* survey also allows us to establish that older men are also supportive of FLFP outside of home: among those over 35, the share agreeing with that statement is 71%. Figure 8 displays the share of men supportive of FLFP outside of home in the different samples discussed in this paper.

5 Conclusion

In this paper, we provide evidence that: i) the vast majority of married men in Saudi Arabia privately support female labor force participation; ii) they substantially underestimate the level of support by other men, including their own neighbors; iii) correcting these beliefs about others affects their wives' labor supply decisions. These results might help us understand the role that social norms play in constraining female labor force participation in Saudi Arabia, but also how these constraints might be lifted by the simple provision of information. Active information provision might be particularly important in less democratic societies, where the availability of other natural aggregators of information (such as elections, referenda and even opinion polls) is more limited.

We end with potential avenues for future research. We view our findings as "proof of concept" of the potential for information provision to change behavior regarding female labor force participation in Saudi Arabia (and potentially in other countries). We believe that expanding the scale and observing how information spreads in networks, and how it affects a large set of outcomes is an important topic for future work. Also, the goal of the study was to understand the opinions and perceptions of male guardians in Saudi Arabia. As a result, we did not examine opinions and perceptions of women. Evidence from the *Arab Barometer* suggests that the vast majority of women in the country are supportive of FLFP outside of home (89%). Future work can enrich the analysis on the women's side of the labor supply decision process. Finally, understanding what is at the root of the stigma associated with female labor force participation might help design policies to address it: what are husbands trying to signal to others by acting in opposition to female labor force participation?

References

Acemoglu, Daron and Matthew O. Jackson, "Social Norms and the Enforcement of Laws," *Journal of the European Economic Association*, forthcoming.

Akerlof, George A. and Rachel E. Kranton, "Economics and Identity," *Quarterly Journal of Economics*, 2000, 115, 715-753.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, 2013, 128(2), 469-530.

Ali, S. Nageeb and Roland Bénabou, "Image Versus Information: Changing Societal Norms and Optimal Privacy," 2016, NBER Working Paper No. 22203.

Baldiga, Katherine, "Gender Differences in Willingness to Guess," *Management Science*, 2014, 60(2), 434-448.

Bénabou, Roland and Jean Tirole, "Laws and Norms", 2011, NBER Working Paper No. 17579.

Bernhardt, Arielle, Erica Field, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer-Moore, "Male Social Status and Women's Work," *AEA Papers and Proceedings*, 2018, 108: 363–67.

Bertrand, Marianne, "New Perspectives on Gender," in Orley Ashenfelter and David Card eds., *Handbook of Labor Economics*, 2010, 4B, 1545–1592.

–, **Emir Kamenica, and Jessica Pan,** "Gender Identity and Relative Income Within Households," *Quarterly Journal of Economics*, 2015, 130(2), 571-614.

Bordalo, Pedro, Katherine B. Coffman, Nicola Gennaioli, and Andrei Shleifer, "Beliefs about Gender," 2016, NBER Working Paper No. 22972.

Bursztyn, Leonardo, Georgy Egorov and Stefano Fiorin, "From Extreme to Mainstream: How Social Norms Unravel," 2017, NBER Working Paper No. 23415.

–, **Leonardo, Thomas Fujiwara, and Amanda Pallais,** "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 2017, 107(11), 3288–3319.

–, **and Robert T. Jensen,** "How Does Peer Pressure Affect Educational Investments?," *Quarterly Journal of Economics,* 2015, 130 (3), 1329–1367.

Cameron, A. Colin, Jonah B. Gelbach and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics,* 2008, 90(3), 414–427.

Coffman, Katherine B., "Evidence on Self-Stereotyping and the Contribution of Ideas," *Quarterly Journal of Economics*, 2014, 129(4), 1625–1660.

Correll, Shelley J., Cecilia L. Ridgeway, Ezra W. Zuckerman, Sharon Jank, Sara Jordan-Bloch, Sandra Nakagawa, "Its the Conventional Thought That Counts: How Third-Order Inference Produces Status Advantage," *American Sociological Review*, 2017, 82(2), 297–327.

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao, "Voting to Tell Others," *Review of Economic Studies*, 2017, 84(1), 143–181.

–, John A. List, and Ulrike Malmendier, "Testing for Altruism and Social Pressure in Charitable Giving," *Quarterly Journal of Economics*, 2012, 127, 1–56.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner, "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences," *Journal of the European Economic Association*, 2011, 9(3), 522–550.

Eckel, Catherine C., and Philip J. Grossman, "Men, Women and Risk Aversion: Experimental Evidence," *Handbook of Experimental Economic Results*, ed. by C. R. Plott, and V. L. Smith, Vol. 1, 2008.

Enikolopov, Ruben, Alexey Makarin, Maria Petrova, and Leonid Polishchuk "Social image, Networks, and Protest Participation," *Working paper*, 2017.

Fernández, Raquel, "Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century," *American Economic Review*, 2013, 103(1), 472-500.

–, "Women, Work, and Culture," *Journal of the European Economic Association*, 2007, 5(2-3), 305–332. '

–, –, and Alessandra Fogli, "Culture: An Empirical Investigation of Beliefs, Work, and Fertility," *American Economic Journal: Macroeconomics*, 2009, 1(1), 146–177.

–, Alessandra Fogli, and Claudia Olivetti, "Mothers and Sons: Preference Formation and Female Labor Force Dynamics," *Quarterly Journal of Economics*, 2004, 119(4), 1249–1299.

Field, Erica, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer Moore, "On Her Account: Can Strengthening Women?s Financial Control Boost Female Labor Supply?," *Working Paper*, 2016,

Goldin, Claudia, "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 2014, 104(4)

González, Alessandra L., *Islamic Feminism in Kuwait: The Politics and Paradoxes,* 2014, New York, NY: Palgrave-Macmillan.

Giuliano, Paola, "Living Arrangements in Western Europe: Does Cultural Origin Matter?," *Journal of the European Economic Association*, 2007, 5(5), 927–952.

International Labour Organization, *ILOSTAT database*, retrieved March 2017.

Jayachandran, Seema, "The Roots of Gender Inequality in Developing Countries," *Annual Review of Economics*, 2015, 7, 63–88.

Katz, Daniel and Floyd Allport, "Students' Attitudes," 1931, *Syracuse, Craftsman.*

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, 75(1), 83–119.

Kuran, Timur, *Private Truths, Public Lies,* Cambridge, MA: Harvard University Press, 1995.

–, "The East European Revolution of 1989: Is It Surprising That We Were Surprised?," *American Economic Review*, 1991, 81 (2), 121–25.

List, John A., Azeem M. Shaikh and Yang Xu, "Multiple Hypothesis Testing in Experimental Economics," 2016, NBER Working Paper No. 21875.

O'Gorman, Hubert J., "Pluralistic Ignorance and White Estimates of White Support for Racial Segregation," *Public Opinion Quarterly*, 1975, 39(3), 313–330.

Perez-Truglia, Ricardo and Guillermo Cruces, "Partisan Interactions: Evidence from a Field Experiment in the United States," *Journal of Political Economy*, 2017, 125(4), 1208–1243.

Prentice, Deborah A. and Dale T. Miller, "Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm," *Journal of Personality and Social Psychology,* 1993, 64(2), 243–256.

Raghavarao, Damaraju and Walter T. Federer, "Block Total Response as an Alternative to the Randomized Response Method in Surveys," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 1979, 41 (1), 40–45.

Tankard, Margaret E. and Elizabeth Levy Paluck, "Norm Perception as a Vehicle for Social Change," *Social Issues and Policy Review*, 2016, 10(1), 181–211.

Voigtländer, Nico and Hans-Joachim Voth, "Nazi indoctrination and anti-Semitic beliefs in Germany," *Proceedings of the National Academy of Sciences*, 2015, 112 (26), 7931–7936.

World Economic Forum, *Global Gender Gap Report 2015*, 2015.

Figures and Tables

Figure 1: Experimental Design
(*Main Experiment*)
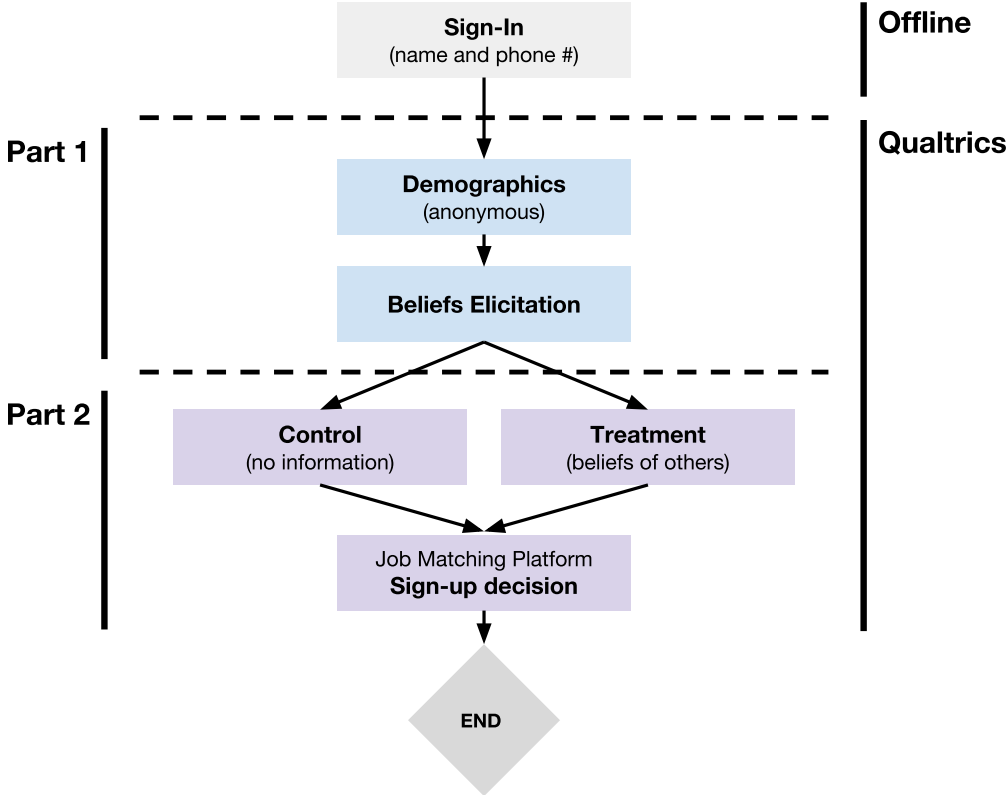
Notes: Experimental design of the main experiment. Dashed lines indicate points in the survey flow where participants were instructed by the session facilitator and/or the survey instructions to wait for all session participants to be ready to start the next section before proceeding.

