BETA
Bureau
d'économie
théorique
et appliquée

Documents de travail

« Norms Behind Closed Doors: A Field Experiment on Gender Norm Misperceptions and Maternal Employment Decisions in Couples»

Auteurs

Marie Boltz, Monserrat Bustelo, Ana María Díaz, Agustina Suaya

Document de Travail n° 2026 – 08
(*Version modifiée du WP 2025-37)*

Mars 2026

Bureau d'Économie
Théorique et Appliquée
BETA

https://www.beta-economics.fr/

Contact :
jaoulgrammare@beta-cnrs.unistra.fr

Université de Strasbourg · cnrs · Université de Lorraine · INRAE · AgroParisTech · Université Haute-Alsace

Norms Behind Closed Doors: A Field Experiment on Gender Norm Misperceptions and Maternal Employment Decisions in Couples

Marie Boltz[*] Monserrat Bustelo[†] Ana María Díaz[‡] Agustina Suaya[§¶]

March 5, 2026

Abstract

We study whether pluralistic ignorance about societal and spousal support for maternal employment sustains gender gaps in women's labour-market outcomes. Using a representative sample of 1,732 cohabiting couples with young children in Bogotá, we document near-universal first-order support for working mothers but substantial underestimation of others' support, especially that of fathers, and frequent misperceptions of the partner's views. We then implement a randomised information intervention that delivers personalised feedback on prevailing local attitudes toward maternal employment. The intervention narrows key second-order belief gaps about community and spousal support, while leaving first-order attitudes essentially unchanged. Treated men are more likely than control men to nominate their wife rather than themselves for a career-building course. One to two months later, treated women report more intensive job search and treated men place greater weight on work–family balance. Effects are concentrated among women who are already active in the labour market, underscoring both the potential and the limits of norm-correcting information in a context with high support for women's work but large misperceptions.

JEL Codes: J16, J21; D91, C93

Keywords: Gender norms, Female Employment, Pluralistic ignorance, RCT.

[*]Corresponding author. University of Strasbourg, BETA, CNRS; CEREQ. 61 avenue de la Forêt Noire, 67085 Strasbourg, France. m.boltz@unistra.fr

[†]Gender and Diversity Division, Inter-American Development Bank. Email: monserratb@iadb.org.

[‡]Department of Economics, Pontificia Universidad Javeriana. Email: a.diaze@javeriana.edu.co.

[§]Gender and Diversity Division, Inter-American Development Bank. Email: agustina.suaya@gmail.com

[¶]We thank an anonymous IADB referee, Karen Macours, Bailey Palmer, Maria Carmela Lo Bue, Eliane El Badaoui, Salvatore di Falco, Habiba Djebbari, Libertad Gonzalez, Diana Contreras, Mathieu Lefèbvre, and Philippe de Vreyer for their helpful suggestions. We also thank participants at NEUDC 2025, SITES conference 2025, and the Care Economy Workshop in 2025 (U. Leiden and Rotterdam), the economic seminar at U. Paris Sud (RITM) for valuable comments. We thank Clara Blech for excellent research assistance. We gratefully acknowledge financial support from the Inter-American Development Bank (IADB). All remaining errors are our own. This study was pre-registered in the AEA Social Science Registry under Trial ID AEARCTR-0014648 and received the IRB certificate from Pontificia Universidad Javeriana on 2024-04-24. During the preparation of this work the authors used Claude Code for editing assistance. The authors reviewed and edited the content as needed and take full responsibility for the publication. The replication data and code are available at https://doi.org/10.7910/DVN/QYWHLA.

1 Introduction

Mothers' labor-force participation remains far below fathers' despite strong stated support for women's employment in most countries (Bursztyn et al., 2023). In Bogotá, Colombia, the setting for our study, 89 percent of fathers with children under six support mothers of young children working outside the home, yet they estimate that only 61 percent of other men agree.[1] The gap between fathers' own support and their perception of other fathers' support defines *pluralistic ignorance*, the systematic underestimation of others' agreement with a position one privately holds (Prentice and Miller, 1993). Bursztyn et al. (2023) document large discrepancies between actual and perceived support for working women in sixty countries. These gaps function as an informational friction, sustaining gender-unequal outcomes even where preferences are progressive (Bicchieri, 2016).

This paper asks whether correcting pluralistic ignorance about support for *maternal* employment, specifically for mothers of young children working outside the home, narrows misperceptions (including misperceptions about one's own partner) and whether these belief updates affect intra-household decisions and women's labor-market behavior. We focus on couples with young children, a stage at which gender gaps in labor-force participation and earnings sharply widen (Cortés and Pan, 2023; Goldin, 2014; Kleven et al., 2019). Misperceived norms are especially consequential within households, where decisions about women's labor supply are negotiated between partners. Even when both spouses privately support maternal employment, their shared underestimation of community norms can create a social-permission friction: each partner may hesitate to advocate for the wife's career if they believe broad societal disapproval is high, regardless of what they know their partner thinks. Within-couple misperceptions in our sample run in the *optimistic* direction — spouses slightly overestimate each other's support — so the primary coordination friction operates at the community level rather than within the couple. This community-level friction is likely to be particularly salient when couples look to other couples as their reference group for what constitutes acceptable household behavior (Cudeville and Recoules, 2014). Under this logic, a husband may hesitate to support his wife's career not because he doubts her wishes, but because he believes most men in comparable households would not do the same. Inaccurate second-order beliefs about the community therefore constrain intra-household bargaining and coordination, not merely reflect them (Bernhardt et al., 2018).

We study these questions using a randomized information experiment with 1,732

[1]These figures come from our representative survey of parents with at least one child under six in Bogotá, described in Section 3. Appendix Table B.1 reports the full set of baseline beliefs.

cohabiting couples (3,464 adults) with at least one child under six in Bogotá, a large middle-income city where female labor-force participation (approximately 64,2%, DANE 2024) exceeds the Latin America and Caribbean average of around 52% (World Bank, 2022) and ranks among the highest of major cities in the region. A distinctive feature of our design is the randomization and surveying *both* partners within each couple, which allows us to measure mutual misperceptions, both about societal norms and about each other's views, and to observe how belief updates translate into intra-household decisions. Both partners participate in the baseline survey, but each receives the informational intervention individually through a personal WhatsApp message. Our design has two stages. In the first stage, we elicit first-order beliefs, community second-order beliefs, and spousal second-order beliefs on whether mothers with young children should be free to work outside the home. This allows us to measure actual support for maternal employment and to quantify the gap between true attitudes and perceived norms in the line of recent approaches to measuring social norms as shared expectations (Bicchieri, 2006, 2016; Krupka and Weber, 2013). In the second stage, we randomize couples to treatment or control. Treated individuals receive personalized WhatsApp messages presenting the true level of support measured in the first stage; control individuals receive information on attitudes toward corporate subsidies for green transport, delivered through the same channel, format, and schedule. We randomize at the couple level and deliver individualized messages to each partner, allowing us to study how spouses differentially respond to the same information.

We evaluate the effects of this intervention along a hypothesized causal pathway linking belief updating to intra-household decisions and labor-related outcomes. Specifically, we examine whether updated second-order beliefs affect (i) own preference and the perceptions of community and spousal support, (ii) the allocation of a scarce career-development opportunity within the household, and (iii) short-run labor-market behaviors and aspirations. The course nomination, a single scarce career-building opportunity that can be allocated to either spouse, serves as a revealed-preference measure of intra-household prioritization of women's careers, while labor behaviors and aspirations capture short-run downstream responses over the one-to-two-month horizon. Our analysis accounts for partial exposure to the intervention and survey attrition through a two-step inverse-probability regression adjustment.

Three main findings emerge. First, private support for mothers of young children (under six) working outside the home is very high at baseline (89% of men; 91% of women), yet both men and women substantially underestimate societal support. The largest gaps concern perceived support among men: husbands estimate that only 61%

of men agree with the statement, against an actual rate of 89%—a gap of 28 percentage points; the corresponding gap for women's perceived support is around 12 percentage points. Misperceptions also extend within couples, where partners frequently misjudge each other's views; women are particularly likely to underestimate their husbands' support for equal task-sharing.

Second, the information intervention corrects specific second-order beliefs while leaving first-order attitudes largely unchanged. Treated respondents revise their estimates of community support toward the true norm, with updates on the order of 7 to 9 percentage points one to two months after information treatment. Community-level information also spills over to spousal beliefs since treated men become 6.3 percentage points more likely to perceive their wife as supportive of maternal employment, and treated women update beliefs about their husband's support for equal task-sharing by a similar margin. Own attitudes, already near ceiling, show little additional movement.

Third, in a real intra-household allocation decision over a scarce career-development opportunity, treated men are 9 percentage points (23%) more likely to nominate their wife rather than themselves —a meaningful shift in a zero-sum choice that carries direct personal cost. Though estimated with some imprecision, we interpret it as credible evidence that corrected norms alter how men prioritize their wife's career investments. Women's allocation choices, already strongly favoring self-nomination in the control group (84%), change little. At the one-to-two-month follow-up, treated women report more intensive job search and treated men place greater weight on work–family balance.

Baseline labor-market attachment strongly conditions treatment effects. The main behavioral responses are concentrated among couples in which the wife is employed or actively job-seeking at baseline, whereas households with inactive women appear largely unresponsive. Effects seem also to depend on within-couple exposure: joint exposure produces the strongest and most coherent belief updates. Indirect exposure through a treated spouse is comparatively weak and often negligible, particularly for men.

A large literature documents how gender norms shape women's economic opportunities (for recent reviews, see Anderson, 2025; Olivetti et al., 2024). We contribute by focusing on a specific friction: systematic underestimation of social and spousal approval (pluralistic ignorance). Global evidence shows misperceptions are widespread and patterned across contexts, with the largest errors about men's support (Bursztyn et al., 2023). In the United States, Cortés et al. (2024) show how perceptions of prevailing social norms about mothers working shape individual attitudes, underscoring the behavioral relevance of second-order beliefs. Because work decisions are often negotiated within couples, belief gaps matter even when private attitudes are supportive. In India,

Bernhardt et al. (2018) show men substantially overestimate the community disapproval of women working, which distorts intra-household bargaining and constrains women's labor-market engagement. Bohren et al. (2023) formalize the distinction between tastes and inaccurate beliefs in discrimination settings, reinforcing the idea that policy should target misperceptions rather than preferences when beliefs are wrong. In our setting, the relevant misperceptions are second-order beliefs about social and spousal approval.

A small set of experiments tests whether correcting misperceived gender norms changes beliefs and behaviors. In Saudi Arabia, Bursztyn et al. (2020) show that correcting young married men's underestimation of peers' support for women working outside the home increases men's costly willingness to facilitate their wives' job search (i.e., signing them up for a job-matching service) and raises wives' subsequent job applications and interviews; in an additional recruitment experiment, informing women about true support induces switching toward a higher-paying outside-the-home job option. In an online experiment in Indonesia, Cameron et al. (2026) first measure perceived norms in a sample of 1,050 panel-recruited respondents, then use the estimated norms from that sample to inform a second online panel of more than 4,000 respondents. They find a 25% increase in the probability of selecting an online career-mentoring course instead of an equal-value shopping voucher.In related work, Laszlo et al. (2025) run a lab-in-the-field experiment in Paraguay that elicits both first-order and second-order (common) beliefs about gender norms, documents substantial misperceptions of others' views, and shows that a norm-shifting intervention increases individuals' normative beliefs toward a more equitable division of household responsibilities.

These studies typically observe only one partner's beliefs and therefore cannot measure mutual misperceptions within couples or test whether belief corrections affect the intra-household allocation of scarce career investments. A broader lesson from this body of work is that the relevant barrier in many contexts is not preferences themselves but inaccurate beliefs about what others prefer — and that correcting the latter can shift behavior even where the former are already progressive. Our study contributes to this literature in three ways.

First, we document persistent pluralistic ignorance in a Latin American urban context with relatively high female labor-force participation, showing that substantial belief gaps remain salient even where private support for maternal employment is very high. Crucially, the credibility of any norm-correction intervention depends on the accuracy of the norm being communicated for the reference group that actually shapes behavior — which in turn requires a representative baseline. For couples making intra-household decisions about maternal employment, the relevant reference group is not society in

the abstract but other comparable couples: misperceptions about what households like theirs do generate a particularly binding coordination friction (Cudeville and Recoules, 2014), even when spouses are privately aligned. We are the first study in this literature to deliver such an intervention to couples using a norm estimated on a probability-sampled population of those very households — ensuring the disclosed norm reflects the actual distribution of attitudes among parents comparable to those in our sample. Focusing on couples with young children, when gender gaps in labor supply and earnings sharply widen (e.g., Goldin, 2021; Kleven et al., 2019), further underscores the relevance of these misperceptions for maternal employment decisions.

Second, we randomize and survey *both* partners within each couple, which allows us to (i) measure mutual misperceptions about community and spousal approval and (ii) study within-couple exposure patterns to the same information. These within-household belief measures are typically absent in single-respondent designs, yet they are central for understanding how norms shape bargaining and coordination (e.g., Bertrand et al., 2015; Cassidy et al., 2021; Majlesi, 2016). We show that information about community norms not only corrects societal misperceptions but also leads individuals to revise beliefs about their own partner's views.

Third, we provide suggestive evidence that correcting misperceived norms can affect the intra-household allocation of a scarce career investment. Because the course nomination is a zero-sum, revealed-preference choice over a scarce asset, it provides direct evidence on intra-household prioritization, a step beyond costly stated support or hypothetical willingness-to-allow outcomes. Consistent with evidence in which misperceived social norms (second-order beliefs) can constrain intra-household bargaining and coordination (Bernhardt et al., 2018), treated men are suggestively more likely to prioritize their wives' careers. Short-run behavioral responses are concentrated among women attached to the labor market. This suggests that information can shift beliefs and some behaviors, but that broader changes in maternal employment may also require complementary policies addressing structural barriers such as childcare availability, workplace flexibility, and labor demand (Kleven et al., 2019; Olivetti and Petrongolo, 2017). Effects are especially limited for women out of the labor force, highlighting scope for future research on sustained exposure to accurate norms and on complementarities between norm-correction and constraint-relaxing interventions.

The finding that misperceptions can sustain gender-unequal outcomes despite progressive private attitudes has implications beyond Latin America. Across contexts where attitudes are modernizing, informational frictions rather than deep preference gaps may be the binding constraint on women's labor supply. Our heterogeneity results, however,

introduce an important qualification: effects concentrate among households where the wife is already labor-market attached, while inactive women respond little. This points to a clear complementarity between norm feedback and policies that relax structural constraints — childcare, workplace flexibility, labor demand — with the two approaches targeting different margins of the same problem. More broadly, in contexts where gender attitude gaps are widening, our results suggest that norm-correcting information may be most effective where conservatism partly reflects misperceptions about peers rather than deeper grievances — a distinction that future research should take seriously.

The remainder of the paper is organized as follows. Section 2 presents the experimental design and the theory of change. Section 3 describes the data with a focus on the baseline beliefs elicitation. Section 4 outlines the empirical approach, including data collection, randomization, and outcome measures. Section 5 discusses the main results, including belief updating, intra-household dynamics, and job market behaviors. Finally, Section 7 concludes.

2 Experimental Design

We study cohabiting heterosexual couples with at least one child under six years of age in Bogotá. The baseline survey was conducted between July and September 2024 and interviewed 1,732 households, comprising 3,464 adults (two partners per household). Households were randomly assigned in equal proportions (1:1) to treatment and control. Randomization was stratified by wife's baseline employment status (inactive, working and unemployed) and the husband's baseline first-order support for maternal employment, using permuted blocks within strata to ensure balance across groups.

Our experimental design has three sequential stages implemented with a consistent sample of cohabiting couples: (i) a baseline survey to elicit first- and second-order beliefs, and socio-economic characteristics; (ii) an informational intervention delivered via WhatsApp; and (iii) a follow-up survey to measure belief updating, job search, and labor market aspirations.

Baseline survey. Between July and September 2024, we interviewed 1,732 cohabiting heterosexual couples in Bogotá, i.e., 3,464 adults, with at least one child under six (Sample (1) in Appendix Figure C.1). Both partners were surveyed, but each was interviewed individually. Data collection combined in-person and telephone interviews, depending on participants' availability.