Figure 2: Misperceptions about Others' Beliefs (Working Outside the Home)
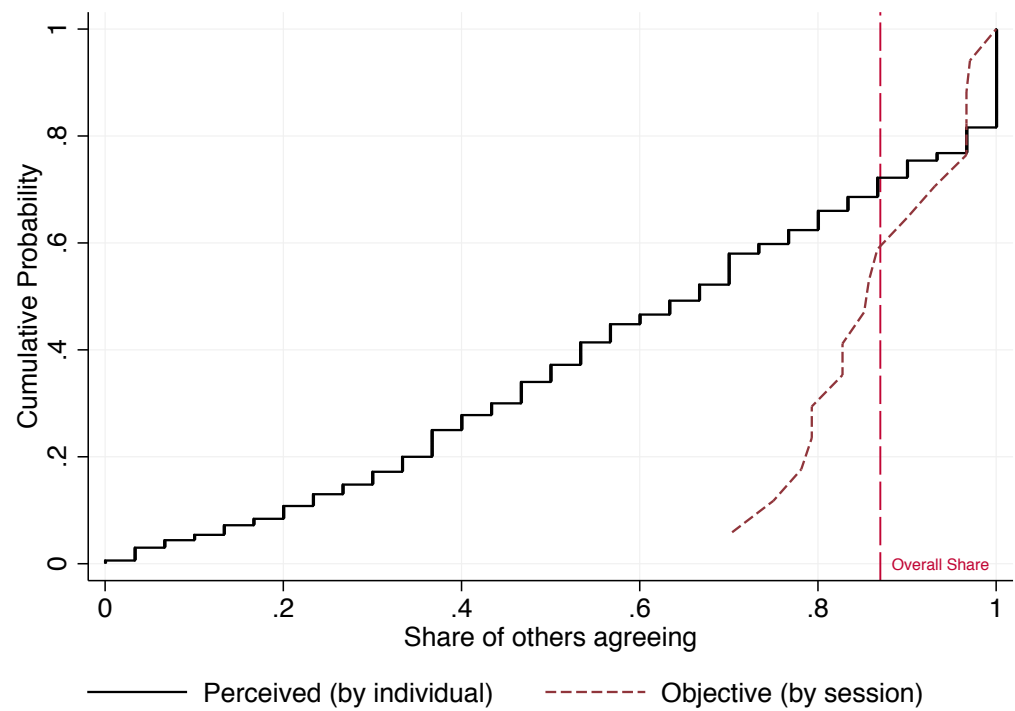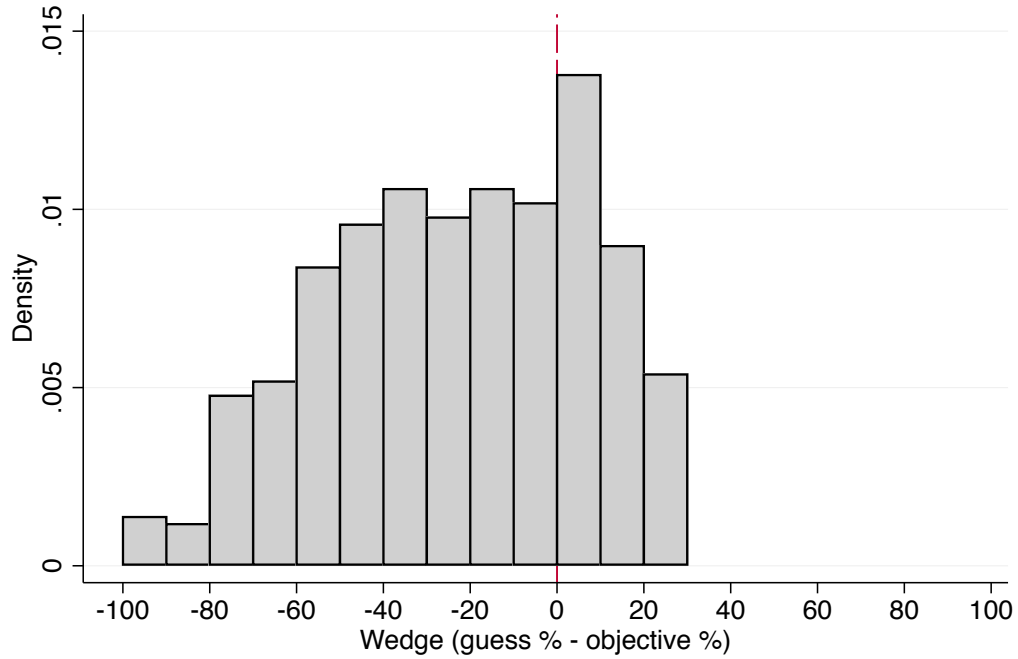(*Main Experiment*)

Figure 3: Wedges in Perceptions of Others' Beliefs (Working Outside the Home)
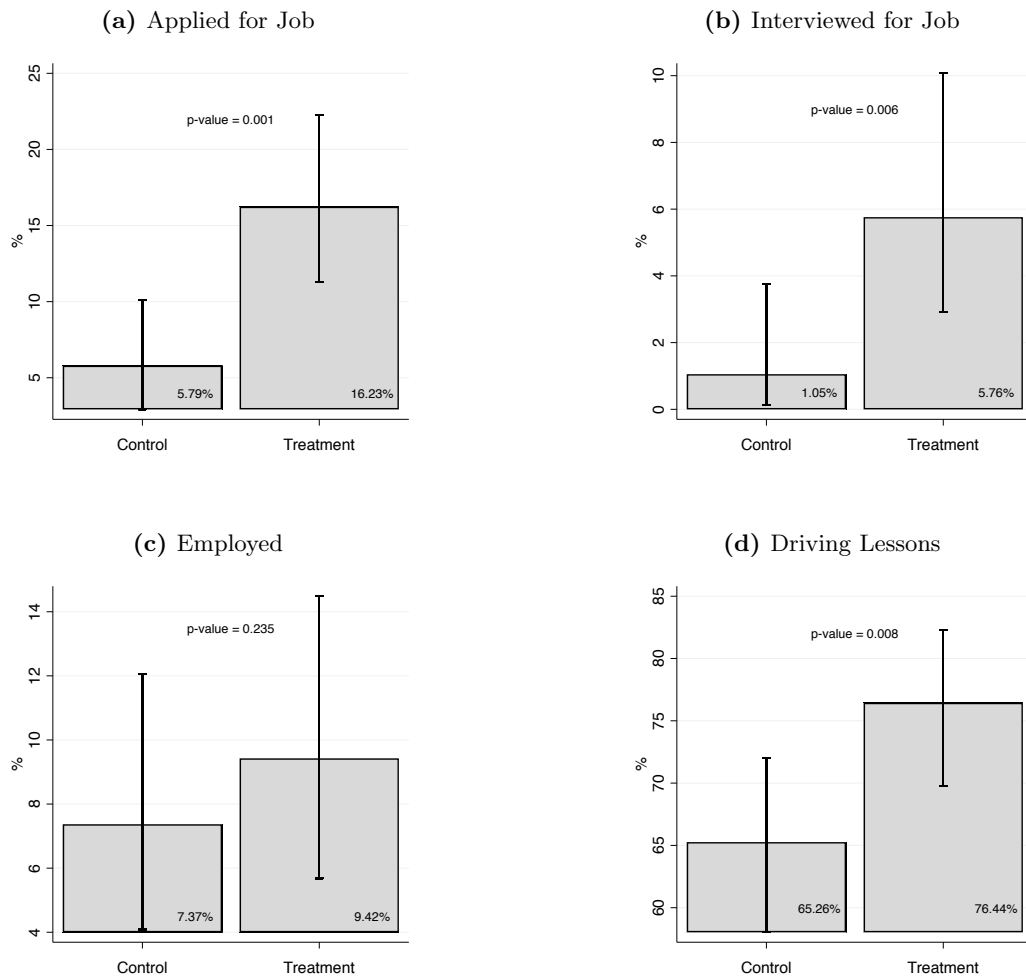(*Main Experiment*)

Notes: The distribution of wedges in perceptions about the beliefs of others regarding whether women should be able to work outside the home. Wedges calculated as (*the respondent's guess about the % of session participants agreeing with the statement*) - (*the true % of session participants agreeing with the statement*).

Figure 4: Job-Matching Service Sign-up
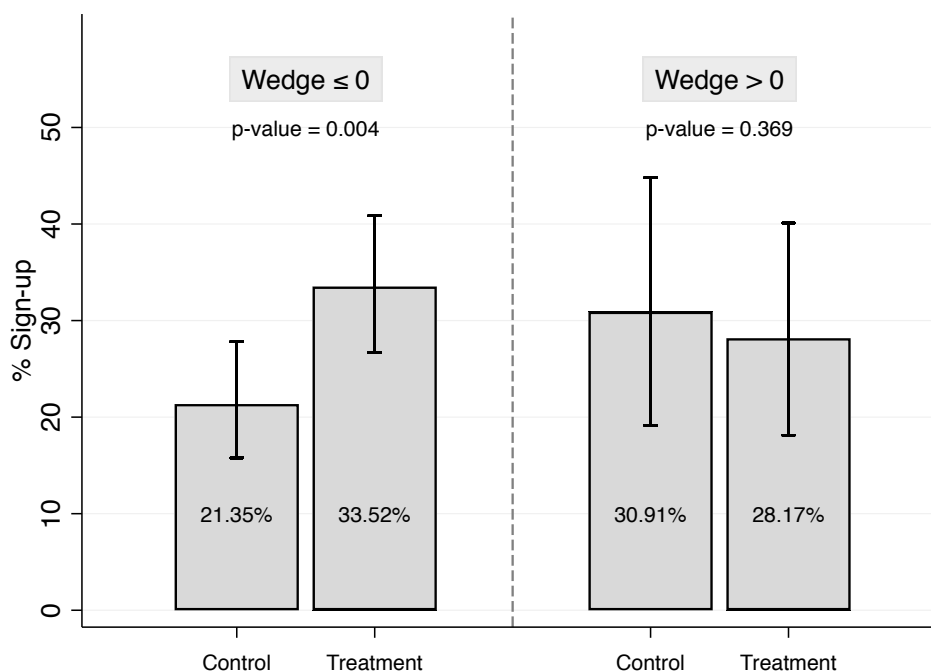(*Main Experiment*)
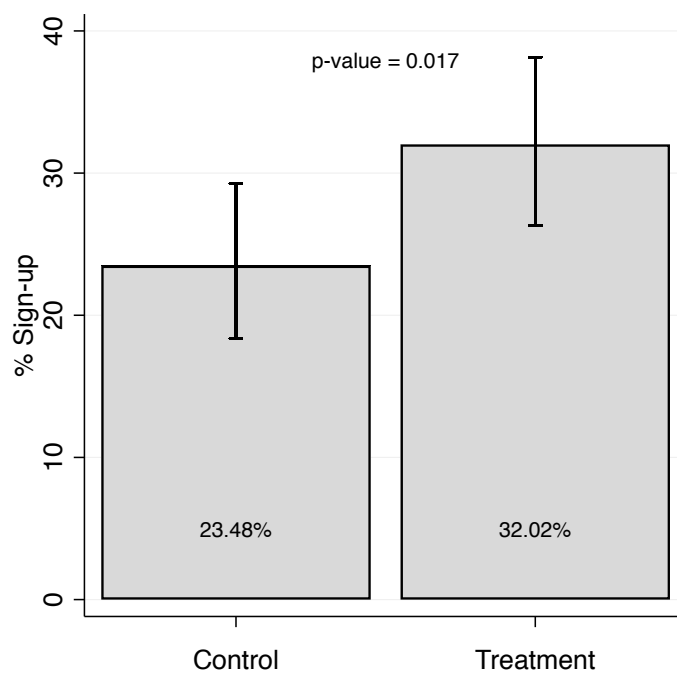
Notes: 95% binomial proportion confidence intervals. *p*-value calculated from testing for equality of proportions.

Figure 5: Long-term Labor Supply Outcomes
(*Follow-up*)

(a) Applied for Job

(b) Interviewed for Job

**(c)** Employed

(d) Driving Lessons

Notes: Labor supply outcomes of participants' wives in the follow-up survey. 95% binomial proportion confidence intervals. *p*-value calculated from testing for equality of proportions. Panels (a) - (c) refer exclusively to job opportunities outside of the home.

Figure 6: Job-Matching Service Sign-up–Heterogeneity by Wedge
(*Main Experiment*)

Notes: Job-Matching service sign-up rates for respondents with non-positive and positive wedges in perceptions about the beliefs of others regarding whether women should be able to work outside the home. Wedges calculated as (*the respondent's guess about the % of session participants agreeing with the statement*) - (*the true % of session participants agreeing with the statement*). 95% binomial proportion confidence intervals. *p*-value calculated from testing for equality of proportions.

Figure 7: Misperceptions about Others' Beliefs
(*National Survey*)
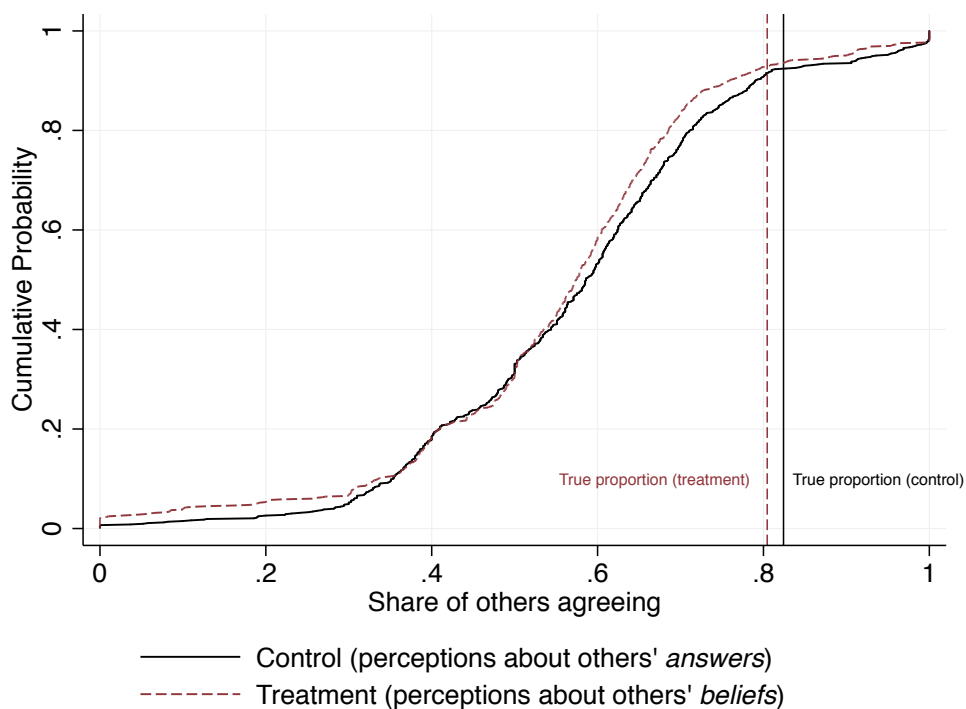
Notes: CDF of respondents' guesses about the share of other participants in the national survey agreeing with the statement that women should be able to work outside the home. Vertical lines show the true proportion of respondents agreeing with the statement. In the treatment group, private beliefs are elicited using a list experiment. The true proportion of respondents agreeing to the statement in the treatment group is then inferred by subtracting the average number of agreements in the control list from the treatment list (a list of statements identical to the control list with the single addition of the sensitive FLFP statement).

Figure 8: Comparison with Arab Barometer
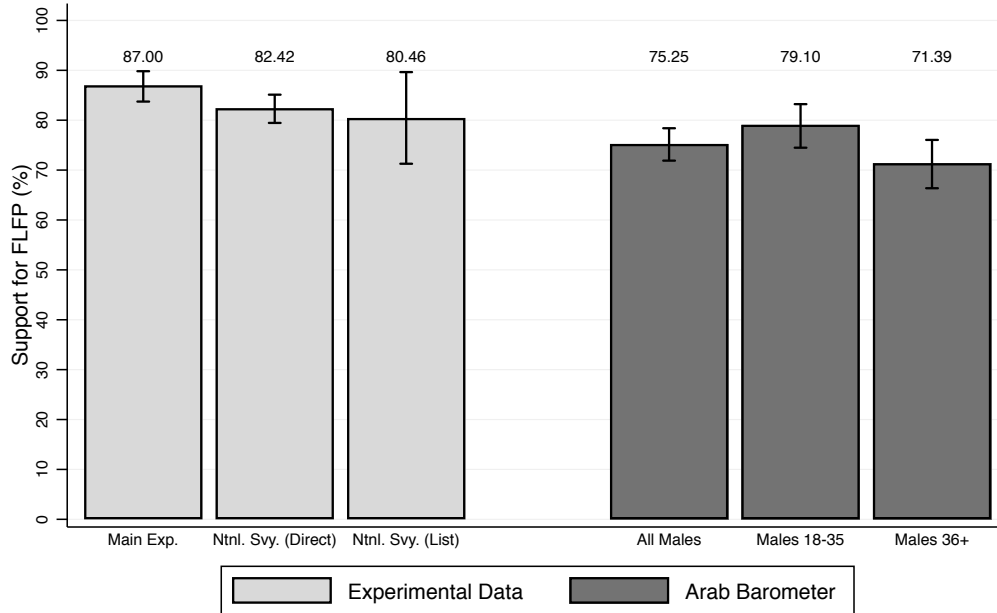
Notes: Comparison of support for female labor force participation in experimental data and the nationally representative Arab Barometer. In each of the experimental conditions, participants were presented with the following statement: "In my opinion, women should be allowed to work outside the home." The corresponding statement presented in the Arab Barometer is: "a married woman can work outside the home if she wishes." The plot shows the percentage of respondents agreeing to the presented statement in each case. In the Arab Barometer data, we define the agreement indicator by pooling respondents who reported "agreeing" or "strongly agreeing." "Ntnl. Svy. (Direct)" refers to the direct elicitation of beliefs conducted in our own nationally representative online survey. "Ntnl. Svy. (List)" refers to the level of support inferred from the list experiment in the same survey by subtracting the average number of agreements in the control group from the average number in the treatment group. 95% binomial proportion confidence intervals. Confidence interval for the level of support inferred from the list experiment calculated by taking the variance of the treatment–control estimator to be the sum of the variance of the sample mean in the control group and the variance of the sample mean in the treatment group.

Table 1: Job-Matching Service Sign-Up
(*Main Experiment*)

	(1)	(2)	(3)	(4)
Treatment (β)	0.0853**	0.0876**	0.0899**	0.0898**
	(0.0399)	(0.0399)	(0.0395)	(0.0404)
Constant	0.235***	0.132*	-0.156	-0.131
	(0.0270)	(0.0739)	(0.0970)	(0.186)
Inference Robustness (β)				
p-value: Robust S.E.	0.033	0.028	0.023	0.027
p-value: Wild Bootstrap	0.013	0.020	0.004	0.011
p-value: Permutation Test	0.036	0.038	0.024	0.025
Session F.E.		✓	✓	✓
Baseline beliefs			✓	✓
Controls				✓
N	500	500	500	491
R^2	0.00907	0.0696	0.108	0.116

Notes: Column (1) reports estimates from an OLS regression of an indicator for whether the respondent signed their wife up for the job-matching service on a treatment dummy. Column (2) includes session fixed effects. Column (3) controls for 1st and 2nd order baseline beliefs. Column (4) adds socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parenthesis. Reported p-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the session level. * p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01.

Table 2: Effect of Treatment on Labor Supply Outcomes
(*Follow-up*)

	K-L-K Index	Employed	Applied	Interviewed	Driving Lessons	Beliefs about Neighbors
Panel A: No controls						
Treatment (β)	0.249***	0.0206	0.104***	0.0471**	0.112**	0.128***
	(0.0495)	(0.0285)	(0.0317)	(0.0185)	(0.0463)	(0.0267)
Constant	-0.119***	0.0737***	0.0579***	0.0105	0.653***	0.439***
	(0.0272)	(0.0190)	(0.0170)	(0.00742)	(0.0346)	(0.0195)
Inference Robustness (β)						
p-value: Robust S.E.	0.000	0.471	0.001	0.011	0.016	0.000
p-value: Wild Bootstrap	0.000	0.523	0.027	0.008	0.019	0.000
p-value: Permutation Test	0.000	0.593	0.002	0.017	0.009	0.000
p-value: L-S-X MHT Corr.	–	0.484	0.003	0.050	0.039	0.000
Lee Attrition Bounds						
Lower Bound:	0.238***	0.014	0.101***	0.046**	0.092	0.119***
	(0.0576)	(0.0288)	(0.0314)	(0.0181)	(0.0564)	(0.0350)
Upper Bound:	0.276***	0.033	0.120**	0.056***	0.111**	0.138***
	(0.0592)	(0.0532)	(0.0555)	(0.0166)	(0.0500)	(0.0392)
N	381	381	381	381	381	381
R^2	0.0626	0.00137	0.0278	0.0168	0.0151	0.0571
Panel B: Session fixed effects, baseline beliefs and socioeconomic controls						
Treatment (β)	0.276***	0.0223	0.115***	0.0469**	0.125***	0.144***
	(0.0528)	(0.0282)	(0.0332)	(0.0182)	(0.0475)	(0.0268)
Constant	-0.141	0.0545	-0.0262	0.0222	0.882***	0.310***
	(0.210)	(0.137)	(0.146)	(0.0795)	(0.225)	(0.114)
Inference Robustness (β)						
p-value: Robust S.E.	0.000	0.429	0.001	0.010	0.009	0.000
p-value: Wild Bootstrap	0.000	0.452	0.016	0.002	0.023	0.001
p-value: Permutation Test	0.000	0.423	0.000	0.006	0.007	0.000
N	375	375	375	375	375	375
R^2	0.164	0.0947	0.166	0.117	0.128	0.229

Notes: Each column reports estimates from OLS regression of the outcome indicated by the column on a treatment dummy. The Kling-Liebman-Katz index, defined in the text, is constructed from all 6 tested outcomes including those not reported in the table. Panel A includes no controls while Panel B adds session fixed effects, baseline beliefs and socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. In Panel A, L-S-X MHT Correction refers to the multiple hypothesis testing procedure presented in List, Shaikh and Xu (2016). Robust standard errors reported in parenthesis. Reported p-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the (main-experiment) session level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 3: Job-Matching Service Sign-Up: Heterogeneity by Wedge
(*Main Experiment*)

	Wedge ≤ 0				Wedge > 0			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment (β)	0.122***	0.122***	0.127***	0.125***	-0.0274	-0.0383	-0.0391	-0.0252
	(0.0459)	(0.0457)	(0.0450)	(0.0463)	(0.0827)	(0.0883)	(0.0919)	(0.114)
Constant	0.214***	0.0511	-0.202*	-0.251	0.309***	0.315*	0.711	0.946
	(0.0297)	(0.0710)	(0.117)	(0.204)	(0.0628)	(0.162)	(0.659)	(0.882)
Inference Robustness (β)								
p-value: Robust S.E.	0.008	0.008	0.005	0.007	0.741	0.665	0.671	0.825
p-value: Wild Bootstrap	0.004	0.003	0.000	0.001	0.776	0.663	0.691	0.840
p-value: Permutation Test	0.014	0.009	0.007	0.010	0.850	0.644	0.651	0.826
Session F.E.		✓	✓	✓		✓	✓	✓
Baseline beliefs			✓	✓			✓	✓
Controls				✓				✓
N	374	374	374	365	126	126	126	126
R^2	0.0186	0.0943	0.135	0.150	0.000890	0.136	0.221	0.271

Notes: Wedge for the given statement (whether women should be able to work outside the home) calculated as (*the respondent's guess about the number of session participants agreeing with the statement*) - (*the true number of session participants agreeing with the statement*). Column (1) reports OLS estimates from regressing an indicator for whether the respondent signed their wife up for the job-matching service on a treatment dummy restricted to respondents with non-positive wedge (those underestimating the progressivism of the other participants). Column (5) reports estimates from the same specification restricted instead to respondents with positive wedge. Columns (2) and (6) include session fixed effects. Columns (3) and (7) additionally control for 1st and 2nd order baseline beliefs. Columns (4) and (8) add socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parenthesis. Reported p-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the session level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 4: Effect of Belief Update on Job-Matching Service Sign-Up
(*Main Experiment*)

	(1)	(2)	(3)
Update $(-\text{Wedge}*\mathbb{1}_{Treatment};\beta)$	0.00758**	0.00648*	0.00682*
	(0.00334)	(0.00342)	(0.00352)
Constant	-0.0255	-0.116	-0.111
	(0.0716)	(0.0979)	(0.182)
Inference Robustness (β)			
$\quad$ *p*-value: Robust S.E.	0.023	0.058	0.053
$\quad$ *p*-value: Wild Bootstrap	0.014	0.017	0.019
$\quad$ *p*-value: Permutation Test	0.028	0.069	0.069
Baseline beliefs and confidence	✓	✓	✓
Session F.E.		✓	✓
Controls			✓
N	500	500	491
R^2	0.0270	0.0798	0.0926

Notes: Column (1) reports estimates from an OLS regression on the update in belief about the support of others induced by the information treatment (for women working outside the home), controlling for own belief, belief about others and confidence in belief about others (for the question asking whether women should be able to work outside the home). The update measure is given by minus the wedge for those in the treatment condition and equals 0 for those in the control condition. Column (2) includes session fixed effects. Column (3) additionally includes socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parenthesis. Reported *p*-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the session level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Online Appendix — Not for Publication

A   A Simple Model of Labor Supply and Stigma

Assume a continuum of agents (husbands or guardians) making an observable labor-supply decision for their wives. Agents can hold one of two mutually exclusive convictions, A (pro-FLFP) or B (anti-FLFP); which we refer to as their type and write $t_i \in \{A, B\}$ for agent i. The share of citizens holding conviction A is p, so $\Pr(t_i = A) = p$. We do not assume that p is known, and instead allow agents to hold an incorrect belief about the share of citizens with conviction A, which we denote by q. To avoid dealing with higher-order beliefs, we assume that q is common knowledge.