The questionnaire included eight statements on gender norms and maternal employment; here we focus on the item "Mothers of children under six years of age should be free to work for pay outside the home." For every statement, we elicited three measures: (i) first-order beliefs, capturing respondents own attitudes (agree/disagree); (ii) community second-order beliefs, elicited as respondents' estimate (on a scale of 0 to 100) of how many out of 100 fathers (respectively, mothers) with young children in Bogotá would agree with the statement; and (iii) spousal second-order beliefs, elicited as respondents' prediction of their partner's position.

The baseline sample was designed to be representative on households with at least one child under six in Bogotá. This feature is key to allow us to compute the *actual* distribution of beliefs among parents of young children and to compare these distributions to the respondent's second-order beliefs. We measure pluralistic ignorance as the gap between a respondent's perceived level of community support and the *true* share of agreement observed in the baseline sample.

The survey also collected detailed demographics characteristics, household composition, employment status, and other socioeconomic covariates, which were used to define the randomization strata, to improve precision in the empirical analysis and correct for attrition in later rounds.

Midline survey and Information intervention Following the baseline survey, at the end of October 2024, the 1,732 households were randomly assigned to treatment or control using the stratified design described above. Randomization was at the household level: both partners in a treated couple received the treatment, and both partners in a control couple received the placebo. We chose couple-level randomization because partners were likely to discuss the messages, and assigning different content to members of the same household would risk confusion or suspicion. Within each treated household, WhatsApp messages were sent individually to each partner through a personal chatbot interaction, with content personalized to the respondent's own baseline answers. The chatbot addressed a single norm: "Mothers of children under six should be free to work for pay outside the home."

The intervention used a discrepancy-based feedback strategy. Each respondent (i) was shown their own baseline estimations of the share of men and women agreeing with the statement; (ii) was asked whether they believed their estimates matched the true values; (iii) was then shown the actual share of agreement in Bogotá, computed from the baseline data and presented visually; and (iv) was asked to rate the discrepancy as "interesting," "irrelevant," or "disappointing." The same sequence was repeated for beliefs about men's

and women's attitudes, with the order of presentation randomized.

Control households (866 couples) received placebo content through the same channel, on the same schedule and with the same format, but on an unrelated topic: attitudes toward corporate subsidies for green transport. Because delivery mode, frequency, and user experience were identical across groups, any differences in the outcomes isolate the effect of correcting misperceived gender norms.

After the information module, the chatbot invited all participants to claim a single per-household slot for an online career-development and skill-building course. Participants were asked whether they wished to enroll themselves or offer the course opportunity to their spouse or partner, given that only one slot was available per household. This choice created a clear intra-household trade-off and serves as a behavioral outcome in our analysis. The course was a real, concrete opportunity (not hypothetical)—an online professional program—so the decision carried genuine stakes for both spouses. The zero-sum design, however, means that men face a direct personal cost when prioritizing their wife; our estimates of men's willingness to do so may therefore represent a lower bound relative to settings where career investments are not mutually exclusive.

Although WhatsApp usage is nearly universal in the sample, engagement with the chatbot was limited. In total, 1,233 of 3,464 adults (36 percent) clicked through the content, meaning saw the treatment information (618) or the placebo (615). We refer to respondents who interacted with the chatbot as Sample (2) (see Appendix Figure C.1). In Section (4.3) we discuss this selective take-up and present a strategy to address it with a two-step inverse-probability weighting procedure.

Endline survey Between November 2024 and January 2025, we recontacted 1,382 individuals from the original baseline sample by phone, a response rate of approximately 40 percent, referred to as Sample (3). Achieving this required sustained effort: interviewers sent a personalized SMS reminder before each call attempt, scheduled calls across multiple days and times to accommodate working respondents, and incrementally increased participation incentives for non-respondents after several failed contact attempts. Phone follow-up response rates in urban household panel studies vary substantially, ranging from roughly 40 to 60 percent depending on topic salience, survey mode, and fieldwork intensity (Ali et al., 2022; Dabalen et al., 2016; World Bank, 2021); our 40-percent rate, achieved without physical tracking and for a non-emergency topic conducted by callers the household does not recognize, falls at the lower end of this range and reflects the structural difficulty of phone-based panel follow-up rather than incomplete effort. WhatsApp access is near-universal in the baseline sample, so low midline engagement reflects

a behavioral choice not to interact with the chatbot rather than a technology barrier. The compliers are therefore the population that a low-cost digital norm-correction campaign deployed at scale would reach.

Among these respondents, 806 individuals (399 treated and 407 control) had previously participated in the midline survey (WhatsApp chatbot) – referred to as Sample (2∩3); while the remaining 576 individuals (280 treated and 296 control) were reached only at endline — referred to as Sample (3\2) (see Appendix Figure C.1).

All endline respondents first completed a module measuring first-order beliefs and community second-order beliefs. This design allows us to assess, among respondents in Sample (2∩3) who participated in both Midline and Endline surveys whether, whether belief and perception updates induced by the WhatsApp intervention persist several months later, prior to any additional exposure.

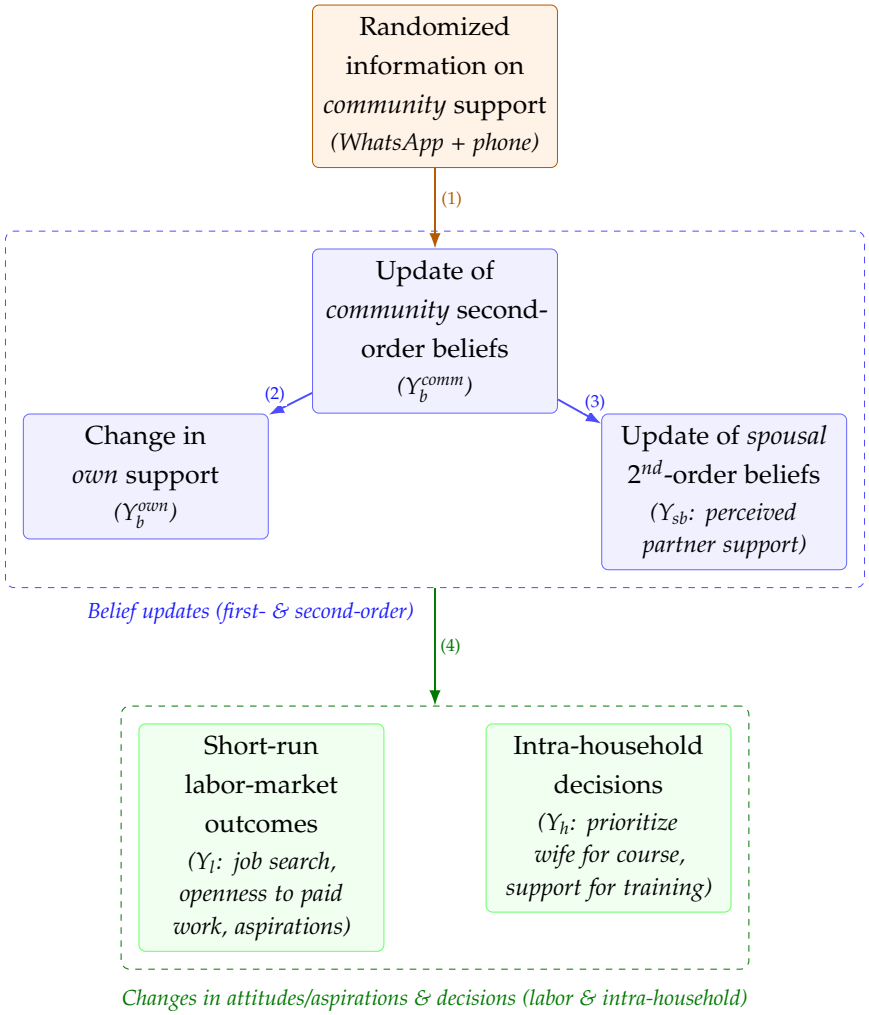
Because engagement with the WhatsApp module was incomplete, we re-administered the information during the endline phone survey to all respondents assigned to the treatment group. Interviewers followed a standardized script replicating the WhatsApp sequence. As a result, all treated respondents received the information at least once, but treatment intensity varied: some respondents were exposed only at endline (Sample (3\2)), while others were exposed twice, at midline and endline (Sample (2 ∩ 3)). Within the endline call, the reinforcement script was administered after respondents had already completed the community-belief module, so that beliefs about societal norms are measured under a single exposure (for Sample 2 ∩ 3) and are not contaminated by the endline re-exposure. Second-order beliefs about the spouse, by contrast, were elicited after the reinforcement script and therefore reflect cumulative exposure; we discuss the implications for interpretation when presenting those results (Section 5.1).

We therefore estimate treatment effects on different samples depending on the outcome of interest, in a manner consistent with experimental timing. In particular, outcomes such as employment status and labor-market aspirations are analyzed primarily among respondents exposed at midline (2∩3). We describe in detail the outcomes in section 4.2.

Theory of change. The analysis tests whether correcting misperceived community norms triggers belief updating, reshapes intra-household expectations, and influences short-run labor-market behaviors (Figure 1). The hypothesized pathway runs in three steps: (i) the information corrects misperceptions about community support for maternal employment; (ii) revised community-norm perceptions prompt updates to first-order beliefs and spousal second-order beliefs; and (iii) updated beliefs reshape intra-household allocation decisions and labor-market behaviors. Because each partner receives the infor-

mation individually, information may also spill over through within-couple discussion, amplifying or attenuating the direct treatment effect on each spouse. In our sample, within-couple misperceptions are small and optimistic (Section 3), so the key mechanism is the removal of a *community-level* social-permission barrier rather than correction of spousal pessimism. This barrier is particularly salient when couples use other couples as their reference group for acceptable household behavior (Cudeville and Recoules, 2014), making misperceptions about what comparable households do the binding friction. The correction of these misperceptions is consistent with *generalized introspection* (Boltz et al., 2025), whereby credible information about social norms realigns perceptions of the broader community and, secondarily, prompts reassessment of the partner's views.

Figure 1. Theory of change: how correcting misperceived gender norms may affect intra-household and labor market outcomes



3 Data description

Sample description

The baseline sample consists of 1,732 cohabiting heterosexual households in Bogotá, comprising 3,464 adults (two partners per household), each with at least one child under six years of age. Table 1 reports descriptive statistics at the household level. Unless otherwise specified, individual-level outcomes are reported at the respondent level using post-stratification weights.

The average household includes 3.8 members and 1.13 children under six; 27.6 percent of the households report that at least one child under six who is not enrolled in childcare or early education, and 32.0 percent report that at least one household member requires permanent care. Households are distributed across income categories as follows: 28 percent low income, 60 percent middle income, and 12 percent high income, closely mirroring the income distribution of the city.[2]

Table 1. Household and Care Context

	Mean	Std. Dev.
Survey modality		
Household composition		
Household size	3.810	1.05
Children under 6	1.129	0.37
Any child <6 not enrolled	0.276	0.45
Member needs permanent care	0.320	0.47
Income category		
Low income	0.280	0.45
Middle income	0.600	0.49
High income	0.120	0.33
Observations	1,732	

Notes: Means are proportions unless stated otherwise. "Any child <6 not enrolled" equals 1 if the household reports at least one unenrolled child under six. "Member needs permanent care" equals 1 if any household member requires ongoing daily assistance.

Table 2 compares baseline characteristics of husbands and wives. Men are, on average, 2.8 years older than women and substantially more likely to be employed (90.5 percent vs. 52.0 percent). Conditional on employment, men work 11.1 more hours per week. Women report more job flexibility and higher work–family compatibility, yet they are more often inactive (41.7% vs. 4.5%).

[2]Interviews were mixed-mode: 68% by telephone and 32% in person. Income categories follow low: less than 1.3 million colombian pesos, middle income 1.3 to 3.9 million colombian pesos and, high income is larger than 3.9 million pesos.

Differences in job-search behavior are also pronounced. Women are more likely to report actively looking for a job, while men are more likely to report preparing to start a business and to express satisfaction with their current employment situation. These gender differences are large and precisely estimated, underscoring the asymmetric labor-market context in which belief updating and intra-household decisions take place.

Table 2. Individual characteristics by gender

	Husband		**Wife**		**Difference**
Variable	Mean (SE)	N	Mean (SE)	N	Diff
Demographics					
Age	34.859 (0.21)	1732	32.033 (0.18)	1732	2.826***
Education Levels					
Low Education	0.143 (0.01)	1732	0.108 (0.01)	1732	0.035***
Medium Education	0.695 (0.01)	1732	0.711 (0.01)	1732	-0.016
High Education	0.162 (0.01)	1732	0.181 (0.01)	1732	-0.020
Employment Status					
Employed	0.905 (0.01)	1732	0.520 (0.01)	1732	0.385***
Unemployed	0.050 (0.01)	1732	0.063 (0.01)	1732	-0.013
Inactive	0.045 (0.01)	1732	0.417 (0.01)	1732	-0.372***
Working Hours	48.730 (0.36)	1666	37.639 (0.49)	1277	11.091***
Job Flexibility					
High Flexibility	0.236 (0.01)	1666	0.330 (0.01)	1277	-0.094***
Some Flexibility	0.272 (0.01)	1666	0.315 (0.01)	1277	-0.043**
No Flexibility	0.489 (0.01)	1666	0.351 (0.01)	1277	0.138***
Work-Family Compatibility					
Compatibility Score	0.774 (0.01)	1666	0.836 (0.01)	1277	-0.062***
Job Search Activities					
Looking for a job	0.106 (0.01)	1709	0.162 (0.01)	1706	-0.055***
Preparing to start a business	0.091 (0.01)	1709	0.072 (0.01)	1706	0.019**
No, but I would like to	0.493 (0.01)	1709	0.518 (0.01)	1706	-0.025
No, satisfied	0.310 (0.01)	1709	0.249 (0.01)	1706	0.062***

Notes: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Education levels: *Low* — None, Preschool, Primary (Grades 1–5), and Lower Secondary (Grades 6–9); *Medium* — Academic High School (Grades 10–11), Technical High School, Normal School, Professional Technician, and Technologist; *High* — University, Specialization, Master's Degree, or Doctorate. Sample sizes vary by item due to non-response or applicability (e.g., only employed respondents were asked about working hours and job flexibility).

Baseline beliefs about maternal employment

Table 3 presents beliefs regarding the statement: *"Mothers of children under six should have the freedom to work outside the home"*. Support for this norm is high: 88.5 percent among husbands and 90.5 percent among wives report agreement. Despite this near-

universal private support, both husbands and wives substantially underestimate societal support—particularly among men. Husbands estimate that 61.0 percent of men agree with the statement, while wives estimate 55.7 percent, compared to an actual agreement rate close to 90 percent in the baseline sample. Perceptions of women's support are higher and closely aligned across genders, at around 80 percent.

To assess within-household misperceptions, we compare each respondent's belief about their partner's support for maternal employment with the partner's own reported belief. Specifically, for husbands the average spousal second-order belief is 93.9 percent, while wives' own reported support is 90.5 percent, yielding a gap of 3.4 percentage points. For wives, the corresponding figures are 89.9 percent and 88.5 percent, resulting in a gap of 1.4 percentage points. These within-couple gaps indicate that misperceptions persist even within households, albeit at a smaller magnitude than at the societal level. Importantly, however, the *direction* of within-couple misperceptions is opposite to the community-level pattern: whereas both spouses underestimate community support by roughly 20 percentage points, they each slightly *overestimate* their partner's support. This contrast matters for mechanism: the dominant informational friction in this setting operates at the community level, not within the couple. Spouses are not systematically pessimistic about each other's views; rather, both privately hold more progressive attitudes than they believe the broader community endorses. The intervention corrects this community-level misperception. Any subsequent updating of spousal beliefs — as measured in Table 6 — therefore reflects a secondary spillover process, not the correction of within-couple pessimism.

In contrast, beliefs about the placebo norm— "companies should subsidize public transport"—exhibit near-universal agreement and minimal gender differences. This comparison confirms that misperceptions are concentrated on gender-role norms rather than reflecting generalized pessimism or survey noise.

These patterns document large and systematic misperceptions at the community level and smaller but non-negligible misperceptions within households. These belief gaps define the informational frictions that the intervention seeks to address.

Table 3. Baseline beliefs for the target and placebo norms

	Husbands (% agreeing) (1)	Wives (% agreeing) (2)	Diff. (1-2)
A. Mothers of children <6 should be free to work			
First-order (own view)	88.50 (0.76)	90.50 (0.68)	–2.00**
Second-order: share of *men* who agree	60.98 (1.19)	55.70 (1.30)	5.28***
Second-order: share of *women* who agree	79.61 (1.10)	80.01 (1.10)	–0.40
Spousal second-order	93.90 (0.52)	89.90 (0.70)	4.10***
B. Placebo norm: companies should subsidise public transport			
First-order (own view)	93.50 (0.60)	94.90 (0.50)	–1.40***

Notes: Weighted means; robust standard errors in parentheses. "Diff." reports two-sided t-tests for equality of male and female means. The placebo norm was asked pre-randomization but shown again only to the control arm during the chatbot. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix Table B.1 reports baseline beliefs for additional gender norms.[3] Across these items, we observe a similar pattern: progressive private attitudes coexist with sizable misperceptions about others' views, particularly regarding men's support. This reinforces the interpretation that pluralistic ignorance is not limited to a single norm but characterizes broader beliefs about gender roles.

4 Empirical strategy

We estimate the causal effects of providing information about societal support for maternal employment using randomized assignment. Let $D_i \in \{0, 1\}$ denote assignment to the information treatment. Throughout, we focus on average treatment effects on the treated (ATT), and on contrasts that condition on observed exposure where explicitly stated. To address selective engagement with the intervention and differential attrition across survey waves, we rely on stabilized inverse-probability weighting combined with regression adjustment (IPWRA). Section 4.3 details the weighting strategy and diagnostics.[4]

[3]We consider statements such as "father and mother should share caregiving equally", "children suffer if mother works", "children suffer if mother works", and "there are problems if wife earns more than the husband".

[4]In Appendix E we complement the main ATT analysis with an exploratory IV-style mediation exercise. For each labor outcome, we treat follow-up perceived societal support for maternal employment and work–family balance as potential mediators and estimate a two-equation 2SLS system in which the mediator M_i is instrumented with the randomized assignment Z_i. With a single instrument and a single mediator, the product $\hat{\alpha}\hat{\pi}$ of the 2SLS slope of M_i on the outcome and the first-stage effect of Z_i on M_i coincides with the reduced-form coefficient of Z_i on the outcome, so the "mediated share" is mechanically equal to one. We therefore interpret the IV mediation results as a consistency check on the sign pattern and as suggestive evidence on the direction of the channels, rather than as precise estimates of the share mediated. All mediation specifications use the same stabilized selection weights, baseline controls, and strata fixed effects as the main IPW-weighted models.

4.1 Baseline specification

For each post-intervention outcome, our main specification in the pooled sample is

$$y_i = \beta_0 + \beta_1 D_i + \rho\, y_{i0} + \mathbf{X}_i'\gamma + \varepsilon_i, \tag{1}$$

where y_i is the post-intervention outcome, y_{i0} is the baseline measure of the same outcome (or its closest pre-specified baseline analogue), and $\mathbf{X}_i$ is a vector of pre-specified baseline covariates (including gender, age, education, and household attributes).

All regressions include stratification fixed effects, standard errors are clustered at the household level, and we report randomization-inference (Fisher) p-values computed within strata. In this specification, β_1 is the ATT of the information intervention. Gender-specific effects are obtained by interacting the treatment indicator with a female indicator in the pooled sample. Our primary approach to multiple comparisons rests on the randomization-inference p-values reported throughout. Because these exact p-values are derived from the permutation distribution of the actual treatment assignment, they provide valid finite-sample inference for each pre-specified hypothesis without relying on asymptotic approximations (Fisher, 1935; Imbens and Rubin, 2015). For additional transparency, Appendix Table H.4 reports Romano-Wolf step-down adjusted p-values (Romano and Wolf, 2005) for each outcome family; these are computed on an unweighted OLS specification and are therefore conservative relative to our preferred IPWRA estimates.

4.2 Outcomes and analysis samples

We analyze four groups of outcomes that correspond to the causal pathway described in Figure 1. Because outcomes are measured at different stages of the study, analytic samples vary by survey timing and treatment exposure. These samples are pre-specified based on participation in the midline WhatsApp module (Sample 2) and completion of the endline phone survey (Sample 3) as shown in Appendix Figure C.1.

- **Beliefs about society** (Y_b) include first-order beliefs and community second-order beliefs regarding support for maternal employment and equal task sharing. These outcomes are measured at the start of the endline survey, prior to treatment reinforcement. The sample includes respondents who participated in both the midline WhatsApp survey and the endline survey (Sample 2∩3).[5]

[5]We augment Sample 2∩3 with all endline participants of the control group, no matter their status in the midline survey as they all did not receive the treatment, to increase sample size and precision. Results remain unchanged besides precision.

- **Spousal perceptions** (Y_s) capture beliefs about the partner's views and within-couple misperceptions. These outcomes are elicited after treatment reinforcement during the endline interview and are analyzed both in the full endline sample (Sample 3) and the overlap Sample $(2 \cap 3)$. Belief updating focuses on two key perceptions: (i) support for maternal employment among mothers and fathers in Bogotá, and (ii) support for equal caregiving.

- **Intra-household allocation** (Y_h) refers to the respondent's decision of whether to claim or transfer the career-development course opportunity. We obtained this choice at midline and it is analyzed using Sample 2.

- **Labor market behaviors** (Y_ℓ) include job search effort, labor-market aspirations, and preferences for work–family balance, are measured at endline and reflect short-run responses (1 to 2 months after treatment). We analyze these outcomes separately for three groups: (i) respondents directly treated at midline and followed up at endline (Sample $2 \cap 3$)[6], (ii) respondents indirectly exposed through a treated partner within Sample 3, and (iii) respondents treated only at endline (Sample $3 \setminus 2$).

Table C.1 in Appendix C summarizes the mapping between outcomes and estimation samples, clarifying how survey timing and treatment exposure define the analytic sets.

4.3 Selection, Compliance, and Weighting

We face two sources of non-random selection. First, not all baseline respondents are observed in each analysis sample due to differential participation in the midline and endline surveys. This attrition might affect representativeness and can threaten internal validity if survey participation is correlated with potential outcomes, potentially in a way that differs by randomized assignment. Second, take-up of the information is incomplete: among respondents assigned to treatment, only a subset engages with the treatment. As a result, within the realized analysis samples, baseline covariates can be imbalanced between the treated and control arms.

We address these potential threats using a two-step inverse-probability weighting strategy, estimated separately for women and for men. In the first step, we model selection into each analysis sample using a rich baseline covariate set and construct stabilized inverse-probability weights to account for attrition. In the second step, within each selected sample, we estimate treatment effects using an inverse-probability–weighted regression adjustment (IPWRA)

[6]See previous footnote.

Step 1: Selection into analysis samples. For each attrition-defined sample

$S \in \{\text{midline (Sample 2), endline (Sample 3), both (Sample } 2 \cap 3\text{), endline only (Sample } 3 \setminus 2)\}$,

we estimate a sample-specific *selection model*

$$p_i^S = \Pr(S_i = 1 \mid D_i, X_i),$$

where D_i denotes the randomized treatment assignment at baseline and X_i is a rich vector of baseline covariates, including demographics, household composition, education, labor status, stratification fixed effects, and the full set of baseline first- and second-order beliefs (approximately 50 covariates).[7] Each selection model is estimated on the full baseline sample using a probit specification with clustering at the household level, and we construct stabilized selection weights as:

$$w_i^S = \frac{\Pr(S = 1)}{\hat{p}_i^S},$$

where $\Pr(S = 1)$ is the unconditional share of baseline respondents who belong to sample S and $\hat{p}_i^S$ is the predicted probability from the selection model.[8] Under the assumption that selection into S is ignorable conditional on (D_i, X_i) and that there is sufficient overlap in these covariates across selected and non-selected units, reweighting by w_i^S recovers the baseline distribution of (D_i, X_i) within each analysis sample. All endline estimates are computed using these selection weights as probability weights unless otherwise specified; baseline descriptive statistics are reported without weights.

Step 2: Treatment effects within samples. Within each analysis sample S, we estimate the causal effect of being assigned to the information intervention, D_i, using an IPWRA specification.[9] The IPWRA estimator combines (i) a *treatment model*,

$$e(X_i^D, \text{strata}) = \Pr(D_i = 1 \mid X_i^D, \text{strata}),$$

[7]The exact list corresponds to the union of non-belief baseline covariates and belief blocks used in the main analysis. The same X_i enters all selection models.

[8]Stabilization centers weights around one and keeps effective sample sizes comparable across specifications. As $\hat{p}_i^S$ varies with (D_i, X_i), $w_i^S > 1$ for units under-represented in S given their covariates and $w_i^S < 1$ for over-represented units.

[9]For Sample 3, the treatment group includes both individuals who interacted with the WhatsApp chatbot and individuals who did not open or respond to the messages but subsequently received the information by phone at endline. We therefore interpret $D_i = 1$ as assignment to receive the information (via WhatsApp or phone) and $D_i = 0$ as assignment to the control condition.

and (ii) an *outcome model* for each treatment status,

$$m_d(X_i^D, \text{strata}) = \mathbb{E}[Y_i \mid D_i = d, X_i^D, \text{strata}], \quad d \in \{0, 1\},$$

where X_i^D is a targeted subset of X_i including demographics, household composition, labor status, and baseline beliefs about societal and spousal support for maternal employment, and *strata* denotes the stratification variables used in the design (wife's labor-market situation and husband's first-order belief about working mothers).

For each analysis sample S, the treatment and outcome models are estimated using the stabilized selection weights w_i^S from Step 1 as probability weights, so that the rich covariate set X_i enters the IPWRA procedure through the selection correction. Let $\hat{e}_i = \hat{e}(X_i^D, \text{strata})$ denote the estimated propensity score and $\hat{m}_d(X_i^D, \text{strata})$ the estimated outcome regression for $d \in \{0, 1\}$. Defining the (selection-adjusted) number of treated units as

$$N_T^w = \sum_i w_i^S D_i,$$

a representation of the doubly robust ATT estimator is

$$\widehat{\text{ATT}} = \frac{1}{N_T^w} \sum_i w_i^S \left[D_i \big(Y_i - \hat{m}_0(X_i^D, \text{strata})\big) + (1 - D_i) \frac{\hat{e}_i}{1 - \hat{e}_i} \big(\hat{m}_1(X_i^D, \text{strata}) - \hat{m}_0(X_i^D, \text{strata})\big) \right].$$

Intuitively, IPWRA combines inverse-probability weighting based on the treatment model $e(\cdot)$ with regression adjustment based on the outcome model $m_d(\cdot)$, thereby reweighting the control group toward the covariate distribution of the treated and adjusting flexibly for remaining covariate differences. The estimator is doubly robust in the sense that the ATT is consistently estimated if either the treatment model or the outcome model is correctly specified.

All outcome specifications include stratification fixed effects and cluster standard errors at the household level; IPWRA results are reported with analytic cluster-robust variances.

Robustness to extreme attrition weights. Because the estimated attrition propensity scores range from 0.06 to 0.76, a small share of observations receive large inverse-probability weights that could unduly influence the IPWRA estimates. Appendix Table H.6 reports results under four alternative weight specifications: (i) the baseline probit-PS weights used throughout; (ii) weights winsorised at the 95th percentile; (iii) a trimmed sample that drops the 1% of observations with estimated propensity scores

below 0.10; and (iv) an alternative logit specification for the attrition model. We discuss the implications for each headline result in the relevant results sections.

Identification and testable implications. Steps 1 and 2 rely on two conditions. First, selection into each analysis sample S is ignorable conditional on (D_i, X_i) with sufficient overlap in these covariates across included and non-included units. Second, conditional on selection and stratification, the propensity score model used in IPWRA provides adequate reweighting to balance observed baseline covariates between treatment and control within each sample. While these conditions are not directly testable, they imply observable restrictions on covariate balance, overlap of estimated propensity scores, and the concentration of the resulting weights. The diagnostics below assess these implications for both the selection and treatment components. Given randomized assignment within strata, the selection adjustment in Step 1, and the diagnostic evidence on balance, overlap, and weight concentration, we view these assumptions as reasonably plausible in our setting.

Diagnostics. We address four potential threats to the validity of our design and estimates. First, baseline covariates may be imbalanced between treatment and control arms within each analysis sample, either because randomization is imperfect or because differential attrition reintroduces imbalance after randomization. Second, attrition into the analysis samples may be non-random and correlated with potential outcomes, threatening internal validity. Third, selection into engagement with the WhatsApp module may differ across treatment and control arms, undermining the placebo comparison. Fourth, the social norm disclosed to treated respondents may not accurately represent the reference group that actually shapes decisions for the subpopulation that engages.

We assess covariate balance, propensity score overlap, and weight concentration for both the selection and treatment models. Specifically, we report standardized mean differences (SMDs) for all baseline covariates, examine densities of the estimated propensity scores and the (weighted) mass outside common support, and summarize weight concentration using the effective sample size (ESS). For the treatment models, we also implement Hansen J-tests of joint balance by recasting the covariate-balance restrictions as overidentifying moments. Standardized mean differences and the J-tests refer to covariate balance, the propensity score densities and mass outside common support inform overlap, and the ESS summarizes the extent to which extreme weights affect precision.

Appendix Table D.1 and Figures D.1–D.3 report diagnostics for selective exposure to treatment (D_i). After weighting, maximum absolute SMDs are generally below 0.10 in

all samples and genders, effective sample sizes remain large, the weighted mass outside common support is small, and the J-tests never reject joint balance where defined.

Additionally, Table 4 reports unweighted OLS balance tests for the baseline treatment assignment Z_i by analysis sample, regressing Z_i on the full set of baseline covariates and stratification fixed effects. A small number of covariates (all shown in the table) are marginally imbalanced in some columns, and with such a rich covariate set some joint F-tests reject at conventional levels. However, effect sizes are small, and the subsequent SMD and weighting diagnostics indicate that any departures from exact balance at baseline are limited and are more than corrected by the selection and treatment weights.

Table 4. Selected Balance Test Results Across Samples (Dependent Variable: Treated = 1)

Variable	**Entire Sample** Sample (1)	**Midline Survey** Sample (2)	**Endline Survey** Sample (3)	**Both Surveys** Sample (2)∩(3)	**Endline Only (No Midline)** Sample (3)\(2)
Female	-0.000	-0.015	0.041*	0.112***	0.091*
	(0.005)	(0.023)	(0.017)	(0.022)	(0.038)
Age	0.001	0.001	0.004	0.003	0.007*
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Telephone access	−0.055*	−0.077*	-0.039	0.024	-0.049
	(0.025)	(0.039)	(0.039)	(0.040)	(0.052)
2nd-order belief (partner): mothers with children under 6 work outside home (SOB2P)	-0.033	-0.080	-0.077	−0.113*	0.027
	(0.033)	(0.051)	(0.051)	(0.056)	(0.087)
Observations	3,458	1,233	1,382	1,102	576
R^2	0.008	0.020	0.028	0.046	0.056
F-stat (all baseline covariates)	0.91	0.95	1.39	2.80	1.73
p-value (all baseline covariates)	0.565	0.516	0.120	0.000	0.029
F-stat (belief variables only)	0.77	0.74	1.79	1.87	1.48
p-value (belief variables only)	0.544	0.567	0.128	0.114	0.208
F-stat (non-belief covariates)	0.96	1.03	1.18	2.93	1.54
p-value (non-belief covariates)	0.495	0.417	0.282	0.000	0.088

Notes. Each column reports OLS estimates from regressing the treatment indicator Z_i on the full set of baseline covariates, including stratification fixed effects. The table lists only baseline covariates that are significant at the 10% level in at least one sample. Standard errors in parentheses, clustered at the household level. The bottom panel reports R^2, sample size, and joint F-tests for (i) all baseline covariates, (ii) belief variables only, and (iii) non-belief covariates. Full results for all baseline covariates are reported in Appendix Table D.2.