Agents are paired with one another; within each pair, there is a sender i and receiver j. Sender i makes a binary labor supply decision $s_i \in \{A, B\}$, observed by a receiver j, where $s_i = A$ means allowing the sender's wife to work outside of home, and $s_i = B$ means not allowing it. Making a decision $s_i \neq t_i$ generates a cost of suboptimal decision-making $c_i \sim U[0, 1]$ and independently of t_i. The sender enjoys a benefit proportional (with intensity factor denoted by a) to his belief that the receiver approves his type. Finally, if a sender of either type chooses $s_i = A$, he receives an additional benefit $y > 0$ corresponding to the wife's income. Consequently, the expected utility of a citizen i with two-dimensional type (t_i, c_i) from choosing action s_i and being observed by receiver j is given by

$$U_i(s_i) = -c_i \mathbf{I}_{\{s_i \neq t_i\}} + y\mathbf{I}_{\{s_i = A\}} + aq \Pr_j(t_i = A \mid s_i) + a(1 - q)(1 - \Pr_j(t_i = A \mid s_i)), \tag{1}$$

where $\Pr_j(t_i = A \mid s_i)$ is the receiver's posterior that the sender is type A conditional on the decision he made, s_i.

In this game, we are interested in Perfect Bayesian Equilibria, which furthermore satisfy the D1 criterion (Cho and Kreps, 1987).

We start by characterizing equilibrium conditions:

Proposition 1. *Denote*

$$z_A(q) = \min \left\{ 1, \max \left\{ 0, \frac{\sqrt{(1 - q + qy)^2 - 4q(y - qy - a + 3aq - 2aq^2)} + (1 - q) + qy}{2q} \right\} \right\},$$

$$z_B(q) = \min \left\{ 1, \max \left\{ 0, \frac{\sqrt{(y - qy - q)^2 - 4aq(1 - 2q)(1 - q)} + y(1 - q) - q}{2(1 - q)} \right\} \right\}.$$

(i) *If $q > \frac{a-y}{2a}$, then all type A agents choose $s_i = A$. Type B agents with cost $c_i < z_B(q)$ choose $s_i = A$; type B agents with cost $c_i > z_B(q)$ choose $s_i = B$. Moreover, if q satisfies $y > aq(1 - 2q) + 1$, then all type B agents choose $s_i = A$.*

(ii) *If $q = \frac{a-y}{2a}$, then all agents choose $s_i = t_i$.*

(iii) *If $q < \frac{a-y}{2a}$, then all type B agents choose $s_i = B$. Type A agents with cost $c_i < z_A(q)$ choose $s_i = B$; type A agents with cost $c_i > z_A(q)$ choose $s_i = A$. Moreover, if q satisfies $y < a(1 - q)(1 - 2q) - 1$, then all type A agents choose $s_i = B$.*

Proof. Let $z_A \in [0, 1]$ denote the threshold such that type-A agents with $c_i < z_A$ choose $s_i = A$ and type-B agents with $c_i > z_A$ choose $s_i = B$.[33] Similarly, let $z_B \in [0, 1]$ denote the threshold such that type-B agents with $c_i < z_B$ choose $s_i = B$ and type-B agents with $c_i > z_B$ choose $s_i = A$. In other words, agents choose $s_i = t_i$ when the cost of suboptimal decision making is high relative to the payoff and choose $s_i \neq t_i$ when the cost is low relative to the payoff.

We begin our analysis with case (ii), the "truthful" equilibrium. In this case, $z_A = z_B = 0$, and thus a receiver places probability 1 on the event that the sender's action matches his type, i.e. $\Pr_j(t_i = A \mid s_i) = 1$ for $s_i = A$, $\Pr_j(t_i = B \mid s_i) = 1$ for $s_i = B$. Then for a type A agent, we have:

$$U_i(s_i = A | t_i = A) = y + aq \text{ and } U_i(s_i = B | t_i = A) = -c_i + a(1 - q),$$

---

[33] Agents with $c_i = z_A$ are indifferent; we will ignore this population of measure 0.

37

and for a type B agent we have:

$$U_i(s_i = A|t_i = B) = y - c_i + aq \text{ and } U_i(s_i = B|t_i = B) = a(1-q).$$

Note that for $z_A = 0$, we must have $U_i(s_i = A|t_i = A) - U_i(s_i = B|t_i = A) \geq 0$ for all type A agents, i.e. $y \geq a(1-2q) - c_i$. Since $\min\{c_i\} = 0$, this inequality must hold at $c_i = 0$, i.e. $y \geq a(1-2q)$. Similarly, for $z_B = 0$, we must have $U_i(s_i = A|t_i = A) - U_i(s_i = B|t_i = A) \leq 0$ for all type B agents, i.e. $y \leq a(1-2q) + c_i \implies y \leq a(1-2q)$. Thus at $y = a(1-2q) \implies q = \frac{a-y}{2a}$, all agents will choose $s_i = t_i$. This proves case (ii).

Next, we will prove case (i). All type A agents choose $s_i = A$ for any $q > \frac{a-y}{2a}$, so for a type B agent, choosing $s_i = A$ results in expected utility

$$U_i(s_i = A) = aq\frac{q}{q + (1-q)z_B} + a(1-q)\frac{(1-q)z_B}{q + (1-q)z_B} - c_i + y,$$

whereas choosing $s_i = B$ results in expected utility

$$U_i(m_i = B) = a(1-q),$$

because only receivers of type B will approve of him. The sender with cost $c_i = z_B$ is indifferent if and only if $U_i(s_i = A) = U_i(s_i = B)$, i.e.

$$aq\frac{q}{q + (1-q)z_B} + a(1-q)\frac{(1-q)z_B}{q + (1-q)z_B} - c_i + y - a(1-q) = 0 \tag{2}$$

The left hand side must be positive at $z_B = 1$, i.e. $y > aq(1-2q) + 1$. As discussed before, the left hand side must be negative at $z_B = 0$, i.e. $y < a(1-2q)$. Thus for $z_B \in (0,1)$, we need $aq(1-2q) + 1 < y < a(1-2q)$. Suppose that y satisfies this inequality; then solving (2) for z_B yields

$$z_B = \frac{\sqrt{(y - qy - q)^2 - 4aq(1-2q)(1-q)} + y(1-q) - q}{2(1-q)} \tag{3}$$

Note that in the special case $q = \frac{1}{2}$, this becomes $z_B = y$. The intuition is as follows. At $q = \frac{1}{2}$, social image concerns become irrelevant, since the expected social utility of undertaking either action is the same. It is easy to see that type A agents will report truthfully when $q = \frac{1}{2}$, since choosing $s_i = B$ would result not only in the cost c_i but also in the loss of income y. Type B agents, on the other hand, must balance the cost c_i against the income y. Agents for whom $c_i < y$ will report $s_i = A$; agents for whom $c_i > y$ will report $s_i = B$.

Finally, we will prove case (iii); the analysis is similar to that of case (i). All type B agents choose $s_i = B$ for any $q < \frac{a-y}{2a}$, so for a type A agent, choosing $s_i = A$ results in expected utility

$$U_i(s_i = A) = y + aq,$$

while choosing $s_i = B$ results in expected utility

$$U_i(s_i = B) = -c_i + aq\frac{qz_A}{qz_A + (1-q)} + a(1-q)\frac{(1-q)}{qz_A + (1-q)}.$$

The sender with cost $c_i = z_A$ is indifferent if and only if $U_i(s_i = A) = U_i(s_i = B)$, i.e.

$$y + aq + z_A - aq\frac{qz_A}{qz_A + (1-q)} - a(1-q)\frac{(1-q)}{qz_A + (1-q)} = 0 \tag{4}$$

The left hand side must be negative at $z_A = 1$, i.e. $y < a(1-q)(1-2q) - 1$. As discussed before, the left hand side must be positive at $z_A = 0$, i.e. $y > a(1-2q)$. Thus for $z_A \in (0,1)$, we need $a(1-2q) < y < a(1-q)(1-2q) - 1$. Suppose that y satisfies this inequality; then solving (4) for z_A yields

$$z_A = \frac{\sqrt{(1 - q + qy)^2 - 4q(1-q)(2aq - a + y)} + (1-q) + qy}{2q} \tag{5}$$

∎

We can now provide a result describing the effect of an experimental intervention that shifts beliefs from $q < \frac{a-y}{2a}$ to some $q' > q$. Note that in our experimental sample, the average measured prior (q) is greater than $1/2$ (which

would imply that all agents in favor of FLFP should choose to have their wives work). Still, we are in a situation where these agents are not choosing to have their wives work, which is consistent with case (iii) in Proposition 1. One way to reconcile these elements is to assume that the measured prior is a proxy for the relevant q, which is lower.

Clearly, if the shift in beliefs from q to q' is sufficiently large, i.e. $q' \geq \frac{a-y}{2a}$, then the new q' will satisfy case (ii) or case (iii) in Proposition 1, and thus the number of A choices will increase. We characterize the effect of a smaller change in q below.

Proposition 2. *For* $q < \frac{a-y}{2a}$, $\frac{dz_A(q)}{dq} \leq 0$. *If* $z_A \neq 1$, *i.e.* $y \geq a(1-q)(1-2q) - 1$, *the inequality is strict.*

Proof. Suppose that $q < \frac{a-y}{2a}$ (case (iii) of Proposition 1). Then all type B agents choose $s_i = B$; type A agents with a cost $c_i < z_A(q)$ choose $s_i = B$, while type B agents with a cost $c_i > z_A(q)$ choose $s_i = A$.

First, suppose that $y < a(1-q)(1-2q) - 1$. Then all type A agents choose $s_i = B$. For any $q' \in (q, q+\epsilon)$, we will also have $y < a(1-q')(1-2q') - 1$, so all type A agents once again choose $s_i = B$. Thus $\frac{dz_A(q)}{dq} = 0$.

Now suppose that $y > a(1-q)(1-2q) - 1$, so that $z_A \in (0,1)$. For a type A agent with cost $c_i = z_A$, we have:

$$U_i(s_i = B) - U_i(s_i = A) = aq\frac{qz_A}{qz_A + (1-q)} + a(1-q)\frac{(1-q)}{qz_A + (1-q)} - z_A - aq - y$$

$$= \frac{a - y - z + 2aq^2 - qz^2 - 3aq + qy + qz - qyz}{-q + qz + 1}$$

Denote the numerator by $f(z_A, q) = (-q)z_A^2 + (q - qy - 1)z_A + (a - y + 2aq^2 - 3aq + qy)$. The agent with cost $c_i = z_A$ must be indifferent between choosing A and B, so the equilibrium z^* solves $f(z^*, q) = 0$. The coefficient of the z_A^2 term is negative, and $a - y + 2aq^2 - 3aq + qy = (1-q)(a - y - 2aq) > 0$ for $q < \frac{a-y}{2a}$. So there is exactly one positive root z_A^*, and the parabola points downward. Thus $\frac{\partial f}{\partial z_A}|_{z_A = z^*} < 0$.