Appendix Table D.3 and Figures D.4–D.6 provide analogous diagnostics for selection into the attrition-defined samples. Selection weights reduce large unweighted SMDs to below about 0.10 within gender, with very limited mass outside common support and high ESS. These results show that, absent weighting, both effective treatment receipt (D_i) and sample inclusion are correlated with baseline covariates in ways that would

bias unweighted estimates and that the combination of selection weights and IPWRA delivers tight postweighted covariate alignment and adequate overlap. We therefore treat the IPW-adjusted estimates as our preferred specification and report unweighted OLS alongside as a benchmark. Molina Millán and Macours (2025) document that only 28% of development-economics RCTs explicitly correct for attrition in estimation, while a further 23% stop at testing whether attrition rates are balanced across arms. Our approach (i.e., combining IPWRA, Lee bounds, and the near-miss timing diagnostics described below) addresses both estimation and diagnostic gaps identified in that review.

Near-miss attrition. Unfortunately, we do not have records of the number of contact attempts per respondent at endline. As an alternative proxy for contact difficulty — in the spirit of Behaghel et al. (2015) and Molina Millán and Macours (2025) — we use the timing of the endline interview. Our phone survey ran from November 18, 2024 to January 20, 2025 (63 days). Field teams worked through the sample sequentially and returned to non-respondents; respondents interviewed in December and January therefore represent, on average, those who required more contact attempts. We compare 379 November respondents (easy to reach) with 492 December–January respondents (harder to reach) on their baseline characteristics.

The near-miss exercise is reassuring on the dimension that matters most for our intervention. The variable the treatment directly corrects — second-order beliefs about men's community support for maternal employment — does not differ between timing groups (November: 58.3; December–January: 59.4; difference of 1.2 percentage points, $p > 0.5$). First-order personal support shows a marginally significant gap of 3.7 percentage points, but two features limit its relevance. First, husband's first-order belief is one of our stratification variables, so any differential loss on this dimension is addressed directly by the randomization design and subsequently by the IPWRA selection correction. Second, to the extent that harder-to-reach respondents hold slightly less uniformly supportive attitudes, residual treatment effects on them would be at least as large as for easier-to-reach respondents — the estimated effects are therefore conservative with respect to the hardest-to-survey group. No other belief measure differs significantly across timing groups for all remaining outcomes). Appendix Table D.6 reports the full covariate balance for the two timing groups.

Engager characterization and placebo-selection diagnostics. A key concern for information experiments is whether the population that engages with the messages differs systematically from the full baseline sample, and whether selection into engagement

differs across treatment and control arms. We assess both directly.

First, engagement with the WhatsApp chatbot was nearly identical across arms: 35.8% of treated adults and 35.6% of control adults engaged (620 vs. 616 individuals), with no meaningful difference by gender (treated men: 27.6%; control men: 26.6%; treated women: 44.0%; control women: 44.6%). The symmetry in engagement rates provides initial evidence against content-driven selection. This symmetry is consistent with the latent-variable response framework of Behaghel et al. (2015), in which equal engagement rates across arms indicate equal distributions of latent reluctance to participate, providing evidence against differential selection into engagement by treatment status.

Second, we characterize the engager population and assess whether demographic selection compromises the accuracy of the disclosed norm. Appendix Table D.4 reports baseline characteristics for engagers ($n = 1{,}236$) and non-engagers ($n = 2{,}228$). Engagers differ from non-engagers on several demographic characteristics consistent with time availability and topic relevance: they are 11 percentage points more likely to be economically inactive and 11 percentage points less likely to be employed, have on average 0.1 more children, and are more likely to live with a household member requiring permanent care. This selection pattern is substantively interpretable: households with heavier caregiving burdens and more flexible schedules are more likely to engage with a chatbot about maternal employment. It also has an implication for external validity that cuts in our favour: because treatment effects are concentrated among labor-market-attached women (Section 6.1), and engagers are disproportionately inactive, the within-engager treatment effect is identified on a population that is harder to move. Effects found among this group therefore likely understate what an equivalent intervention would achieve in a less selected sample. The variable that matters for intervention validity is the second-order belief about men's community support for maternal employment, the specific statistic the treatment corrects. Engagers and non-engagers hold virtually identical priors on this dimension (58.1 vs. 58.6, a difference of 0.5 percentage points that is not statistically distinguishable from zero), confirming that the disclosed norm is accurate for the engager subpopulation's reference group. Any information experiment based on sharing social-norm statistics depends critically on the accuracy of those statistics for the relevant population (Bicchieri, 2016; Bursztyn et al., 2023); the near-identical second-order belief profiles across engagement groups directly support this accuracy condition.

Third, we directly test the validity of the placebo comparison by examining whether treatment and control engagers differ in baseline beliefs. A potential concern is that individuals who chose to click through a message about gender norms may differ from those who clicked through a message about green-transport subsidies, undermining

the placebo as an attention control. Appendix Table D.5 compares the 620 treatment engagers and 616 control engagers on the full set of baseline covariates. No belief variable differs significantly between the two groups; point differences are at most 2.5 percentage points on any single belief measure. The combination of near-identical engagement rates, negligible between-arm differences in baseline beliefs, and the pre-randomization nature of the baseline beliefs together support the interpretation of the placebo comparison as a valid control for message-receipt and chatbot interaction.overthink

Reference-group norm accuracy. A related concern is whether the Bogotá average norm disclosed to respondents accurately represents the norm within each person's relevant reference group. If individuals weight the views of socioeconomically similar peers more heavily than the city-wide average, a population-level statistic could over- or understate the corrective signal for particular subgroups. We test this by examining how baseline second-order beliefs vary across observable strata.

Second-order beliefs about men's support for maternal employment are broadly homogeneous across socioeconomic and demographic subgroups. Across socioeconomic strata, the mean of second-order beliefs ranges from 57.6 (low SES, stratum 1–2) to 61.7 (high SES, stratum 4–6), a spread of 4.1 percentage points. Across education groups, the range is slightly wider: from 55.4 (primary education) to 60.5 (tertiary), a spread of 5.1 percentage points. Among women, the range across employment categories is only 2.8 percentage points. Crucially, within each subgroup, engagers' baseline beliefs are statistically indistinguishable from those of non-engagers (all $p > 0.15$), confirming that chatbot participants are representative of their respective reference groups.

These estimates bound the reference-group accuracy of the disclosed norm. The maximum deviation of any subgroup's mean second-order beliefs from the city-wide average (the mean being equal to 58.4$) is 3.3 percentage points (high-SES respondents). By comparison, the misperception being corrected is approximately 28 percentage points (actual first-order support: 89%; perceived support among men: 61%). The reference-group mismatch is therefore less than 20% of the corrected misperception, and its sign is conservative: high-SES respondents are told a norm that slightly *understates* their reference group's true support, so the intervention if anything underestimates the potential corrective scope for that group. Appendix Figure D.7 displays these patterns. Together, these results support the interpretation that the Bogotá average is a valid and sufficiently accurate proxy for the relevant norm across the heterogeneous households in our sample.

5 Main results

We organize the results according to the theory of change described in Section 2. If credible information on the prevailing norms updates people's perceptions of societal and spousal support for maternal employment, these belief changes may, in turn, affect intra-household allocation decisions and, potentially, short-term labor-market behaviors and aspirations. The analysis therefore proceeds in three steps.

First, we examine treatment effects on beliefs, distinguishing between first-order attitudes and second-order beliefs about community norms and spousal views (Section 5.1). Second, we assess whether these belief updates translate into changes in a concrete intra-household decision involving a career-enhancing opportunity (Section 5.2). Third, we test whether the intervention affects short-run labor-market behaviors and stated preferences measured at endline (Section 5.3).

Throughout, our main estimates rely on IPWRA to account for selective engagement with the WhatsApp module and survey attrition; OLS and IPW-weighted OLS analogues are reported in Appendix Section 7.1.

5.1 Changes in beliefs

First-order beliefs and community second-order beliefs, Y_b. We first test whether exposure to accurate information about the share of adults in Bogotá who support mothers working outside home updates respondents' own attitudes (first-order beliefs) and their perceptions of societal support (community second-order beliefs). These outcomes were measured at the beginning of the endline phone survey and before the endline reinforcement. This implies that treated individuals were exposed to the information only once, via the WhatsApp chatbot at midline. The sample includes respondents observed at both midline and endline (Sample $2 \cap 3$ ($N = 1{,}102$; 453 men and 649 women).

Table 5 reports IPWRA estimates. Panel A shows the results for the full sample, while Panel B reports gender-specific ATTs. Columns (1)–(3) present treatment effects on first-order beliefs and community second-order beliefs. Columns (4)–(5) use binary misperception indicators equal to one if the respondent underestimates true support among men (column 4) or among women (column 5) by at least five percentage points.

We find no evidence that the intervention shifts first-order beliefs about whether mothers with young children should be allowed to work outside the home. The ATT for the pooled sample is essentially zero and statistically insignificant (column 1), and the gender-specific estimates are similarly small and imprecisely estimated. This is consistent

with a ceiling effect: baseline support for maternal employment is already very high, with around 90 percent of control respondents agreeing with the statement (0.90 overall, 0.87 among men and 0.93 among women).

By contrast, the intervention increases perceived societal support for maternal employment. Treated individuals revise upward their beliefs about men's support by approximately 2.7 percentage points, about 4% relative to a control mean of 63.2% (column 2, Panel A). Perceived support among women increases by 3.6 percentage points, about 5% relative to the control mean of 75.8% (column 3, Panel A). Gender-specific estimates indicate that belief updating is more pronounced among men: treated men increase perceived support among women by about 4.5 percentage points, whereas women's estimates are positive but smaller and not statistically significant.

Columns (4) and (5) summarize these changes using misperceptions indicators. In the pooled sample, the probability of underestimating men's support declines by 5.4 percentage points, implying a 6.5% reduction relative to the control mean of 83% (column 4, Panel A).These averages mask substantial gender heterogeneity. Among men, the probability of underestimating other men's support falls by 11 percentage points (column 4, Panel B), while the corresponding estimate for women is small and statistically insignificant. For misperceptions about women's support (column 5), point estimates are negative but imprecise for either gender.

In sum, Table 5 shows that a one-time exposure to accurate social information about prevailing norms leaves respondents' own attitudes unchanged, but it does improve the accuracy of beliefs about community support. The clearest adjustments concern perceptions of men's support for maternal employment, where baseline misperceptions were largest. Tables E.1 and E.2 report OLS and weighted-OLS benchmarks with similar magnitudes.

Second-order beliefs about spouse, Y_s. We next examine whether the intervention affected beliefs about one's partner's attitudes, i.e., within-household belief updating. Respondents were not informed about their spouse's true baseline belief. Any change in spousal perceptions must therefore arise indirectly, either through updated perceptions of broader community norms, or because the intervention increases the salience of these topics and triggers within-couple discussion.

The timing of these outcomes differs from those in Table 5. Spousal second-order beliefs were elicited after the endline reinforcement, meaning that respondents in Sample $2 \cap 3$ had been exposed to the information twice — once through the midline WhatsApp chatbot and once again via the scripted endline reminder — before reporting their

Table 5. Impact of Information Treatment on Beliefs and Misperceptions about Maternal Employment (IPWRA Estimates, Sample 2 ∩ 3)

	First-Order Belief	Second-Order Beliefs about		Misperceptions (D) about	
		Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)
Panel A. ATT for All Treated (Sample 2 ∩ 3)					
Treatment (ATT All)	0.001	2.747*	3.582**	–0.054*	–0.047
{Analytical SE}	{0.020}	{1.613}	{1.485}	{0.029}	{0.035}
p-value	(0.951)	(0.089)	(0.016)	(0.061)	(0.179)
Mean Dep. Var (Controls)	0.902	63.194	75.791	0.832	0.553
N (All)	1102	1102	1102	1102	1102
Panel B. ATT by Gender					
Male (ATT)	–0.004	2.818	4.497*	–0.110**	–0.047
{Analytical SE}	{0.040}	{2.576}	{2.370}	{0.051}	{0.055}
p-value	(0.921)	(0.274)	(0.058)	(0.031)	(0.393)
Mean Dep. Var (Controls)	0.874	65.618	75.291	0.830	0.566
N (Men)	453	453	453	453	453
Female (ATT)	–0.002	2.572	2.296	–0.018	–0.040
{Analytical SE}	{0.020}	{1.951}	{1.893}	{0.034}	{0.043}
p-value	(0.939)	(0.187)	(0.225)	(0.595)	(0.349)
Mean Dep. Var (Controls)	0.930	61.426	76.754	0.840	0.536
N (Women)	649	649	649	649	649

Notes: Analytical standard errors in braces {}. Stars denote significance based on p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Outcomes: (1) First-order belief that mothers with children under 6 should be allowed to work outside the home; (2)–(3) Second-order beliefs about support among men and women in Bogotá; (4)–(5) indicators for underestimating true support among men and women by at least 5 percentage points. All IPWRA models use stabilized inverse-probability weights for selection into Sample 2 ∩ 3 (WhatsApp and Endline) and for treatment assignment, and control for baseline covariates (age, income, education, education, employment status, baseline beliefs, and stratification blocks). Corresponding OLS and OLS-with-weights estimates are reported in Appendix Tables E.1 and E.2. For additional transparency, Romano-Wolf step-down p-values for community belief outcomes (Family F1) are reported in Appendix Table H.4; these supplement rather than replace the Fisher exact p-values, which remain our primary inference framework. Lee (2009) sharp bounds are reported in Appendix Table H.5 (Panel C); all bounds include zero, consistent with the null findings.

perceptions of their partner. A limitation of this design is that the estimates in Table 6 cannot be decomposed into a WhatsApp-only effect and a reinforcement effect: they capture the cumulative impact of both exposures. Table 5 offers a partial benchmark: community beliefs, measured at the start of the endline call and therefore before the reinforcement, show effects of 3–5 percentage points under a single WhatsApp exposure. The spousal belief effects reported here — approximately 6 percentage points on perceived partner support for maternal employment — are larger, consistent with a second exposure amplifying the initial update and/or with within-couple discussion occurring in the weeks between midline and endline. This framing is consistent with the design

intent: the reinforcement was administered precisely to address incomplete WhatsApp take-up. The implication is that the estimates in Table 6 should be interpreted as effects of the full information package rather than of WhatsApp exposure alone, and that the paper cannot quantitatively trace a clean WhatsApp → spousal belief updating channel.

Table 6 reports IPW estimates for two statements: support for maternal employment (columns 1–3) and support for equal sharing of household tasks (columns 4–6). For each statement, we consider three outcomes: (i) own attitudes (first-order beliefs), (ii) perceived spousal support (second-order beliefs about the partner), and (iii) a binary misperception indicator equal to one if the respondent is wrong about the partner's beliefs.

Own attitudes towards both working mothers and equal task sharing are already very favorable and remain essentially unchanged. For maternal employment, the ATT on first-order beliefs is close to zero and imprecisely estimated for both men and women. For equal task sharing, the pooled effect is about 2.1 percentage points and is driven by women (2.5 percentage points), from a control mean of 97.5 %. Given the high baseline approval rates, we view these changes as tightening around a ceiling rather than a meaningful shift in attitudes.

By contrast, the intervention increases perceived spousal support along several dimensions. For maternal employment, treated respondents revise upward their perception of their partner's approval by 6.3 percentage points in the pooled sample, relative to a control mean of 88.5% (column 2, Panel A). The magnitude is similar across gender: treated men and women increase perceived partner support by approximately 5.9 and 6.3 percentage points, respectively (column 2, Panels B and C), with both estimates statistically significant. Misperception indicators (column 3) move in the expected direction, though estimates are less precise.

For equal task sharing, perceived spousal support increases by 3.8 percentage points in the pooled sample (column 5, Panel A), from a control mean of 89.9%. This effect is driven by women, who revise upward their perception of their husband's support by about 6.1 percentage points (column 5, Panel C). In contrast, the corresponding estimate for men is small and statistically indistinguishable from zero (column 5, Panel B). Misperception indicators (column 6) again point in the expected direction but are not precisely estimated.

Overall, Table 6 shows that although the intervention targets community-level norms, reinforced exposure leads respondents to revise beliefs about their partner's attitudes. The clearest and most precisely estimated effects are on perceived spousal support for maternal employment for both genders, and on perceived support for equal task sharing among women. Together with Table 5, the pattern is consistent with a belief-updating channel in which corrected social information prompts individuals to reassess

expectations about close others, particularly where baseline underestimation is more common.