To show that $\frac{dz_A}{dq}|_{z_A = z^*} < 0$, it suffices to prove $\frac{\partial f}{\partial q}|_{z = z^*} < 0$. We have that $\frac{\partial f}{\partial q}\Big|_{z = z^*} = -(z^*)^2 + (1-y)z^* + (y - 3a + 4aq)$. We solve our expression $f(z^*, q) = 0$ to compute $z^{*2} = q^{-1}[(q - qy - 1)z^* + (a - y + 2aq^2 - 3aq + qy)]$. Plugging this expression for z^* into $\frac{\partial}{\partial q}(\cdot)|_{z = z^*}$, we derive

$$\frac{\partial}{\partial q}(\cdot)\Big|_{z = z^*} = \frac{z^* - a + y + 2aq^2}{q}$$

Denoting the numerator by $g(z)$, it suffices to prove that $g(z^*) < 0$. Since $\frac{dg(z)}{dz} > 0$ and g admits an inverse g^{-1}, it suffices to prove that $g^{-1}(0) > z^*$. $g^{-1}(0) = a - y - 2aq^2$; we claim that $g^{-1}(0)$ is positive, since $g^{-1}(0) = a - y - 2aq^2 > a - y - 2aq > a - y - 2a\frac{a-y}{2a} = 0$. Since $f(z, q) > 0$ for $z \in (0, z^*)$ and $f(z, q) < 0$ for $z > z^*$, it suffices to prove that $f(g^{-1}(0), q) < 0$.

$$f(g^{-1}(0), q) = -aq(a - 4q - y - 4aq^2 + 4aq^4 + 2q^2y + 2q^2 + 2)$$

Denoting the second factor by $h(y)$, i.e. $h(y) = (a - 4q - y - 4aq^2 + 4aq^4 + 2q^2y + 2q^2 + 2)$, it suffices to show that $h(y) > 0$. Simplifying, we get

$$h(y) = (2q^2 - 1)y + (a - 4q - 4aq^2 + 4aq^4 + 2q^2 + 2)$$

Notice that this is linear in y, so to show that this is positive for all possible values of y, it suffices to check the minimal value ($y_1 = 0$) and the maximal possible value found from the condition $q < \frac{a-y}{2a}$, i.e. $y_2 = a(1 - 2q)$. Recalling that $q < \frac{a-y}{2a}$ guarantees that $q < \frac{1}{2}$, we have:

$$h(y_1) = (1 - 2q^2)^2 a + 2(1-q)^2 > 0, \text{ and } h(y_2) = 2(1-q)(q(1 - 2q^2)a + (1-q)) > 0$$

Thus we conclude that $h(y) > 0$ for all y, and therefore that $\frac{dz_A}{dq}|_{z_A = z^*} < 0$. In other words, assuming we start at an interior solution for z_A, a positive update in q results in more A actions being undertaken in equilibrium.

∎

Appendix Figure and Tables

Figure A1: Misperceptions about Others' Beliefs (Semi-segregated Environments)
(*Main Experiment*)
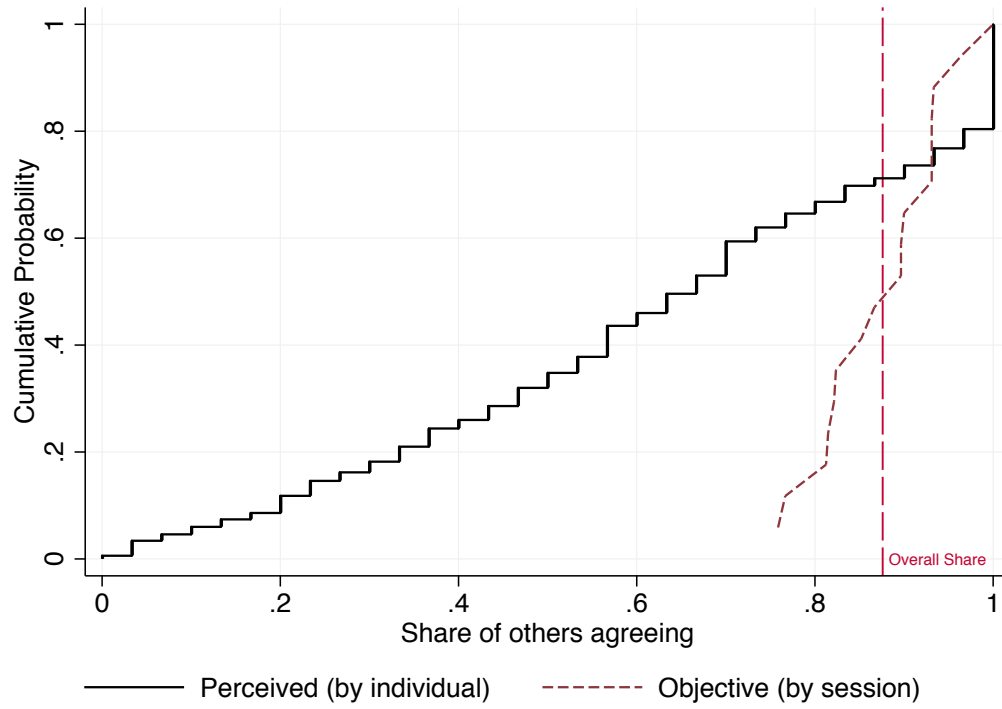
Figure A2: Wedges in Perceptions of Others' Beliefs (Semi-segregated Environments)
(*Main Experiment*)
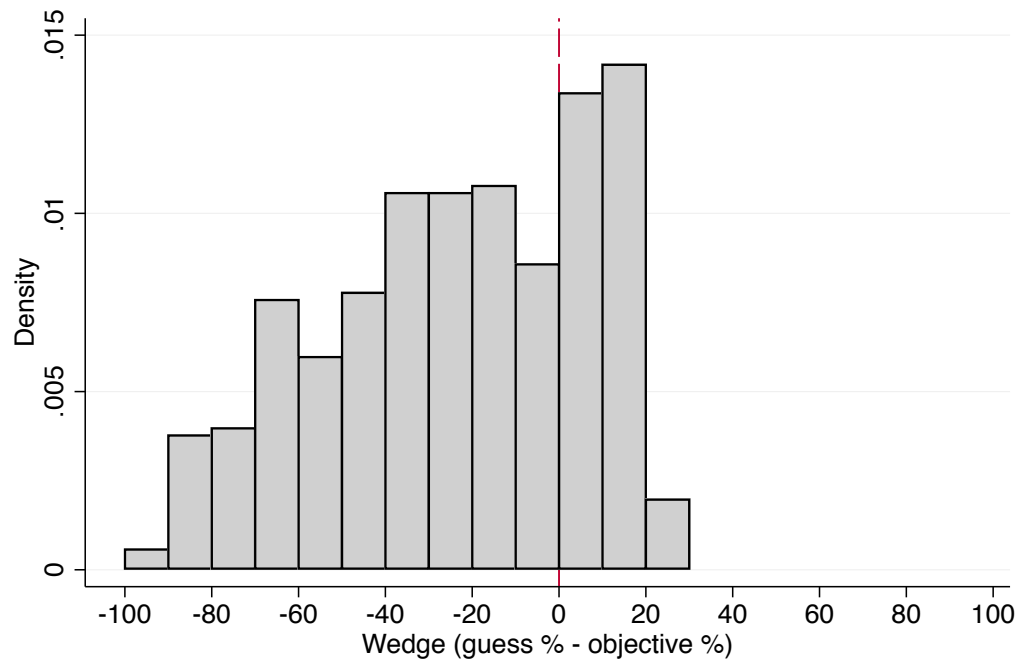
Notes: The distribution of wedges in perceptions about the beliefs of others regarding whether women should be able to work in semi-segregated environments. Wedges calculated as (*the respondent's guess about the % of session participants agreeing with the statement*) - (*the true % of session participants agreeing with the statement*).

Figure A3: Confidence and Social Connections
(*Main Experiment*)

(a) Confidence and Accuracy

(b) Connections and Confidence

(c) Connections and Accuracy
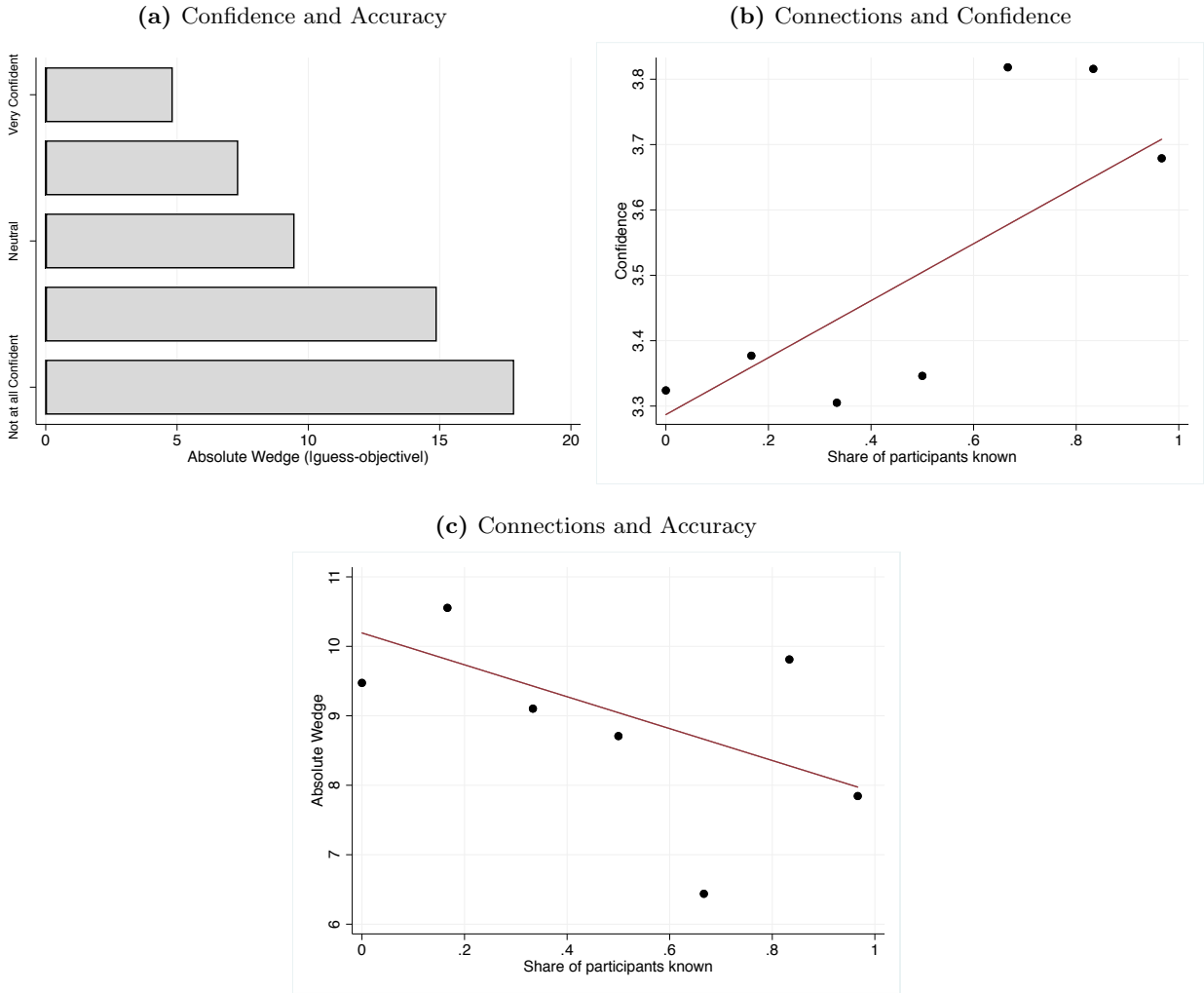
Notes: Panel (a) shows the average absolute wedge in perceptions about the beliefs of others for each confidence level. Panel (b) and (c) show binned scatterplots of confidence and absolute wedge, respectively, against the share of other participants the respondent claimed to know. Absolute wedge calculated as |*the respondent's guess about the number of session participants agreeing with the statement*) - (*the true number of session participants agreeing with the statement*)|. Confidence measured on a scale of 1-5.