Table 6. Impact of Information Treatment on Beliefs and Misperceptions about Spouse's Attitudes (IPW Estimates, Sample 2 ∩ 3)

	Working Mothers			Equal Task Sharing		
	First-order (1)	2nd-order (Spouse) (2)	Misperception (D) (3)	First-order (4)	2nd-order (Spouse) (5)	Misperception (D) (6)
Panel A. ATT for All Treated (Sample						
Treatment (ATT All)	.009	.063***	−.028	.021**	.038**	−.027
{Analytical SE}	{.019}	{.018}	{.034}	{.011}	{.017}	{.023}
p-value	(.628)	(.001)	(.413)	(.048)	(.023)	(.248)
Mean Dep. Var (Controls)	.900	.885	.184	.965	.899	.104
N (All)	1102	1102	746	1102	1102	746
Panel B. ATT by Gender: Men						
Male (ATT)	.010	.059**	−.046	.013	−.004	−.002
{Analytical SE}	{.037}	{.026}	{.048}	{.021}	{.017}	{.021}
p-value	(.783)	(.021)	(.336)	(.542)	(.822)	(.932)
Mean Dep. Var (Controls)	.875	.909	.155	.956	.982	.028
N (Men)	453	453	341	453	453	341
Panel C. ATT by Gender: Women						
Female (ATT)	.004	.063**	−.004	.025***	.061**	−.040
{Analytical SE}	{.021}	{.026}	{.046}	{.009}	{.025}	{.041}
p-value	(.836)	(.015)	(.939)	(.006)	(.016)	(.325)
Mean Dep. Var (Controls)	.925	.868	.205	.975	.841	.164
N (Women)	649	649	405	649	649	405

Notes: Analytical standard errors in braces {}. Stars denote significance based on p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Outcomes for "Working Mothers": (1) first-order belief about the spouse's approval of mothers with young children working outside the home; (2) second-order belief about the spouse's approval; (3) misperception indicator equal to one if the respondent is mistaken about the spouse's true support. Outcomes for "Equal Task Sharing": (4) first-order belief about the spouse's approval of equal sharing of household tasks; (5) second-order belief about the spouse's approval; (6) misperception indicator defined analogously. All estimates use inverse-probability weighting for selection into Sample 2 ∩ 3 (WhatsApp and Endline) and for treatment assignment, and control for baseline covariates and stratification blocks, as in the main specification. Corresponding OLS and weighted-OLS estimates are reported in Appendix Tables E.3 and E.4. For additional transparency, Romano-Wolf step-down p-values for community and spousal belief outcomes (Families F1–F2) are reported in Appendix Table H.4; these supplement rather than replace the Fisher exact p-values, which remain our primary inference framework. Lee (2009) sharp bounds are not reported for spousal belief outcomes because the estimation sample (Sample 2 ∩ 3, requiring completion of both the WhatsApp module and the endline survey) has a mechanical treatment imbalance by design: treated households are more likely to have completed the WhatsApp module and therefore appear in Sample 2 ∩ 3 at a higher rate, which would violate the monotonicity assumption underlying Lee bounds. Attrition robustness for this table is instead supported by the negligible treatment effect on endline completion alone (-0.011, $p > 0.5$; Appendix Table D.3).

5.2 Changes in Intra-Household Preferences for Career-enhancing Course Enrollment

We next assess whether the information treatment affected a concrete within-household allocation decision: whether the single career-development course slot was assigned to the wife rather than the husband. This outcome captures a short-run choice with potential implications for future labor-market trajectories. The decision was elicited during the WhatsApp module, immediately after treatment exposure at midline. The estimation sample therefore includes all respondents who completed the chatbot module (Sample 2), regardless of subsequent endline completion.

Table 7 reports IPWRA estimates and follows the same structure as previous tables. Columns (1)–(3) correspond to: (i) prioritizing the wife for the course, (ii) the respondent's own stated interest in the course, and (iii) the respondent's belief about their partner's interest.

Pooling men and women, we find no average treatment effect on prioritizing the wife for the course (column 1). The pooled estimate, however, masks heterogeneity by gender. Among men, the treatment increases the probability of prioritizing the wife by 9.1 percentage points (Panel B), relative to a control mean of 40.2%—an increase of roughly 23 percent. Under our preferred IPWRA specification the conventional clustered p-value is 0.104, which lies above the 10 percent threshold; we therefore interpret this as suggestive rather than conclusive evidence.[10]

Among women, the estimate in Panel C is close to zero. In the control group, women already prioritize the wife for the course in 84.1 percent of cases, leaving little scope for further increases.

Columns (2) and (3) show no systematic treatment effects respondents' own interest in the course or on their beliefs about their partner's interest. Nevertheless, the control means reveal a striking within-couple pattern. Both men and women report high own interest in the course (74.3% for men and 81.9% for women), yet substantially underestimate their partner's interest (57.4% for men's beliefs about their wife; 37.4% for women's beliefs about their husband). Thus, while the intervention appears to shift men's allocation choices in favor of their wives, it does not measurably narrow the broader gap between

[10]Under our preferred IPWRA specification the clustered p-value is 0.104. A Fisher exact p-value of 0.011 is obtained from a randomization test applied to the OLS specification (Appendix Table E.6), and OLS point estimates are very close to the IPWRA estimate. As shown in Section 5.4, however, the 9-percentage-point estimate is remarkably stable in sign and magnitude across all robustness checks: Lee bounds are strictly positive under a trimming fraction of only 0.8%, and point estimates range narrowly between 0.091 and 0.094 across all alternative weight specifications, crossing conventional significance thresholds under winsorised weights.

individuals' own aspirations and their beliefs about their partner's aspirations.

Overall, the results in Table 7 provide suggestive evidence that correcting misperceived norms affects intra-household allocation decisions along a margin where men retain discretion. Consistent with earlier belief-updating results, the intervention appears to relax men's perceived constraints on supporting their wives' career investments, while leaving women's already strongly expressed preferences largely unchanged.

5.3 Labor Market Actions and Attitudes

We conclude the main analysis by examining whether the information intervention translated into short-run labor-market behaviors and stated preferences. Table 8 reports treatment effects on three sets of endline outcomes measured one and two months after the WhatsApp intervention: (i) job mobility (whether the respondent changed jobs or started a business), (ii) aspirations to improve one's labor-market situation, and (iii) stated preferences for balancing work and family life.

Because labor-market behaviors may plausibly respond only to information received before the relevant decision window, we report across samples that vary in exposure timing and intensity. Columns (1), (4), and (6) restrict the sample to respondents observed both at midline (WhatsApp module) and endline (Sample $2 \cap 3$), who therefore had time between midline and endline to adjust behaviors. Columns (2), (5), and (7) use broader endline-based samples. Column (2) includes endline respondents in households where at least one spouse responded to the midline WhatsApp survey, capturing both *direct* and *indirect* exposure through a treated partner. Columns (5) and (7) include all respondents from the endline sample (Sample 3), irrespective of midline participation. Finally, Column (3) uses the endline-only sample (Sample $3 \setminus 2$) as a timing placebo: respondents in this group receive the information only during the endline survey, so the treatment cannot have affected earlier decisions to start changing jobs or to start a business, and this specification primarily helps us assess potential social-desirability responses induced by the treatment.

Job search activities The clearest effects are for women's job mobility. In column (1), in Sample $2 \cap 3$, treated women are 9.6 percentage points more likely to report having started to look for a job, to change job or to start a business than women in the control group, accounting for a 13% increase. Including in the sample, women treated at midline indirectly through the participation of their husband in the WhatsApp chatbot (column (2)), we find that the effect remains for women with a magnitude of 7.5 percentage points.

Table 7. Changes in interest about wife's LFP – IPW estimates

Dependent Variable	Wife Should Attend Course	Are You Interested?	Is your Partner Interested?
Panel A. ATT for All			
Treatment (ATT)	.023	–.014	–.022
Std. Err.	(.031)	(.030)	(.035)
P-value	[.460]	[.648]	[.532]
Mean Dep. Var (Controls)	.688	.793	.460
N (All)	1017	984	973
Panel B. ATT by Gender: Men			
Treatment (ATT)	.091	–.046	.018
Std. Err.	(.056)	(.052)	(.056)
P-value	[.104]	[.376]	[.746]
Mean Dep. Var (Controls)	.402	.743	.574
N (Men)	373	360	356
Panel C. ATT by Gender: Women			
Treatment (ATT)	–.006	.007	–.019
Std. Err.	(.032)	(.035)	(.041)
P-value	[.842]	[.850]	[.637]
Mean Dep. Var (Controls)	.841	.819	.374
N (Women)	644	624	617

Notes: Regressions weighted by inverse probability of being in the WhatsApp sample (selection IPW). Coefficients with clustered standard errors in parentheses ( ). Clustered p-values in brackets [ ]. Stars denote significance based on clustered p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Outcomes: (1) Equals 1 if the spouse chooses to give priority to the wife for the course; (2) equals 1 if the spouse declares being interested in the course; (3) equals 1 if the spouse believes the other spouse to be interested in the course. All estimates use inverse-probability weighting for selection into Sample 2 (WhatsApp) and for treatment assignment, and control for baseline covariates and stratification blocks, as in the main specification. Corresponding OLS and weighted-OLS estimates are reported in Appendix Table E.6. For additional transparency, Romano-Wolf step-down p-values for course-allocation outcomes (Family F3) are reported in Appendix Table H.4. Lee (2009) sharp bounds on the treatment effect, which are robust to differential attrition under the monotonicity assumption, are reported in Appendix Table H.5; trimming fractions are below 3% and both bounds for men's wife-prioritized outcome are positive. Sensitivity of the IPWRA estimate to attrition weight specification (winsorised, trimmed, logit-PS) is reported in Appendix Table H.6, Panel A; the estimate is stable across all specifications.

The placebo outcome in column (3) shows no statistically significant effect for women (Panel C); if anything, the sign is reversed. This suggests that the previous results are not driven by social-desirability bias induced by the treatment. In contrast, and as expected given that the treatment is unlikely to affect men's job mobility, estimated effects for men are close to zero and not statistically significant across all specifications (Panel B).

Labor-market aspirations Concerning aspirations to improve labor market conditions (columns (4) and (5)), estimated treatment effects are small and statistically indistinguishable from zero for both men and women. Across samples, we cannot reject the null of no effect, suggesting that the intervention did not materially shift broad labor-market aspirations in the short run.

Work-family balance preferences The work–family balance outcome exhibits a different pattern. Among men, we find increases in the stated preference for balancing work and family life. In Sample $2 \cap 3$, the estimated effect for men is 11 percentage points, about 35% relative to a control mean of 31.7% (column 6). In the full endline sample, the effect remains positive at 6.6 percentage points (column 7). For women, treatment effects are close to zero and imprecisely estimated. We note an interpretive ambiguity in the men's result: a stated preference for "balancing work and family" could reflect couple-oriented support for the wife's labor-market engagement (consistent with the treatment's intent), or alternatively a desire by men to personally work less and invest more in family time. The two readings have different welfare implications. The heterogeneity result in Table F.2—where work–family balance effects concentrate among households with inactive wives—may be more consistent with the latter interpretation, since information about others' progressive norms could prompt men to re-evaluate their own work intensity when the wife's employment is not the immediate margin. We interpret the result as an attitudinal shift toward greater valuation of family time, while acknowledging that the direction of this shift with respect to women's employment remains unclear. Unlike job mobility, where the timing placebo (Sample $3 \setminus 2$, column 3) argues against social-desirability responses, work–family balance is a stated preference and some demand effect from the progressive-norms message cannot be fully ruled out; this caveat reinforces the conservative interpretation of the result.

These results highlight the value of the couple-based design for distinguishing gender-specific responses and tracing distinct behavioral margins. The most pronounced behavioral response appears among women, who become more likely to report a job change or business start following exposure to norm-correcting information. By contrast, the

strongest attitudinal response appears among men, who report a greater preference for balancing work and family life. While these effects are measured over a short horizon (1 to 2 months), they arise from a light-touch informational intervention and are consistent with the view that updated beliefs about social support can prompt reassessment of household roles and labor-market choices, particularly where participation is already feasible.

Table 8. Impact of Information Treatment on Job Search and Labor-Market Aspirations (IPWRA Estimates)

	Job Mobility			Aspires Better LM Situation		Wants Work–Family Balance	
	Sample 2 ∩ 3	Direct or Indirect T	Sample 3\2	Sample 2 ∩ 3	Sample 3	Sample 2 ∩ 3	Sample 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. ATT for All Treated							
ATT	0.058*	0.043	–0.056	0.005	0.000	0.052	0.038
{Analytical SE}	{0.031}	{0.029}	{0.048}	{0.036}	{0.027}	{0.033}	{0.026}
p-value	(0.064)	(0.138)	(0.243)	(0.900)	(0.995)	(0.116)	(0.144)
Mean Dep. Var (Controls)	0.700	0.695	0.733	0.508	0.502	0.334	0.345
N (All)	1102	1233	445	1102	1382	1102	1382
Panel B. ATT by Gender: Men							
ATT	0.012	0.010	–0.048	–0.042	–0.027	0.110*	0.066*
{Analytical SE}	{0.054}	{0.044}	{0.076}	{0.060}	{0.042}	{0.057}	{0.040}
p-value	(0.817)	(0.822)	(0.533)	(0.481)	(0.524)	(0.054)	(0.099)
Mean Dep. Var (Controls)	0.664	0.666	0.677	0.492	0.504	0.317	0.320
N (Men)	453	549	229	453	603	453	603
Panel C. ATT by Gender: Women							
ATT	0.096***	0.075**	–0.087	0.054	0.018	–0.006	0.015
{Analytical SE}	{0.035}	{0.034}	{0.056}	{0.044}	{0.037}	{0.041}	{0.036}
p-value	(0.006)	(0.028)	(0.119)	(0.223)	(0.626)	(0.891)	(0.673)
Mean Dep. Var (Controls)	0.725	0.725	0.792	0.507	0.502	0.361	0.366
N (Women)	649	684	216	649	779	649	779

Notes: Analytical standard errors in braces {}. Stars denote significance based on p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Columns (1)–(3) refer to indicators of having started to change jobs since baseline: main specification using Sample 2 ∩ 3, an alternative coding based on direct/indirect exposure at endline (Sample 3), and a placebo baseline outcome, respectively. Columns (4)–(5) capture whether the respondent aspires to a better labor-market situation, and columns (6)–(7) whether they prioritize balancing work and family, each measured for Sample 2 ∩ 3 and for the endline (Sample 3), respectively. All estimates are obtained using inverse-probability weighting for selection into the relevant analysis sample and for treatment assignment, and control for baseline covariates and stratification blocks, as in the main belief specifications. Sample 2 ∩ 3 comprises respondents observed in both the *WhatsApp* module and the follow-up phone survey (1,102 individuals: 453 men and 649 women). Corresponding OLS and weighted-OLS estimates are reported in Appendix Tables E.7 and E.8. For additional transparency, Romano-Wolf step-down p-values for labor-market outcomes (Family F4) are reported in Appendix Table H.4; these supplement rather than replace the Fisher exact p-values, which remain our primary inference framework. Lee (2009) sharp bounds are reported in Appendix Table H.5; both bounds are positive for women's job mobility (trim = 2.5%), and the upper bound for men's work-family balance is significantly positive despite a 10.2% trimming fraction. Sensitivity of the IPWRA estimates to attrition weight specification is reported in Appendix Table H.6, Panels B–C: women's job-search result is stable across all weight variants; men's work-family balance is robust to winsorising and a logit-PS alternative but sensitive to dropping the 1% highest-leverage observations.

5.4 Robustness

The main results are robust to a range of specification checks, detailed in the Appendix and summarized here. For each check we consider four headline results: (i) women's job mobility, (ii) men's work–family balance preference, (iii) men's course prioritization of their wife, and (iv) community and spousal belief updating.

Multiple testing correction. Romano-Wolf step-down p-values (Appendix Table H.4) correct for multiple testing within each outcome family. They are qualitatively consistent with the main findings: women's job mobility remains significant at the 5 percent level among women (Romano-Wolf p=0.013). Men's work–family balance preference clears the 10 percent threshold (Romano-Wolf p=0.077). HFor men's course prioritization, the Romano-Wolf p-value is 0.151, consistent with the marginal conventional p-value already noted in the main results. Community and spousal belief updates are not significant after correction, reflecting the modest magnitude of individual belief shifts relative to their standard errors.