42

Figure A4: Persistence in Beliefs Update
(*Follow-up*)
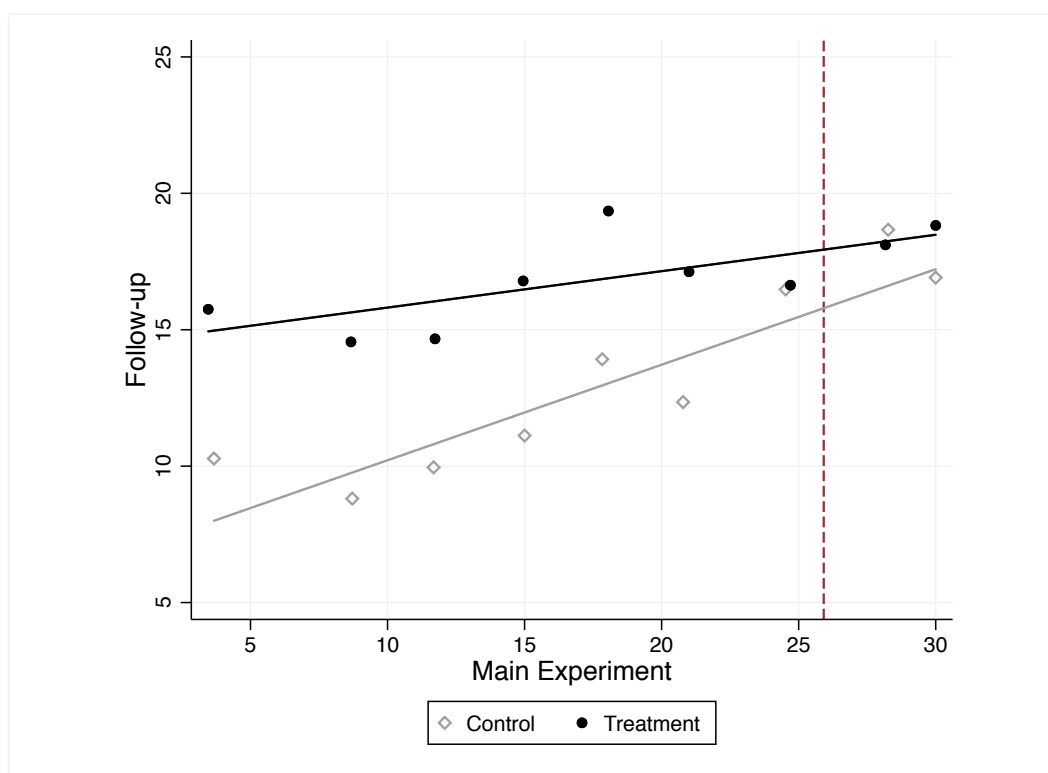
Notes: Guesses of the number of randomly selected neighbors (out of 30) who would support FLFP plotted against participants' guesses of the number of participants (out of 30) supporting FLFP in their main experiment session. The red dotted line shows the overall share of private support for FLFP in the main experiment.

Figure A5: Confidence and Accuracy
(*National Survey*)

(a) Control Group **(b)** Treatment Group
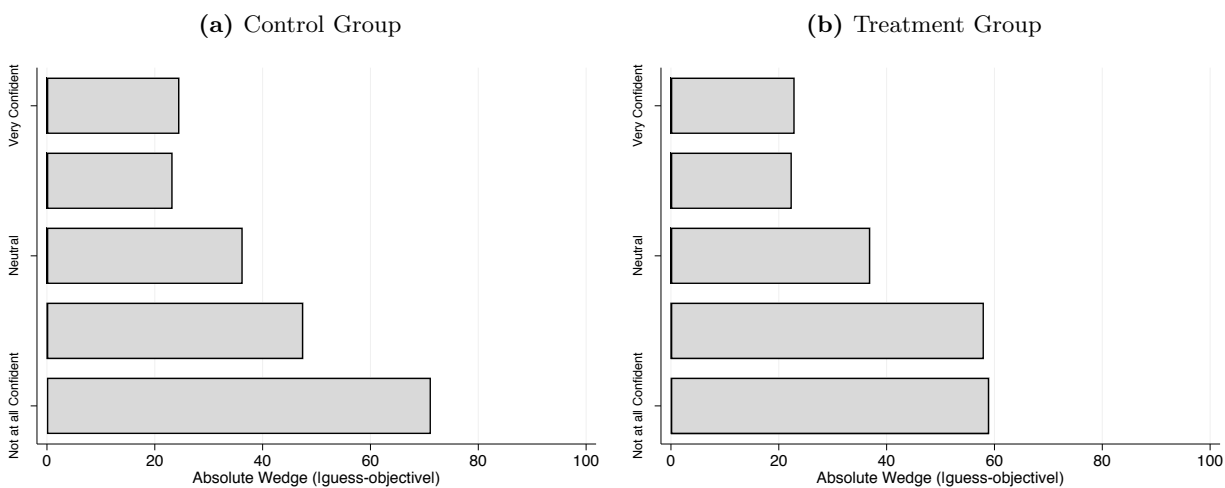
Notes: Respondents in the control group were asked to guess the share of other participants *reporting* agreeing with the statement that women should have the right to work outside the home. Respondents in the treatment group were asked to guess the share of other participants *privately* agreeing with the statement. Absolute wedge calculated as |*the respondent's guess about the share of survey participants responding affirmatively (agreeing privately) to the statement*) - (*the true (or inferred) share of survey participants agreeing with the statement*)|. Confidence measured on a scale of 1-5.

44

Table A1: Attrition
(*Follow-up*)

	All	Successful follow-up	No follow-up
N	500	381	119
Treatment (%)	50.60	50.13	52.10
Age	24.78	24.81	24.68
	(4.21)	(4.29)	(3.94)
Number of Children	1.71	1.69	1.77
	(1.72)	(1.70)	(1.82)
College Degree (%)	56.20	55.91	57.14
Employed (%)	86.60	85.83	89.08
Wife Employed (%)	65.20	63.52	70.59
Job-Matching Service Sign-up (%)	27.80	29.40	22.69
Other Participants Known (%)	51.19	49.72	55.88
	(38.24)	(38.71)	(36.46)
Other Participants with Mutual Friends (%)	38.64	38.03	40.59
	(34.94)	(35.03)	(34.71)

Notes: Summary statistics of respondent characteristics for the subsample that was successfully reached for a follow-up and for the subsample that was not successfully reached. Asterisks in the no follow-up column refer to p-values from two-tailed t-tests of equality with the successful follow-up group; asterisks in the full sample column test the successful follow-up group against the full sample mean. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Attrition by Treatment
(*Follow-up*)

	Control			Treatment		
	All	Follow-up	Dropped	All	Follow-up	Dropped
N	247	190	57	253	191	62
Age	24.64	24.68	24.54	24.91	24.94	24.82
	(3.99)	(4.16)	(3.41)	(4.41)	(4.43)	(4.41)
Number of Children	1.64	1.65	1.60	1.77	1.72	1.94
	(1.70)	(1.83)	(1.22)	(1.74)	(1.56)	(2.22)
College Degree (%)	55.06	54.21	57.89	57.31	57.59	56.45
Employed (%)	87.45	86.84	89.47	85.77	84.82	88.71
Wife Employed (%)	65.59	61.58	78.95*	64.82	65.45	62.90
Job-Matching Service Sign-up (%)	23.48	24.74**	19.30	32.02	34.03	25.81
Other Participants Known (%)	49.68	49.05	51.75	52.66	50.38	59.68
	(38.60)	(39.16)	(36.93)	(37.92)	(38.36)	(35.90)
Other Participants with Mutual Friends (%)	37.62	37.32	38.65	39.63	38.74	42.37
	(34.62)	(34.94)	(33.79)	(35.29)	(35.20)	(35.73)

Notes: Summary statistics of respondent characteristics for the subsample that was successfully reached for a follow-up and for the subsample that was not successfully reached, split by treatment condition in the original experiment. Asterisks in the dropped respondent columns refer to p-values from two-tailed t-tests of equality with the corresponding treatment condition among respondents with successful follow-ups; asterisks in the successful follow-up columns test the sample indicated by the column against the full sample mean. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3: Sample Summary Statistics
(*Main Experiment*)

	All	Control	Treatment
N	500	247	253
Age	24.78	24.64	24.91
	(4.21)	(3.99)	(4.41)
Number of Children	1.71	1.64	1.77
	(1.72)	(1.70)	(1.74)
College Degree (%)	56.20	55.06	57.31
Employed (%)	86.60	87.45	85.77
Wife Employed (%)	65.20	65.59	64.82
Other Participants Known (%)	51.19	49.68	52.66
	(38.24)	(38.60)	(37.92)
Other Participants with Mutual Friends (%)	38.64	37.62	39.63
	(34.94)	(34.62)	(35.29)

Notes: Summary statistics of respondent characteristics in the field experiment. Asterisks in the treatment column refer to p-values from two-tailed t-tests of equality with the control group. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Effect of Treatment on Labor Supply Outcomes: Heterogeneity by Wedge (No Controls)
(*Follow-up*)

	K-L-K Index	Employed	Applied	Interviewed	Driving Lessons	Beliefs about Neighbors
Panel A: Wedge ≤ 0						
Treatment (β)	0.261***	0.00357	0.125***	0.0438**	0.108**	0.142***
	(0.0609)	(0.0331)	(0.0388)	(0.0221)	(0.0524)	(0.0314)
Constant	-0.124***	0.0828***	0.0621***	0.0138	0.676***	0.401***
	(0.0334)	(0.0230)	(0.0201)	(0.00972)	(0.0390)	(0.0223)
Inference Robustness (β)						
p-value: Robust S.E.	0.000	0.914	0.001	0.048	0.040	0.000
p-value: Wild Bootstrap	0.000	0.929	0.037	0.036	0.063	0.005
p-value: Permutation Test	0.000	1.000	0.002	0.039	0.044	0.000
p-value: L-S-X MHT Correction	–	0.904	0.006	0.111	0.117	0.000
Lee Attrition Bounds						
Lower Bound:	0.252***	-0.005	0.122***	0.043**	0.089	0.136***
	(0.0704)	(0.0339)	(0.0386)	(0.0218)	(0.0656)	(0.0395)
Upper Bound:	0.290***	0.024	0.151**	0.057***	0.118**	0.152***
	(0.0715)	(0.0623)	(0.0664)	(0.0196)	(0.0567)	(0.0476)
N	284	284	284	284	284	284
R^2	0.0623	0.0000412	0.0361	0.0141	0.0148	0.0680
Panel B: Wedge > 0						
Treatment (β)	0.214***	0.0709	0.0517	0.0577*	0.134	0.0673
	(0.0791)	(0.0545)	(0.0517)	(0.0327)	(0.0978)	(0.0458)
Constant	-0.104**	0.0444	0.0444	3.12e-17	0.578***	0.562***
	(0.0405)	(0.0310)	(0.0310)	(1.57e-09)	(0.0744)	(0.0352)
Inference Robustness (β)						
p-value: Robust S.E.	0.008	0.196	0.320	0.081	0.175	0.145
p-value: Wild Bootstrap	0.017	0.253	0.455	0.196	0.197	0.202
p-value: Permutation Test	0.008	0.288	0.460	0.258	0.232	0.136
p-value: L-S-X MHT Correction	–	0.346	0.323	0.486	0.436	0.498
Lee Attrition Bounds						
Lower Bound:	0.199**	0.061	0.043	0.050	0.106	0.043
	(0.0889)	(0.0980)	(0.0980)	(0.0930)	(0.1022)	(0.0575)
Upper Bound:	0.235**	0.067	0.049	0.056*	0.111	0.081
	(0.0930)	(0.0541)	(0.0511)	(0.0318)	(0.1167)	(0.0635)
N	97	97	97	97	97	97
R^2	0.0665	0.0165	0.00993	0.0276	0.0195	0.0225

Notes: Each column reports estimates from OLS regression of the outcome indicated by the column on a treatment dummy. For estimates of the full specification with fixed effects, socioeconomic controls and baseline beliefs included as covariates (but without L-S-X multiple hypothesis testing correction and Lee attrition bounds), see Table A4. Wedge for the given statement (whether women should be able to work outside the home) calculated as (*the respondent's guess about the number of session participants agreeing with the statement*) - (*the true number of session participants agreeing with the statement*). Panel A restricts to the subsample of respondents with a non-positive wedge in perceptions about others (those underestimating support for FLFP among the other participants) while Panel B restricts to the subsample of respondents with a positive wedge in perceptions about others. The Kling-Liebman-Katz index, defined in the text, is constructed from all 6 tested outcomes including those not reported in the table. L-S-X MHT Correction refers to the multiple hypothesis testing procedure presented in List, Shaikh and Xu (2016). Robust standard errors reported in parenthesis. Reported p-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the (main-experiment) session level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Effect of Treatment on Labor Supply Outcomes: Heterogeneity by Wedge (Full Specification)
(*Follow-up*)

	K-L-K Index	Employed	Applied	Interviewed	Driving Lessons	Beliefs about Neighbors
Panel A: Wedge ≤ 0 (Session fixed effects, baseline beliefs and socioeconomic controls)						
Treatment (β)	0.317***	0.0138	0.151***	0.0476**	0.145***	0.162***
	(0.0641)	(0.0318)	(0.0407)	(0.0217)	(0.0531)	(0.0317)
Constant	-0.460*	-0.172	-0.0528	-0.0392	0.648**	0.241*
	(0.243)	(0.152)	(0.178)	(0.0987)	(0.265)	(0.145)
Inference Robustness (β)						
p-value: Robust S.E.	0.000	0.664	0.000	0.030	0.007	0.000
p-value: Wild Bootstrap	0.000	0.649	0.010	0.030	0.030	0.005
p-value: Permutation Test	0.000	0.639	0.000	0.025	0.009	0.000
N	278	278	278	278	278	278
R^2	0.202	0.131	0.202	0.123	0.156	0.264
Panel B: Wedge > 0 (Session fixed effects, baseline beliefs and socioeconomic controls)						
Treatment (β)	0.150*	0.0716	0.0108	0.00478	0.0703	0.115*
	(0.0756)	(0.0671)	(0.0570)	(0.0270)	(0.112)	(0.0638)
Constant	2.675***	1.592**	0.251	0.397	3.858***	0.865
	(0.741)	(0.714)	(0.549)	(0.370)	(0.929)	(0.610)
Inference Robustness (β)						
p-value: Robust S.E.	0.051	0.290	0.850	0.860	0.531	0.076
p-value: Wild Bootstrap	0.033	0.218	0.836	0.678	0.649	0.068
p-value: Permutation Test	0.039	0.271	0.861	0.862	0.517	0.078
N	97	97	97	97	97	97
R^2	0.582	0.501	0.530	0.596	0.491	0.333

Notes: Each column reports estimates from OLS regression of the outcome indicated by the column on a treatment dummy, session fixed effects, baseline beliefs, socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. For estimates of the basic specification without fixed effects or additional covariates but including L-S-X multiple hypothesis testing correction and Lee attrition bounds, see Table A5. Wedge for the given statement (whether women should be able to work outside the home) calculated as (*the respondent's guess about the number of session participants agreeing with the statement*) - (*the true number of session participants agreeing with the statement*). Panel A restricts to the subsample of respondents with a non-positive wedge in perceptions about others (those underestimating support for FLFP among the other participants) while Panel B restricts to the subsample of respondents with a positive wedge in perceptions about others. The Kling-Liebman-Katz index, defined in the text, is constructed from all 6 tested outcomes including those not reported in the table. Robust standard errors reported in parenthesis. Reported p-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the (main-experiment) session level. * p < 0.1, ** p < 0.05, *** p < 0.01.

	K-L-K Index	Employed	Applied	Interviewed	Driving Lessons	Beliefs about Neighbors
Panel A: Baseline beliefs and confidence						
Update $(-\text{Wedge}*\mathbb{1}_{Treatment};\beta)$	0.0191***	-0.00227	0.00917***	0.00235	0.00812**	0.0103***
	(0.00590)	(0.00241)	(0.00288)	(0.00152)	(0.00387)	(0.00209)
Constant	-0.204**	0.103*	0.0905	0.0475	0.590***	0.222***
	(0.0967)	(0.0565)	(0.0692)	(0.0358)	(0.106)	(0.0561)
Inference Robustness (β)						
p-value: Robust S.E.	0.001	0.347	0.002	0.123	0.036	0.000
p-value: Wild Bootstrap	0.000	0.484	0.053	0.113	0.001	0.000
p-value: Permutation Test	0.003	0.377	0.002	0.139	0.047	0.000
N	381	381	381	381	381	381
R^2	0.0531	0.00428	0.0398	0.0286	0.0140	0.127
Panel B: Baseline beliefs, confidence, session fixed effects and demographic controls						
Update $(-\text{Wedge}*\mathbb{1}_{Treatment};\beta)$	0.0207***	-0.00242	0.00943***	0.00248*	0.00842**	0.0118***
	(0.00598)	(0.00262)	(0.00304)	(0.00146)	(0.00401)	(0.00221)
Constant	-0.272	0.153	-0.0982	0.0444	0.617***	0.213*
	(0.193)	(0.129)	(0.143)	(0.0726)	(0.219)	(0.118)
Inference Robustness (β)						
p-value: Robust S.E.	0.001	0.357	0.002	0.089	0.036	0.000
p-value: Wild Bootstrap	0.002	0.492	0.052	0.092	0.014	0.000
p-value: Permutation Test	0.000	0.364	0.006	0.125	0.040	0.000
N	375	375	375	375	375	375
R^2	0.143	0.0905	0.157	0.0902	0.112	0.213

Notes: Each column reports estimates from OLS regression of the outcome indicated by the column on the update in belief about the support of others for women working outside the home induced by the information treatment. The update measure is given by minus the wedge for those in the treatment condition and takes value 0 for those in the control condition. The Kling-Liebman-Katz index, defined in the text, is constructed from all 6 tested outcomes including those not reported in the table. Panel A controls for the respondent's own belief about whether women should be able to work outside the home, how many others they thought supported women working outside the home as well as their confidence in this guess. Panel B adds session fixed effects and socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parenthesis. Reported p-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the (main-experiment) session level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A7: Perceptions of Labor Demand
(Main Experiment)

	(1)	(2)	(3)	(4)
Treatment (β)	2.264	2.361	2.151	2.963
	(2.527)	(2.501)	(2.219)	(2.229)
Constant	54.22***	56.85***	17.87***	22.97**
	(1.759)	(4.515)	(5.911)	(10.59)
Inference Robustness (β)				
p-value: Robust S.E.	0.371	0.346	0.333	0.185
p-value: Wild Bootstrap	0.217	0.191	0.279	0.145
p-value: Permutation Test	0.389	0.366	0.313	0.190
Session F.E.		✓	✓	✓
Baseline beliefs			✓	✓
Controls				✓
N	486	486	486	477
R^2	0.00165	0.0795	0.317	0.345

Notes: Column (1) reports estimates from an OLS regression of estimated labor demand on a treatment dummy. Column (2) includes session fixed effects. Column (3) controls for 1st and 2nd order baseline beliefs. Column (4) adds socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parenthesis. Reported p-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the session level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Perceptions of Labor Demand and Job-Matching Service Sign-up (*Main Experiment*)

	(1)	(2)	(3)	(4)
Expected Labor Demand (β)	0.0506*	0.0443	0.0192	0.0248
	(0.0286)	(0.0301)	(0.0316)	(0.0331)
Constant	0.248***	0.241**	-0.0847	-0.0375
	(0.0282)	(0.111)	(0.151)	(0.294)
Inference Robustness (β)				
p-value: Robust S.E.	0.079	0.142	0.544	0.455
p-value: Wild Bootstrap	0.077	0.130	0.402	0.243
p-value: Permutation Test	0.077	0.160	0.545	0.435
Session F.E.		✓	✓	✓
Baseline beliefs			✓	✓
Controls				✓
N	236	236	236	233
R^2	0.0129	0.0773	0.108	0.147

Notes: Control group only. Column (1) reports estimates from an OLS regression of an indicator for whether the respondent signed their wife up for the job-matching service on estimated labor demand (standardized to mean 0 and s.d. 1) and a treatment dummy. Column (2) includes session fixed effects. Column (3) controls for 1st and 2nd order baseline beliefs. Column (4) adds socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parenthesis. Reported p-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the session level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A9: Persistence of Beliefs Update
(*Follow-up*)

	(1)	(2)	(3)	(4)
Treatment (β)	4.376***	4.300***	8.011***	8.859***
	(0.791)	(0.784)	(1.706)	(1.727)
Treatment*Baseline Belief about Others (γ)			-0.209**	-0.257***
			(0.0874)	(0.0894)
Constant	15.12***	7.340***	5.126**	6.618*
	(1.786)	(2.269)	(2.452)	(3.562)
Inference Robustness (β)				
p-value: Robust S.E.	0.000	0.000	0.000	0.000
p-value: Wild Bootstrap	0.001	0.004	0.000	0.000
p-value: Permutation Test	0.000	0.000	0.000	0.000
Inference Robustness (γ)				
p-value: Robust S.E.			0.017	0.004
p-value: Wild Bootstrap			0.051	0.015
p-value: Permutation Test			0.020	0.004
Session F.E	✓	✓	✓	✓
Baseline beliefs		✓	✓	✓
Controls				✓
N	381	381	381	375
R^2	0.123	0.200	0.212	0.247

Notes: Column (1) reports estimates from an OLS regression of respondents estimate of the number of people out of 30 randomly selected people in their neighborhood who would support FLFP (measured during the follow-up) on a treatment dummy (as determined in the main experiment) and session fixed effects. Column (2) controls for 1st and 2nd order baseline beliefs. Column (3) additionally includes the interaction between a dummy for treatment status and the respondents original belief about the number of people in their main experiment session supporting FLFP. Column (4) adds socioeconomic controls for age, education, employment status (of both respondent and wife), number of children and the share of people in the session room the respondent reported knowing and having mutual friends with. Robust standard errors reported in parenthesis. Reported p-values for wild bootstrap and permutation tests derived from running 1000 replications in each case. Wild bootstrap clustered at the (main-experiment) session level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A10: Sample Summary Statistics
(*National Survey*)

	All	Control	Treatment
N	1460	728	732
Age	28.42	28.47	28.36
	(4.31)	(4.31)	(4.31)
Number of Children	2.94	2.95	2.93
	(1.20)	(1.18)	(1.22)
College Degree (%)	59.52	58.65	60.38
Employed (%)	79.66	78.98	80.33
Wife Employed (%)	42.60	43.41	41.80
Wife Employed Outside of Home (%)	10.13	9.18	11.11

Notes: Summary statistics of respondent characteristics in the online survey. Asterisks in the treatment column refer to p-values from two-tailed t-tests of equality with the control group. * p < 0.1, ** p < 0.05, *** p < 0.01.

Survey Scripts

Main Experiment

SURVEY 1 BEGIN

[Consent Form (all)]

If you have any additional questions please let the facilitator know now. After beginning the survey, please keep additional questions until the end of the session to minimize disruptions. I have read this form and the research study has been explained to me. I have been given the opportunity to ask questions and my questions have been answered. If I have additional questions, I have been told whom to contact. I agree to participate in the research study described above and have received a copy of this consent form.

    I consent
    I do not consent <-{cannot continue}

Q1 Please enter the last three digits of your phone number: (This information will not be linked to any personally identifiable information)

Q2 Please fill in the following fields.
    What is your age?

Q3 What is your marital status?
    Single (1)
    Married (2)
    Widowed (3)
    Divorced (4)

Q4 What is the highest education level you have achieved?
    None (1)
    Primary (2)
    Secondary (3)
    Middle (4)
    Diploma (5)
    Bachelor (6)
    Masters (7)
    Ph. D (8)
    Other Post-Graduate (9)

Q5 Have you ever been employed?
    Yes (1)
    No (2)

Q6 Are you currently employed?
    Yes (1)
    No (2)

Q7 Is your wife currently employed?
    Yes (1)
    No (2)

Q8 What is your zipcode of residence?

_____
don't know

Q9 How many children do you have?
    (dropdown, 0-8+)

Q10 How many of the other participants in the room did you know before today's survey?
    0
    1-5
    6-10
    11-15
    16-20
    21-25
    26-29

Q11 Based on your initial impressions, with how many of the other participants in the room would you guess you have a mutual friend?
    0
    1-5
    6-10
    11-15
    16-20
    21-25
    26-29

Q12 We would like to ask you some questions related to employment and the Saudization of the labor market. If you are not sure about your answer, please answer to the best of your ability. All your answers are completely anonymous.

Q13 Do you agree with the following statement?
In my opinion, the current unemployment insurance system (Haafez) is good for the economy.
    Yes (1)
    No (2)

Q14 Do you agree with the following statement?
In my opinion, Saudi nationals should receive privileged access to job vacancies before expatriate workers.
    Yes (1)
    No (2)

Q15 On each of the following pages, you will be presented with a statement. First, you will be asked whether you personally agree with the statement. You will then be asked to guess how many of the other participants in the room agree with the statement. The participant who guesses most accurately across all the statements will receive a $20 Amazon gift card.