Attrition-robust Lee bounds. Lee (2009) sharp bounds address the possibility that differential attrition between treatment and control groups biases estimated treatment effects. Under the monotonicity assumption — that treatment assignment can only increase (or only decrease) the probability of remaining in the sample — they recover the tightest possible bounds on the average treatment effect by trimming the outcome distribution of the group with higher retention. In Appendix Table H.5, both bounds are strictly positive for women's job mobility (trimming fraction 2.5%) and for men's course prioritization of their wife (trimming fraction 0.8%), providing attrition-robust support for both results even in the absence of conventional significance for the latter. The upper bound for men's work–family balance is significantly positive despite a larger trimming fraction of 10.2%. Bounds for community belief outcomes all include zero, consistent with the null findings on first-order attitudes.

Alternative attrition weight specifications. Appendix Table H.6 reports IPWRA estimates under four alternative weight specifications: the baseline probit propensity score, winsorised weights capped at the 95th percentile, a trimmed sample dropping the 1% highest-leverage observations, and a logit propensity score. Women's job mobility is stable across all four specifications, with point estimates ranging from 0.091 to 0.096 and p-values below 0.01 throughout. For men's course prioritization, point estimates range narrowly from 0.091 to 0.094 and the result crosses conventional significance thresholds

under winsorised specification, confirming that the finding is not an artifact of the weighting procedure. Men's work–family balance is robust to winsorising and the logit-PS alternative but sensitive to the trimmed specification; we interpret this as grounds for additional caution in reading that result.

6 Exploring mechanisms

We extend the main analysis in three ways. First, we examine treatment-effect heterogeneity by baseline labor-market attachment of the wife (Section 6.1). Second, we study heterogeneity by intra-household exposure patterns to the information (Section 6.2). Third, we also provide suggestive evidence on potential channels linking belief updating to behavioral responses (Section 6.3).

6.1 Heterogeneity along labor-market attachment of women

We first examine whether treatment effects differ by the wife's baseline labor-force status, focusing on whether the wife was inactive (out of the labor force) at baseline. This dimension is policy-relevant because inactivity plausibly reflects tighter constraints (e.g., childcare and weaker labor-market attachment), which may limit the scope for norm-correcting information to translate into changes in beliefs or behavior.

We estimate interacted specifications of the form

$$Y_i = \alpha + \beta D_i + \gamma Z_i + \delta(D_i \times Z_i) + X_i'\theta + \lambda_s + \varepsilon_i,$$

where D_i denotes treatment assignment, Z_i indicates whether the wife is inactive at baseline, X_i is the baseline covariate set, and λ_s are stratification fixed effects. The treatment effect when $Z_i = 0$ is β; when $Z_i = 1$ it is $\beta + \delta$. Because wife inactivity is a stratification variable, these differences reflect causal heterogeneity in treatment effects. Tables F.1 and F.2 in Appendix 7.2 report heterogeneous effects; we use interactions rather than subgroup-specific estimators to preserve power.

Table F.1 shows that belief updating concentrates among households where the wife is active at baseline. The treatment raises first-order agreement by about 3 percentage points when the wife is active, but the implied effect when she is inactive is close to zero. Heterogeneity is sharpest for men's second-order beliefs about other men's support: updating is sizable when the wife is active but significantly attenuated when she is inactive. By contrast, beliefs about women's support and spousal beliefs show smaller

and less precisely estimated differences by inactivity status.

The same pattern carries over to behavioral outcomes (Table F.2). For the midline course-allocation decision, the treatment increases husbands' likelihood of prioritizing their wife by 11–13 percentage points when she is active, but the net effect is close to zero when she is inactive. At endline, increases in job-search effort and aspirations are concentrated among active wives while the interaction terms imply these gains largely disappear, or even reverse, when the wife is inactive. The exception is stated preferences for work–family balance, where treatment effects concentrate among households with an inactive wife, consistent with information shifting priorities toward balance rather than labor-market entry when participation constraints bind.

Inactive wives face multiple constraints (e.g., childcare, skills, demand conditions), and our data do not allow us to separate whether inactivity reflects binding structural barriers, a joint household preference, or the woman's own choice. Our results shows where norm correcting information works rather than to the effect of changing employment status per se. Correcting misperceptions can shift beliefs and household priorities, but only when structural barriers are already low. On its own, information is unlikely to move women into the labor force. Pairing norm feedback with policies that relax binding constraints—such as childcare access—may be necessary to affect inactive women. This interpretation aligns with evidence that unpaid care work remains a central barrier to women's labor supply in Latin America (Frisancho et al., 2023).

6.2 Exploring variations in intra-household exposure to treatment

We next explore whether responses vary with how information is distributed within the couple. Because we did not randomize which spouse engaged with the WhatsApp module, exposure patterns are endogenous; we therefore interpret these results as descriptive. We distinguish three configurations: *direct exposure*, where the respondent receives the message; *indirect exposure*, where only the spouse receives it; and *joint exposure*, where both partners receive the message.

Appendix Table G.1 shows that belief updating depends strongly on exposure configuration. Direct exposure replicates the main results where treated respondents revise community second-order beliefs upward by 3–5 percentage points, especially beliefs about men's support, and increase perceived spousal support by 6–9 percentage points. Under indirect exposure, adjustments are weaker and gender-asymmetric. Men indirectly exposed through their wife show a decline in first-order support (4–6 percentage points) with little updating in second-order or spousal beliefs. Women indirectly exposed

through their husband update perceived support among men upward (6–8 percentage points) and report higher perceived spousal support (4–5 percentage points). Under joint exposure, belief updating is largest and most coherent: perceived support among men and women increases by 5–9 percentage points, and perceived spousal support rises by 8–9 percentage points, while first-order attitudes remain unchanged. These patterns indicate that information does not automatically diffuse within couples and that bilateral exposure is associated with more consistent belief revisions.

Table G.2 examines whether exposure patterns translate into differences in course allocation and job-search effort. In the full sample, the treatment increases the probability that the course is assigned to the wife by 7–8 percentage points, driven primarily by men prioritizing their wives, while job-search effort shows no clear average response. When only one spouse participates, effects on course prioritization are smaller and imprecise for men, and indirect exposure does not generate detectable job-search effects. Among couples where both spouses participate, treatment effects on course prioritization are largest for men (9–10 percentage points), with suggestive evidence of increased job-search effort among women. Taken together, these results suggest that ensuring both partners receive the information directly — rather than relying on within-couple diffusion — may be important for maximizing the impact of norm-correcting interventions on intra-household decisions.

6.3 Exploring channels

To assess whether belief updating is directionally consistent with the labor-market responses, we conduct an exploratory "consistency check" that relates treatment-induced changes in perceived societal support to endline outcomes (job-search effort, labor-market aspirations, and work–family balance). Using assignment as an instrument for follow-up perceptions of societal support, we find small and often imprecise associations; when non-negligible, signs align with the theory of change (greater perceived support is associated with higher job-search effort among women and stronger work–family balance preferences among men). Given the single-instrument, single-mediator structure and modest first stages, we interpret these results strictly as suggestive and report the full setup and estimates in Appendix 7.4. A further caveat concerns the magnitude of belief updating itself: the 3–5 percentage-point improvements documented in Table 5 close roughly 15–25% of the initial 20-percentage-point misperception gap, leaving substantial residual error. This partial correction means the evidence establishes directional consistency between belief updating and behavioural change, rather than a clean belief-to-behaviour causal chain; the

treatment may partly operate through correlated channels that accompany information provision, such as reduced stigma salience or shifts in within-couple communication dynamics.

7 Conclusion

This paper studies whether pluralistic ignorance about support for maternal employment persists within couples in an urban Latin American context, and whether accurate information about prevailing norms can reduce misperceptions and affect short-run household decisions and labour-related behaviours. Using a representative sample of cohabiting couples with at least one child under six in Bogotá, we document a stark wedge between private attitudes and perceived norms. Support for mothers' employment is high among both women and men, yet respondents substantially underestimate societal support—especially men's support—and often misperceive their partner's views. These patterns point to salient informational frictions around gender norms even in a relatively "progressive" urban setting.

Our WhatsApp-based intervention reduces these frictions. In follow-up surveys conducted one to two months after the intervention, providing respondents with feedback on local attitudes shifts second-order beliefs upward and narrows spousal misperceptions, without changing respondents' own views, consistent with a primarily corrective (rather than persuasive) mechanism. The belief updates are largest for perceptions of fathers' support and, to a lesser extent, beliefs about one's spouse.

Do corrected beliefs matter for behaviour? Our evidence is consistent with a belief-updating channel, though—as discussed in Section 6.3—we cannot formally establish that beliefs are the operative mechanism; the treatment may also work through alternative channels, such as enhanced intrahousehold discussion. In a real intra-household allocation decision, looking a single slot per couple in a career-development course, there is suggestive evidence that treated men are more likely to prioritise their wife over themselves (+9.1 pp; IPWRA p = 0.10), and this effect is concentrated among couples in which the wife is active at baseline (employed or searching for work); when the wife is initially inactive (out of the labour force), men's allocation choices respond little. Downstream labour-related outcomes show a similar pattern: treated women who are active at baseline report more intensive job search and stronger labour-market aspirations shortly after the intervention, whereas initially inactive women show little change in job-search behaviour. Across outcomes, a consistent pattern emerges: the information treatment reliably updates beliefs, and associated behavioural responses—where present—are concentrated

when constraints are not fully binding, i.e., among "near-marginal" women and through men's intra-household prioritisation.

Our couple-based design also reveals that effects depend on how information reaches the household. Bilateral exposure, where both partners receive the message, generates the most consistent belief updating, while indirect exposure through a spouse produces weaker and sometimes asymmetric responses. This suggests that interventions targeting couples may benefit from ensuring both partners receive information directly.

Several limitations deserve mention. Our outcomes are measured one to two months after treatment—sufficient to detect belief updating, allocation decisions, and early behavioral responses, but too short to assess sustained employment effects. The labour-related behavioural outcomes we report—job-search activity, labour-market aspirations, and work–family balance preferences—capture behavioural intent and early job mobility rather than transitions into stable employment; these should not be interpreted as evidence of changes in employment rates. That said, the time horizon compares favorably with related experiments: **?** measure job applications and interviews several months after treatment in Saudi Arabia, while Cameron et al. (2026) rely on immediate online choices. The zero-sum course design (one slot per household) provides a particularly conservative revealed-preference measure of intra-household prioritization: men face a direct personal cost of nominating their wife, with no possibility of compromise. In practice, many intra-household decisions — childcare arrangements, flexible working, career interruptions — involve partial trade-offs rather than zero-sum choices, suggesting that our estimates likely represent a lower bound on men's willingness to support their wives' careers in less constrained settings. Engagement with the WhatsApp module was low, and we rely on inverse-probability weighting to address selective take-up and attrition.

These findings align with evidence that correcting men's misperceptions about others' support can shift intentions and household dynamics (Bernhardt et al., 2018; Bursztyn et al., 2020), and with the broader view that social norms are one component of the barriers to women's work (Jayachandran, 2021). From a policy perspective, calibrated norm feedback is best viewed as a complement, and not a substitute, to interventions that relax structural barriers, and may be most effective when targeted to households where women are already close to the margin of work. An important direction for future research is to test bundled programmes that combine norm-correcting information with childcare or job-access components, and to track couples over longer horizons to assess sustained employment and bargaining effects.

References

Ali, Joseph, Madhuram Nagarajan, Erisa S. Mwaka, Elizeus Rutebemberwa, Andres I. Vecino-Ortiz, Angelica Tórres Quintero, Mariana Rodriguez-Patarroyo, Vidhi Maniar, Gulam Muhammed Al Kibria, Alain B. Labrique, George W. Pariyo, and Dustin G. Gibson, "Remote Consent Approaches for Mobile Phone Surveys of Non-Communicable Disease Risk Factors in Colombia and Uganda: A Randomized Study," *PLOS ONE*, 2022, *17* (12).

Anderson, Siwan, "Gender Biased Norms," CEPR Discussion Paper 20211, CEPR 2025.

Behaghel, Luc, Bruno Crépon, Marc Gurgand, and Thomas Le Barbanchon, "Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models," *Review of Economics and Statistics*, 2015, *97* (5), 1070–1080.

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*, 363–367.

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

Bicchieri, Cristina, *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge: Cambridge University Press, 2006.

_ , *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*, New York: Oxford University Press, 2016.

Bohren, J Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope, "Inaccurate statistical discrimination: An identification problem," *Review of Economics and Statistics*, 2023, pp. 1–45.

Boltz, Marie, Ana Maria Diaz, and Luz Magdalena Salas, "Hidden Biases: Pluralistic Ignorance in Perceptions of Women's Work Among Couples in Colombia," *Revue d'Economie du Développement*, 2025.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived social norms: Women working outside the home in Saudi Arabia," *American economic review*, 2020, *110* (10), 2997–3029.

_ , **Alexander W Cappelen, Bertil Tungodden, Alessandra Voena, and David H Yanagizawa-Drott**, "How are gender norms perceived?," Technical Report, National Bureau of Economic Research 2023.

Cameron, Lisa, Diana Contreras Suarez, and Diahhadi Setyonaluri, "Leveraging Women's Views to Influence Gendered Attitudes to Women Working: Evidence from an Online Intervention in Indonesia," *Journal of Development Economics*, 2026, p. 103722.

Cassidy, Rachel, Marije Groot Bruinderink, Wendy Janssens, and Karlijn Morsink, "The Power to Protect: Household Bargaining and Female Condom Use," *Journal of Development Economics*, 2021, *153*, 102745.

Cortés, Patricia and Jessica Pan, "Children and the Remaining Gender Gaps in the Labor Market," *Journal of Economic Literature*, 2023, *61* (4), 1359–1409.

_ , **Gizem Koşar, Jessica Pan, and Basit Zafar**, "Should Mothers Work? How Perceptions of the Social Norm Affect Individual Attitudes toward Work in the US," *Review of Economics and Statistics*, 2024, pp. 1–28.

Cudeville, Elisabeth and Magali Recoules, "Household Behaviour and Social Norms: A Conjugal Contract Model with Conformism.," *Annals of Economics and Statistics*, 2014, *Special volume on the Economics of Gender*, à paraître.

Dabalen, Andrew, Alvin Etang, Johannes Hoogeveen, Elvis Mushi, Youdi Schipper, and Johannes von Engelhardt, *Mobile Phone Panel Surveys in Developing Countries: A Practical Guide for Microdata Collection*, Washington, DC: World Bank, 2016.

Fisher, Ronald A., *The Design of Experiments*, Edinburgh: Oliver and Boyd, 1935.

Frisancho, Verónica, Monserrat Bustelo, and Mariana Viollaz, "What Is the Labor Market Like for Women in Latin America and the Caribbean?," Technical Report, Inter-American Development Bank 2023.

Goldin, Claudia, "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

_ , *Career and family: Women's century-long journey toward equity*, Princeton University Press, 2021.

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge: Cambridge University Press, 2015.

Jayachandran, Seema, "Social Norms as a Barrier to Women's Employment in Developing Countries," *IMF Economic Review*, 2021, *69* (3), 576–595.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard, "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, October 2019, *11* (4), 181–209.

Krupka, Erin L. and Roberto A. Weber, "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?," *Journal of the European Economic Association*, 2013, *11* (3), 495–524.

Laszlo, Sonia, Carolina De Miranda, Franque Grimard, Tania Insfrán, Thomas Kokossou, Laura Morínigo, and Jim Warnick, "Norm-Shifting Interventions and Beliefs about Gender Norms in Paraguay," *Economic Development and Cultural Change*, 2025, *73* (4), 000–000.

Majlesi, Kaveh, "Labor Market Opportunities and Women's Decision-Making Power within Households," *Journal of Development Economics*, 2016, *119*, 34–47.

Molina Millán, Teresa and Karen Macours, "Attrition in Randomized Controlled Trials: Using Tracking Information to Correct Bias," *Economic Development and Cultural Change*, 2025, *73* (2), 811–834.

Olivetti, Claudia and Barbara Petrongolo, "The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries," *Journal of Economic Perspectives*, 2017, *31* (1), 205–230.

_ , Jessica Pan, and Barbara Petrongolo, "The Evolution of Gender in the Labor Market," in "Handbook of Labor Economics," Vol. 5, Elsevier, 2024, pp. 619–677.