Q16 Do you agree with the following statement?   {Looped statement}
    Yes (1)
    No (2)

Q17 If you had to guess, how many people among the 29 other study participants in this room do you think agree with the same statement?   {Looped statement}
    0 (1)

1-2 (2)
3-4 (3)
5-6 (4)
7-8 (5)
9-10 (6)
11-12 (7)
13-14 (8)
15-16 (9)
17-18 (10)
19-20 (11)
21-22 (12)
23-24 (13)
25-26 (14)
27-28 (15)
29 (16)

Q18 How confident are you about your guess regarding the opinions of the other participants?
1 (1) <-Not at all confident
2 (2)
3 (3)
4 (4)
5 (5) <-Very confident

{Looped statements:

1)   In my opinion, the minimum wage for Saudis (SAR 3000) should be kept at its current level.
2)   In my opinion, women should be allowed to work outside of the home.
3)   In my opinion, a woman should have the right to work in semi-segregated environments.}

Q19 We will determine the overall prize winner after tabulating the responses. If you would like to receive a $20 Amazon gift card reward in case you are selected as the winner, please enter your email below so that we may email you the reward code after the survey concludes (the winner will be sent a reward code within 5 business days):

SURVEY 1 END

SURVEY 2 BEGIN

Q20 If the facilitator has announced the start of the second survey, please continue. If not, please WAIT for further instructions before continuing.

Q21 This second survey continues our questionnaire on employment and the Saudization of the labor market.

--**TREATMENT**--
Q22 The charts on the following pages show the responses of the other participants in the room to the questions you just answered. Please carefully review all the information displayed before answering the question that follows. You may wish to position your device in landscape orientation for optimal display. Depending on your device, you may need to scroll to reach the bottom of the page where you will see the option to continue the survey.

//Embedded Feedback Links//]

Q23 Please read the following information about a software platform called ███.

(Description of the job-matching service.)

███ provides an online platform and mobile app that connects Saudi women to job opportunities. ███ provides access to a large pool of job postings and to career fairs. You now have the opportunity to choose between two options:

      A) Sign your wife up for free access to the premium version of the ███ platform and app.
  OR
      B) Receive a $5 Amazon gift card

{if A}:
Q24 You have chosen to give your wife free access to the premium version of the ███ platform and app. Please enter her contact information in the fields below. Your wife will be contacted directly by ███ to sign up for the service.
    Wife's Name (1)
    Wife's Email (2)
    Wife's Phone Number (3)

{if B}:
Q25 You have chosen to receive a $5 Amazon gift card. Please enter your email below so that we may email you the reward code after the survey concludes (within 5 business days):
    Email (1)

Q26 What percent of private sector firms do you think have semi-segregated environments?
    (0-100) %

Q27 Thank you for your participation. You will be invited to participate in a follow-up survey in the coming months.

<div align="center">

SURVEY 2 END

</div>

-END-

Follow-up Script

Hello,

Am I speaking with {participant name}?
　　{if NO:} -> What would be the best way to reach them?
　　　　{record CONTACT/SCHEDULE info}
　　　　Thank you for your time! {-END CALL-}
　　{if YES, continue}

My name is _____ and I am calling to follow up about a study you participated in at the ▮▮▮▮▮ *about the Saudi labor market 3 months ago. This research is being conducted to understand the effect that new companies such as* ▮▮ *, which provide job opportunities for Saudi women, are having on the labor market. I would like to ask you a few more questions about the employment of you and your wife to conclude our study.*

Please keep in mind that your participation is voluntary. The information you provide will remain completely anonymous and you will not be identified by any of the answers you provide during this call or the answers provided during the previous study session. This conversation is NOT being recorded but notes will be taken to record your answers. All information disclosed during this survey will be kept confidential and stored in a secure location. I can supply you with contact information regarding this study upon request.

Do you have any questions before we get started?
{Record CONSENT: YES or NO}
　　{If NO:} Thank you for your time. {-END CALL-}
　　{if YES, continue}

Ok, let's begin.

1. *Are you currently employed?* {YES/NO}
2. *Was your wife employed 3 months ago?* {YES/NO}
　　{if YES:}　　-> Was she working at home or outside the home?
　　　　{if OUTSIDE:} -> Was this job outside the home more or less than 30 hours/week?
　　　　　　-> What was her occupation?
3. *Is your wife currently employed?* {YES/NO}
　　{if YES:}　　-> Is she working at home or outside the home?
　　　　{if OUTSIDE:} -> Is this job outside the home more or less than 30 hours/week?
　　　　　　-> What is her occupation?
4. *Has your wife applied for a job in the last 3 months?* {YES/NO/UNCERTAIN}
　　{if YES:}　　-> Is this for a job at home or outside the home?
　　　　{if OUTSIDE:} -> Is this job outside the home more or less than 30 hours/week?
5. *Has your wife been interviewed for a job in the last 3 months?* {YES/NO/UNCERTAIN}
　　{if YES:}　　-> Is this for a job at home or outside the home?
　　　　{if OUTSIDE:} -> Is this job outside the home more or less than 30 hours/week?
6. *Does your wife have any job interviews currently scheduled?* {YES/NO/UNCERTAIN}
　　{if YES:}　　-> Is this for a job at home or outside the home?
　　　　{if OUTSIDE:} -> Is this job outside the home more or less than 30 hours/week?
7. *What district in Riyadh do you live in?*
8. *If given the opportunity, would you sign up your wife for driving lessons?*

For each statement, please tell me how many people you think would agree with the statement if we asked 30 randomly selected people from your neighborhood.

If we asked 30 randomly selected residents of your neighborhood if they agreed with the following statement, how many do you think would agree?

9.   *In my opinion, the minimum wage for Saudis (SAR 3000) should be kept at its current level.*

What about the following statement?

{Looped statements:}
10.   *In my opinion, women should be allowed to work outside of the home.*
11.   *In my opinion, a woman should have the right to work in semi-segregated environments.*

Thank you–this concludes the questions in this follow-up survey. Thank you very much for your participation. Do you have any further questions?

> {if participant has further questions:} -> *I would be glad to provide you with further information describing the purpose of this study and the way the information you have provided will be used. Would you like me to send you a detailed information sheet by email?*
> > {if YES:}   -> *What is your email address?* {Record EMAIL}

{prompt for additional notes}

-END-

Online Survey

[Consent Form (all)]

I consent (1)
I do not consent (2) <-{cannot continue}

Q1 Please fill in the following fields.
   What is your age? (1)

Q2 What is your marital status?
      Single (1)
      Married (2)
      Widowed (3)
      Divorced (4)

Q3 What is the highest education level you have achieved?
      None (1)
      Primary (2)
      Secondary (3)
      Middle (4)
      Diploma (5)
      Bachelor (6)
      Masters (7)
      Ph. D (8)
      Other Post-graduate (9)

Q4 Have you ever been employed?
      Yes (1)
      No (2)

Q5 Are you currently employed?
      Yes (1)
      No (2)

    Q5.1 *(if Q5 == "Yes")* Is this job at home or outside the home?
        At home (1)
        Outside the home (2)

        Q5.2 *(if Q5A == "Outside")* Is this job outside the home more or less than 30 hours/week?
           More than 30 hours/week (1)
           Less than 30 hours/week (2)

Q6 Is your wife currently employed?
      Yes (1)
      No (2)

    Q6.1 *(if Q5 == "Yes")* Is this job at home or outside the home?
        At home (1)
        Outside the home (2)

Q6.2 *(if Q5A == "Outside")* Is this job outside the home more or less than 30 hours/week?

      More than 30 hours/week (1)

      Less than 30 hours/week (2)

Q7 What city do you live in?

1. Riyadh
2. Jeddah
3. Mecca
4. Medina
5. Al Ahsa
6. Ta'if
7. Dammam
8. Buraidah
9. Khobar
10. Tabuk
11. Qatif
12. Khamis Mushait
13. Ha'il
14. Hafar Al-Batin
15. Jubail
16. Al-Kharj
17. Abha
18. Najran
19. Yanbu
20. Al Qunfudhah
21. Other

Q8 What district do you live in within {city}?

      _____ (1)

Q8.1 What is your monthly household income (SR)?

      0-3,000 (1)

      3,000-6,000 (2)

      6,000-9,000 (3)

      9,000-12,000 (4)

      12,000-15,000 (5)

      15,000-20,000 (6)

      20,000-25,000 (7)

      25,000-30,000 (8)

      30,000-40,000 (9)

      40,000-50,000 (10)

      50,000-75,000 (11)

      75,000-100,000 (12)

      100,000-150,000 (13)

      150,000+ (14)

Q9 How many children do you have?

      (dropdown, 0-8+)

---**CONTROL**---

We would like to ask you some questions related to employment and the Saudization of the labor market. If you are not sure about your answer, please answer to the best of your ability.

All your answers are completely anonymous.

Q10C In this question, you will be presented with **three** statements. Please read all three statements carefully and consider whether or not you agree with each statement. Then, please indicate the number of statements (from 0 to 3) that you agree with. Note that we are not interested in which statements you agree with, only **how many**.

{in randomized order}
1. *In my opinion, the minimum wage for Saudis (SAR 3000) should be kept at its current level.*
2. *In my opinion, the current unemployment insurance system (Haafez) is good for the economy.*
3. *In my opinion, Saudi nationals should receive privileged access to job vacancies before expatriate workers.*

How many of the statements above do you agree with?

    0 (1)
    1 (2)
    2 (3)
    3 (4)

On the following page, you will be presented with a single statement. First, you will be asked whether you personally agree with the statement. You will then be asked to guess how many of the other participants in the study agreed with the statement. The participant who guesses most accurately will receive a **$50** Amazon gift card.

Q11C Do you agree with the following statement?
In my opinion, women should be allowed to work outside of the home.
    Yes (1)
    No (2)

Q12C There are 1500 total participants in this study who are all married Saudi males. These participants have been recruited from different cities across Saudi Arabia to be a representative sample of all married males in Saudi Arabia aged 18-35. If you had to guess, what percentage of the other participants do you think reported agreeing with the same statement?
In my opinion, women should be allowed to work outside of the home.
    0-100 % (1)

Q13C How confident are you about your guess regarding the answers of the other participants?
    1 (1) <-Not at all confident
    2 (2)
    3 (3)
    4 (4)
    5 (5) <-Very confident

Q14C We will determine the overall prize winner after tabulating the responses. If you would like to receive a $50 Amazon gift card reward in case you are selected as the winner, please enter your email below so that we may email you the reward code after the survey concludes:
    _____(1)

---**TREATMENT**---

We would like to ask you some questions related to employment and the Saudization of the labor market. If you are not sure about your answer, please answer to the best of your ability.

All your answers are completely anonymous.

Q10T In this question, you will be presented with **four** statements. Please read all four statements carefully and consider whether or not you agree with each statement. Then, please indicate the number of statements (from 0 to 4) that you agree with. Note that we are not interested in which statements you agree with, only **how many**.

{in randomized order}
1.   *In my opinion, the minimum wage for Saudis (SAR 3000) should be kept at its current level.*
2.   *In my opinion, the current unemployment insurance system (Haafez) is good for the economy.*
3.   *In my opinion, Saudi nationals should receive privileged access to job vacancies before expatriate workers.*
4.   *In my opinion, women should be allowed to work outside of the home.*

How many *of the statements above do you agree with?*

    0 (1)
    1 (2)
    2 (3)
    3 (4)
    4 (5)

Q12T There are 1500 total participants in this study who are all married Saudi males. These participants have been recruited from different cities across Saudi Arabia to be a representative sample of all married males in Saudi Arabia aged 18-35. If you had to guess, what percentage of the other participants do you think would privately agree with the same statement?
In my opinion, women should be allowed to work outside of the home.
    0-100 % (1)

Q13T How confident are you about your guess regarding the private opinions of the other participants?
    1 (1) <-Not at all confident
    2 (2)
    3 (3)
    4 (4)
    5 (5) <-Very confident

-END-