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.

Romano, Joseph P. and Michael Wolf, "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 2005, *73* (4), 1237–1282.

World Bank, "COVID-19 High-Frequency Phone Surveys in Latin America: Technical Note on Sampling Design, Weighting, and Estimation," Technical Note, World Bank 2021.

Online Appendix

Appendix A: Chatbot Flow for Treated Participants

Message 1: Introduction
Hi, [NAME]!
Thank you for participating in our survey with Universidad Javeriana and IPSOS about household characteristics in Bogotá.
Now, we'd like to invite you to discover the results and reflect on your answers. Your opinion continues to make a difference!
You may also receive a special resource.
Are you in? It will only take 5 minutes
Options: **Yes [Continue to Message 2] / No [Continue to Reminder]**

Random Block 1: Perceptions about Women

Message 2
Question: Out of 100 women with children under 6 in Bogotá, how many do you think believe mothers should have the freedom to work outside the home?

Message 3
Your answer was: [N] out of 100

Message 4
Do you think you were correct?
1. Yes
2. No
3. Not sure

Message 5
The reality was: [X] out of 100 women really support this idea

Message 6
What do you think about this difference?
1. Interesting
2. Irrelevant
3. Disappointing

Random Block 2: Perceptions about Men

Message 7
Question: Out of 100 men with children under 6 in Bogotá, how many do you think

believe mothers should have the freedom to work outside the home?

Message 8

Your answer was: [N] out of 100

Message 9

Do you think you were correct?

1. Yes
2. No
3. Not sure

Message 10

Find out what Bogotá's fathers really think

X

out of 100 actually support this idea

Message 11

What do you think about this difference?

1. Interesting
2. Irrelevant
3. Disappointing

Final Block: Course and Household Decisions

Message 12

As a thank-you for participating, we offer you an online course to develop tools for accessing new job opportunities and improving your job search

Message 13

If there are more interested people than spots, we may assign one per household.

Who do you think should participate?

1. Me
2. My partner

Message 14

Would you be interested in taking the course?

1. Yes
2. No

Message 15

Do you think your partner would be interested in the course?

1. Yes
2. No

3. I don't know

Message 16
We value your opinion! We'd like to ask two more questions based on hypothetical scenarios.

Message 17
A woman has the chance to accept a well-paid job . To do so, she must pay for childcare, which would leave the family income unchanged
What should she do?
Accept the job
Decline the job

Message 18
Alternatively, instead of paying for childcare, her husband could work more from home to care for the child, but he would earn less, again leaving household income unchanged
What should the woman do?
Accept the job
Decline the job

Closing Message
Thank you for participating! We'll call you soon with a few follow-up questions and more information about the course for you or your partner. Stay tuned!

Chatbot Flow for Control Group Participants Message 1: Introduction
Thank you for participating in our survey with Universidad Javeriana and IPSOS about household characteristics in Bogota.
Now, we would like to invite you to discover the results and reflect on your answers. Your opinion continues to make a difference.
You may also receive a special resource.
Are you in? It will only take 5 minutes.
Options: Yes [Continue to Message 2] / No [Go to Reminder]

Reminder
We understand you might be short on time, but your participation is very valuable to us. We would love for you to see the survey results and what they reveal about households in Bogota. It is a chance to compare your responses with others and learn more.
Options: Yes, I will participate [Continue to Message 2] / No, I will not continue [Go to Exit Question]

Block 1: Social Perception on Environmental Policy

Message 2

Do you think companies should subsidize employees who use public transport or carpool to reduce commute-related emissions?

Message 3

Your answer was: [Agree / Disagree]

Participants were then randomly assigned to receive perceptions about women or men.

Message 4

Do you think women in Bogota share your opinion?

1. Yes
2. No
3. I do not know

Message 5

$[X]$ out of 100 women with children under 6 in Bogota support this idea.

Message 6

What do you think of this information?

1. Interesting
2. Irrelevant
3. Disappointing

Message 7

Do you think men in Bogota share your opinion?

1. Yes
2. No
3. I do not know

Message 8

X

out of 100 men with children under 6 in Bogota support this idea.

Message 9

What do you think of this information?

1. Interesting
2. Irrelevant
3. Disappointing

Final Block: Course and Household Decisions

Message 10
As a thank-you for participating, we offer you an online course to develop tools for accessing new job opportunities and improving your job search.

Message 11
If there are more interested people than spots, we may assign one per household.
Who do you think should participate?
1. Me
2. My partner

Message 12
Would you be interested in taking the course?
1. Yes
2. No

Message 13
Do you think your partner would be interested in the course?
1. Yes
2. No
3. I do not know

Message 14
We value your opinion. We would like to ask two more questions based on hypothetical situations.

Message 15
A woman has the opportunity to accept a well-paid job. To do so, she would need to pay for childcare, keeping household income unchanged.
What should she do?
1. Accept the job
2. Decline the job

Message 16
Alternatively, her husband could work more from home to care for the child, earning less, and keeping household income unchanged.
What should the woman do?
1. Accept the job
2. Decline the job

Closing Message
Thank you for participating. In the coming days, we will contact you with some follow-up

questions and more information about the course for you or your partner.

Appendix B: Beliefs

Table B.1. Baseline beliefs by belief type and respondent gender

Statement	**Husbands** Mean (s.e.)	**Wives** Mean (s.e.)	**Diff.** (H–W)
A. First-order beliefs (own view)			
Women should be allowed to work	96.1 (0.32)	98.6 (0.18)	–2.5***
Fathers should share caregiving equally	97.9 (0.31)	98.9 (0.22)	–1.0*
Children suffer if mother works	28.3 (1.09)	34.0 (1.14)	–5.7***
Men should have job priority in scarcity	28.2 (1.09)	23.9 (1.04)	4.3***
Problems if wife earns more than husband	22.1 (1.01)	26.4 (1.08)	–4.3***
B1. Second-order beliefs: *other men*			
Women should work	62.7 (1.17)	57.5 (1.28)	5.2***
Equal caregiving	61.9 (1.26)	49.9 (1.34)	12.0***
Higher-earning wife causes problems	57.2 (1.27)	59.8 (1.28)	–2.6**
B2. Second-order beliefs: *other men* (SDB framing)			
Women should work	62.2 (1.17)	58.0 (1.28)	4.2***
Equal caregiving	60.6 (1.26)	50.2 (1.33)	10.4***
Higher-earning wife causes problems	56.0 (1.28)	58.7 (1.29)	–2.7**
B3. Second-order beliefs: *other women*			
Women should work	80.3 (1.00)	80.5 (1.00)	–0.2
Equal caregiving	86.6 (0.88)	88.2 (0.84)	–1.6**
Higher-earning wife causes problems	55.5 (1.27)	58.6 (1.25)	–3.1***
C. Spousal second-order beliefs			
Partner – Equal caregiving	98.5 (0.24)	88.3 (0.79)	10.2***
Partner – Higher- earning wife causes problems	24.8 (1.02)	27.6 (1.07)	–2.8**

Notes: Weighted means; robust standard errors in parentheses. "SDB framing" repeats the second-order item with the prompt "if other men answered honestly" to gauge social desirability bias (SDB). Items "children suffer" and "male job priority" were not asked under SDB or spousal frames; cells are omitted. "Diff." gives the two–sided t-test for equality of male and female means. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. True population values are computed from respondents' first-order beliefs within the same sample, separately for men and women.

Appendix C: Experimental Protocol

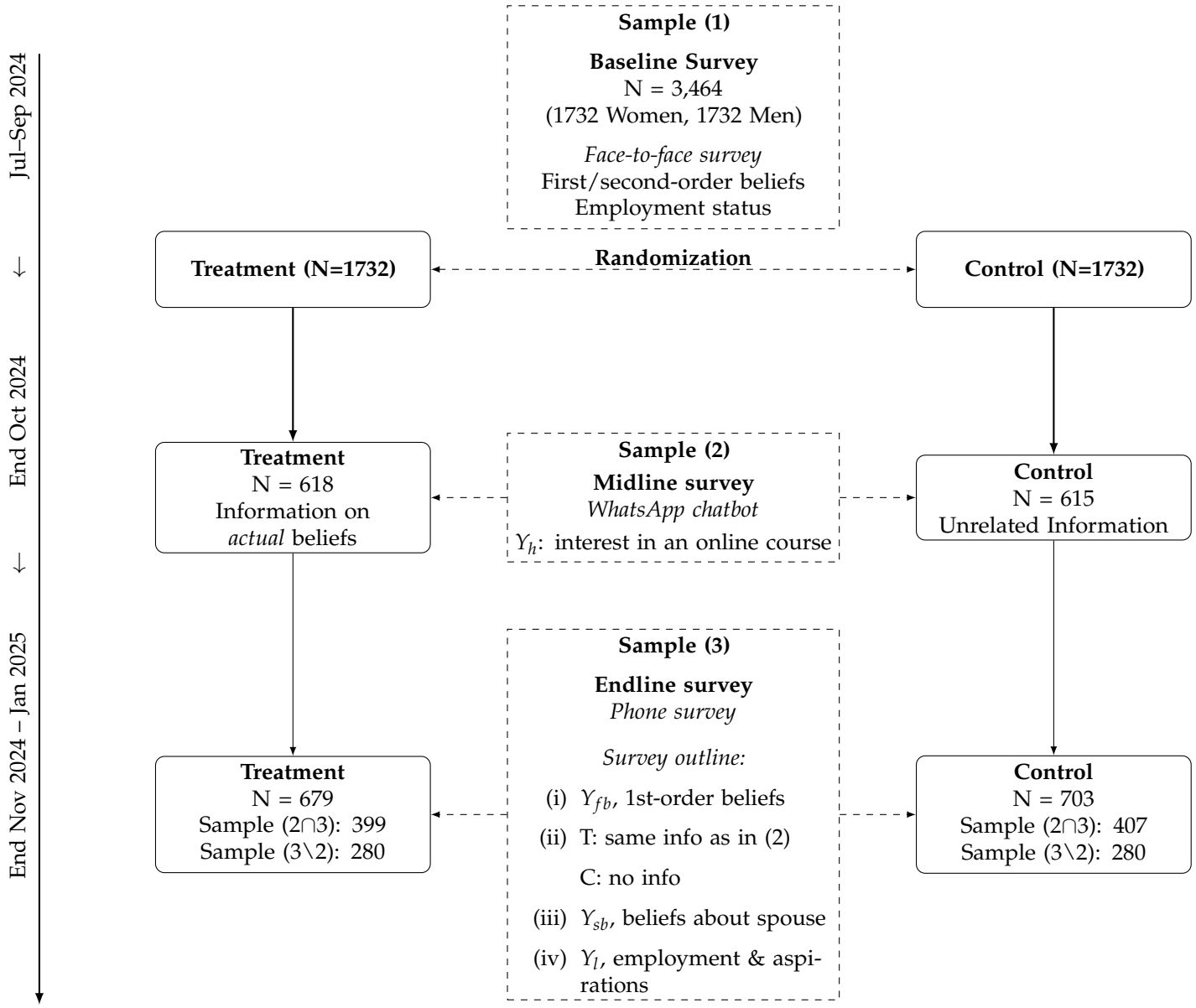


Figure C.1. Experimental design and samples

Table C.1. Mapping of Outcome Variables to Estimation Samples

Outcome Type	Outcome Variables	Estimation Sample
Beliefs about Society (Y_b)	First-order and second-order beliefs about societal support for maternal employment	Sample $2 \cap 3$: Respondents who engaged with the WhatsApp chatbot and completed the endline phone survey.
Beliefs about Spouse (Y_s)	Second-order beliefs about partner views and misperceptions about partner support for maternal employment and for intra-household equal task sharing	Sample 3 (full endline) and Sample $2 \cap 3$; measured after reinforcement.
Course Enrollment Preferences (Y_h)	Wife/self prioritization, own interest in attending, and perceived partner interest	Sample 2: All respondents who completed the midline WhatsApp module, regardless of later survey completion.
Labor-Market Behavior and Attitudes (Y_l)	Changed job / started business Desire to improve labor-market status Preference for work–family balance	Three groups analyzed separately: (i) directly treated at midline and followed up at endline (Sample $2 \cap 3$), (ii) indirectly exposed via a treated partner (within Sample 3), and (iii) placebo: treated only at endline (Sample $3 \setminus 2$).

Notes: Sample 2 comprises respondents who engaged with the WhatsApp intervention during the midline module. Sample 3 includes individuals who completed the phone-based endline survey. Outcomes are matched to samples depending on timing of measurement and treatment exposure (direct, indirect through spouse, or placebo)..

Appendix D: Selection Compliance and Weighting

We report several sets of diagnostics to assess covariate balance, propensity score overlap, and weight concentration for both the selection and treatment models, as implied by the identification assumptions above.

First, we compute standardized mean differences (SMDs) for each covariate X_k,

$$\text{SMD}_k = \frac{\bar{X}_{1k} - \bar{X}_{0k}}{s_k}, \qquad s_k^2 = \tfrac{1}{2}\left(\sigma_{1k}^2 + \sigma_{0k}^2\right),$$

where subscripts 1 and 0 index the relevant arms (e.g. $D = 1$ vs. $D = 0$, or $S = 1$ vs. $S = 0$). Our benchmark is $|\text{SMD}_k| < 0.10$ after weighting, and we aim for values below 0.05 whenever possible.

Second, we assess propensity score overlap by plotting arm-specific densities of the estimated propensity score before and after weighting, and by reporting the weighted mass outside the realized common support. For the treatment models these densities correspond to $\Pr(D_i = 1 \mid X_i^D, \text{strata})$; for the selection models they correspond to $\Pr(S_i = 1 \mid Z_i, X_i)$. For each specification we explicitly compute the share of the (weighted) sample with propensity scores outside the intersection of the treated and control supports.

Third, we summarize weight concentration using the effective sample size

$$\text{ESS} = \frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2},$$

computed for the stabilized treatment weights (conditional on selection) and for the selection weights themselves. This provides a compact measure of how much precision is lost due to extreme weights.

Finally, for the treatment models we implement a Hansen J-test of joint balance by recasting the covariate-balance restrictions as overidentifying moments in a GMM formulation; large p-values indicate that any residual discrepancies in covariate moments are not systematic. When the statistic cannot be computed (typically in the smallest subsamples) we report this explicitly.

Standardized mean differences and the Hansen J-tests speak to covariate balance, the propensity score densities and mass outside common support inform overlap, and the effective sample size summarizes the extent to which extreme weights affect precision.

Selective exposure (treatment) diagnostics. Appendix Table D.1 reports post-weighting diagnostics for selective exposure to treatment, i.e. for D_i, in the pooled sample and

separately by gender, for each of the four analysis samples (midline, endline, both, and endline-only). In Panel A (pooled sample), the maximum weighted |SMD| lies between 0.06 and 0.08 across samples, well below our 0.10 benchmark, and the Hansen J-tests deliver large p-values wherever available. Effective sample sizes remain large (above 500 even in the endline-only sample), and the weighted mass outside common support is small (below 3% in all cases).

Panel B (women) and Panel C (men) show that this good performance carries over to gender-specific analyses. For women, the maximum weighted |SMD| ranges from about 0.04 to 0.11 across samples, with most covariates much closer to zero; the J-tests never reject joint balance, and the mass outside common support is modest, albeit somewhat higher (around 8.5%) in the endline-only sample where the ESS is correspondingly smaller. For men, weighted |SMD| values remain below roughly 0.13 in all samples; residual imbalance and heavier tails are somewhat more pronounced in the both-survey and endline-only samples, as reflected in lower ESS and slightly larger mass near the extremes of the propensity score, but still within standard benchmarks for acceptable overlap.

Figures D.1–D.3 provide the corresponding visual evidence. In the raw panels, the propensity score densities for $D = 1$ and $D = 0$ are clearly shifted in several samples, especially for men in the both-survey and endline-only samples. In the weighted panels, the treated and control densities almost coincide over the interior of the support in every sample, with divergence confined to the tails. This pattern is fully consistent with the small post-weighting SMDs, the limited mass outside common support, and the ESS values reported in Table D.1. Overall, selective exposure appears to be a relatively minor threat that is fully addressed by the weighting strategy.

Attrition (selection into samples) diagnostics. Appendix Table D.3 reports analogous diagnostics for selection into each attrition-defined sample, based on the stabilized selection weights w_i^S. For the pooled sample, the maximum weighted |SMD| does not exceed about 0.03 in any sample. Within gender, the largest weighted |SMD| remains below roughly 0.10 in every case, even though unweighted SMDs can be as large as 0.24 before applying the selection weights. This pattern indicates that, conditional on X_i, included and non-included units are closely comparable covariate by covariate, both overall and within gender.

Figures D.4–D.6 plot the arm-specific densities of the estimated selection propensity $\Pr(S_i = 1 \mid Z_i, X_i)$ for the pooled sample and separately for women and men. In the raw panels, the $S = 1$ and $S = 0$ distributions display modest shifts, especially in the endline-based samples, consistent with non-random retention conditional on X_i. After

reweighting with the selection IPW, the right-hand panels display tight alignment of the two densities across the support, with only mild tails at the extremes of the propensity score. This visual evidence is in line with the near-zero mass outside common support and the high effective sample sizes documented in Table D.3.

The selective exposure and attrition diagnostics jointly show that, absent weighting, both effective treatment receipt (D_i) and sample inclusion are correlated with baseline covariates in ways that would bias unweighted estimates. The combination of selection weights and IPWRA for D_i delivers tight post-weighting covariate alignment, adequate overlap, and moderate weight concentration, with the Hansen J-tests supporting joint balance wherever they are defined. We therefore treat the IPW-adjusted estimates as our preferred specification and report unweighted OLS alongside as a robustness check.

Table D.1. Selective Exposure: Post-weighting Diagnostics — Pooled and by Gender

	Max |SMD| (raw)	Max |SMD| (weighted)	p-value J-test	ESS (All)	Mass outside CS (%)
Panel A. Pooled sample					
Midline Survey (Sample 2)	0.110	0.060	0.998	1,169	0.8
Endline Survey (Sample 3)	0.154	0.060	0.967	1,326	0.2
Both (Sample 2∩3)	0.235	0.075	0.958	993	0.2
Endline Only (Sample 3\2)	0.198	0.071	—	511	2.5
Panel B. Women					
Midline Survey (Sample 2)	0.140	0.042	0.997	727	0.8
Endline Survey (Sample 3)	0.229	0.075	0.958	740	1.4
Both (Sample 2∩3)	0.239	0.106	0.835	598	0.6
Endline Only (Sample 3\2)	0.346	0.109	0.982	192	8.5
Panel C. Men					
Midline Survey (Sample 2)	0.170	0.097	0.995	411	0.4
Endline Survey (Sample 3)	0.193	0.091	0.978	576	0.5
Both (Sample 2∩3)	0.208	0.124	—	370	1.1
Endline Only (Sample 3\2)	0.214	0.092	—	289	2.6

Notes: Each row corresponds to one follow-up sample: (2) Midline Survey; (3) Endline Survey; (2∩3) Both; (3\2) Endline Only. Panel A reports diagnostics for the pooled sample (men and women), Panel B for women only, and Panel C for men only. Diagnostics are computed from stabilized ATT inverse-probability weights for treatment D based on a rich baseline covariate set X (21 covariates: demographics, household composition, education and labor status, and baseline first- and second-order beliefs). "Max |SMD|" is the largest absolute standardized mean difference across all covariates, before and after weighting. The J-test p-value comes from a Hansen overidentification test of joint covariate balance; when the statistic is not available for a given sample, we report —. ESS (All) denotes the effective sample size $(\sum_i w_i)^2 / \sum_i w_i^2$, and "Mass outside CS" is the weighted share of observations with propensity scores outside the common-support interval.

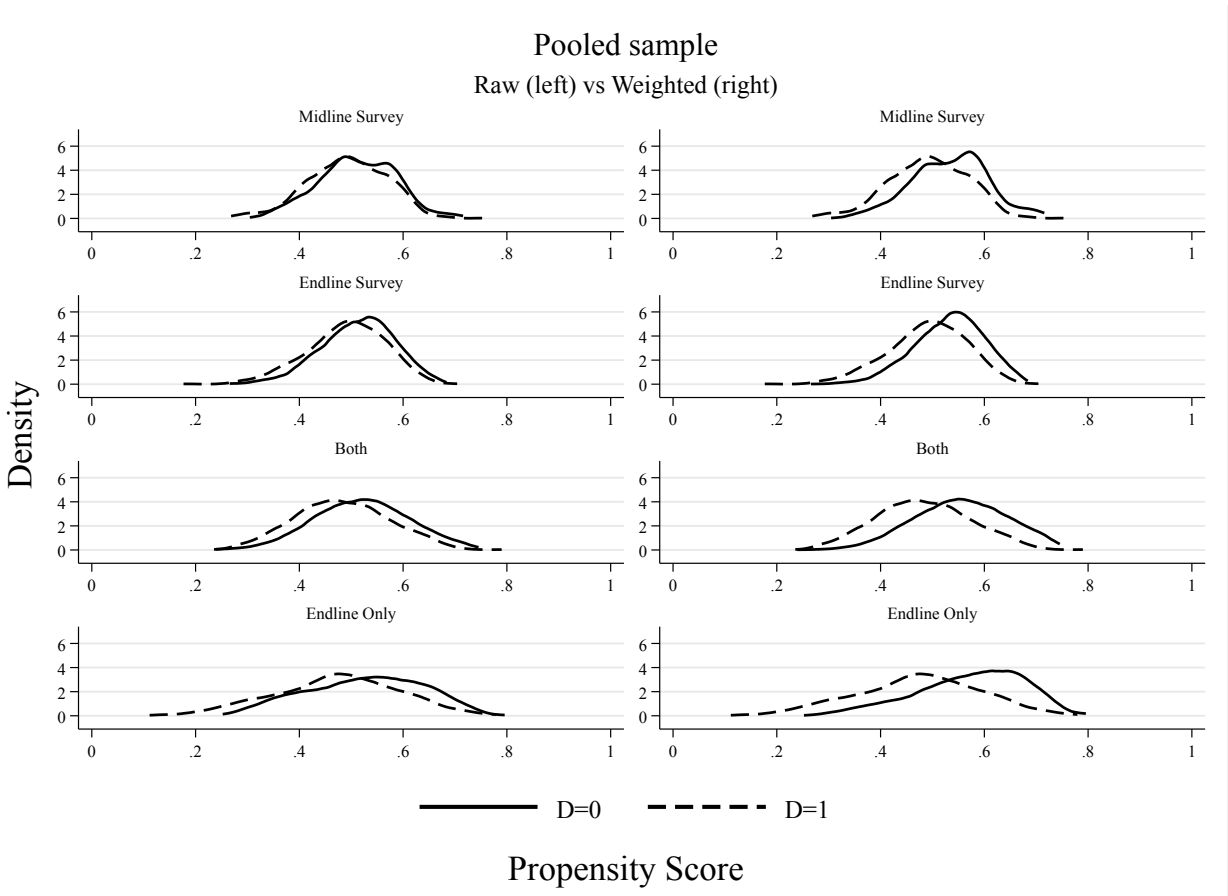


Figure D.1. Propensity Score Densities - Total Sample

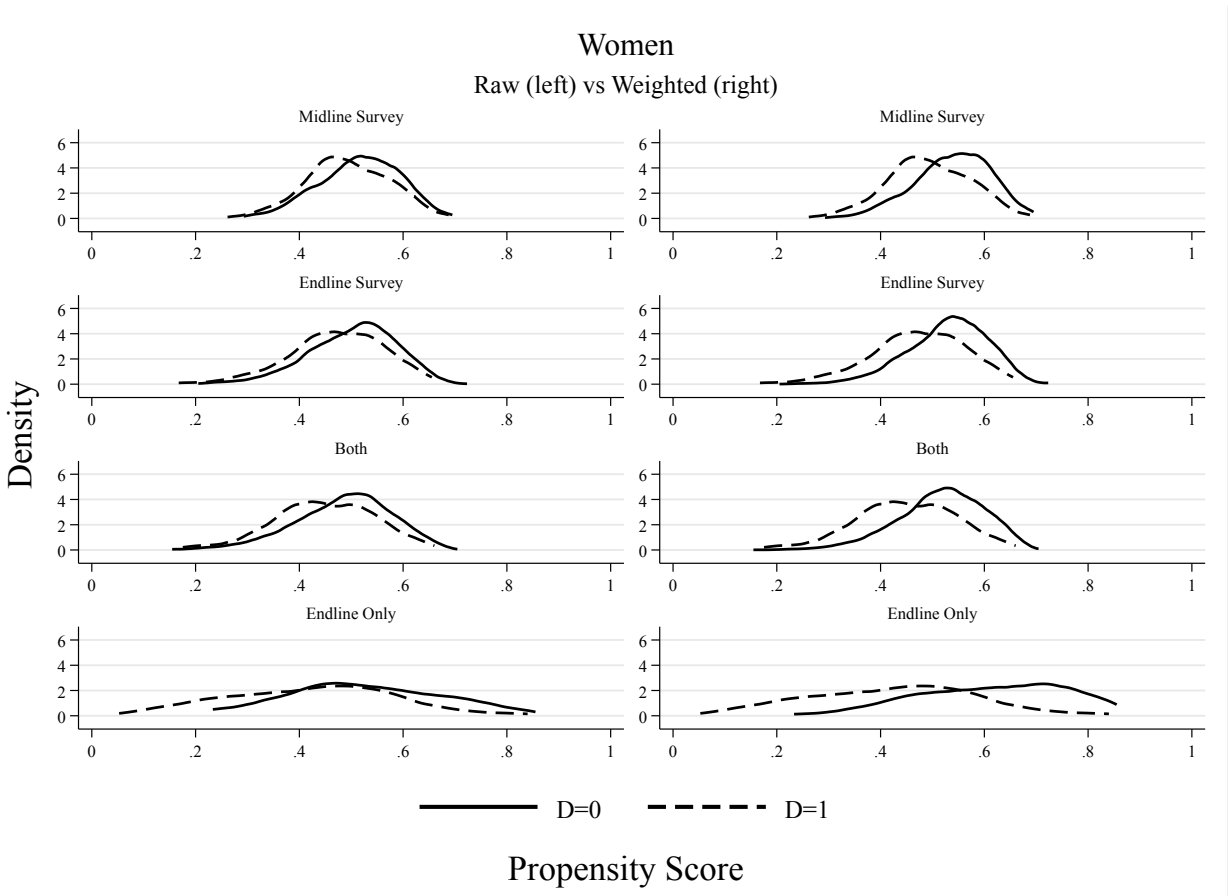


Figure D.2. Propensity Score Densities - Women

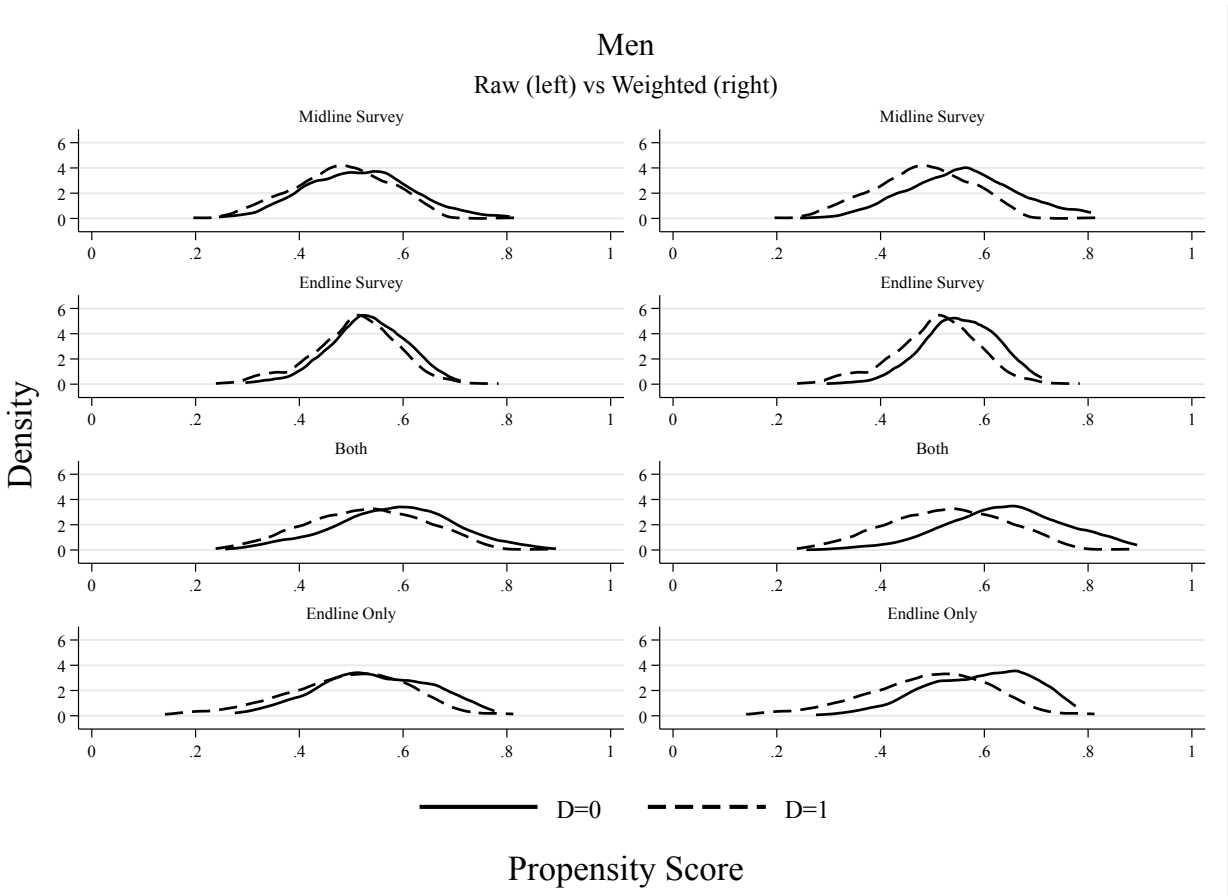


Figure D.3. Propensity Score Densities - Men

Table D.2. Balance Test for Different Samples (Dependent Variable: Treated = 1)

Variable	**Entire Sample** Sample (1)	**Midline Survey** Sample (2)	**Endline Survey** Sample (3)	**Both Surveys** Sample (2)∩(3)	**Endline Only (No Midline)** Sample (3)\(2)
Female	-0.000	-0.015	0.041*	0.112***	0.091*
	(0.005)	(0.023)	(0.017)	(0.022)	(0.038)
Household members	0.005	0.000	0.002	0.003	-0.020
	(0.021)	(0.027)	(0.028)	(0.028)	(0.038)
Adults older than 60 years	0.044	0.048	0.086	0.089	0.104
	(0.048)	(0.069)	(0.070)	(0.076)	(0.090)
Children under 18 years	0.025	0.021	0.028	0.031	0.054
	(0.027)	(0.035)	(0.036)	(0.036)	(0.049)
Permanent care (1 = Yes)	-0.001	0.037	0.007	0.039	-0.047
	(0.026)	(0.036)	(0.037)	(0.038)	(0.051)
Age	0.001	0.001	0.004	0.003	0.007*
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Age difference (husband–wife)	0.001	-0.000	-0.003	-0.004	-0.005
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
Low education	-0.048	-0.070	-0.005	-0.063	0.083
	(0.039)	(0.064)	(0.058)	(0.061)	(0.080)
Medium education	-0.033	-0.027	-0.024	-0.040	0.041
	(0.028)	(0.045)	(0.041)	(0.044)	(0.059)
Wife more educated than husband	0.015	0.049	-0.036	-0.038	-0.060
	(0.032)	(0.046)	(0.046)	(0.045)	(0.060)
Husband inactive	0.030	0.020	-0.002	0.048	-0.038
	(0.058)	(0.077)	(0.076)	(0.080)	(0.101)
Husband unemployed	-0.027	-0.119	-0.032	-0.050	0.060
	(0.055)	(0.075)	(0.074)	(0.071)	(0.102)
Stratum 1 & 2	-0.045	-0.078	-0.072	-0.021	-0.101
	(0.040)	(0.058)	(0.056)	(0.057)	(0.074)
Stratum 3	-0.047	-0.079	-0.034	0.018	0.005
	(0.041)	(0.058)	(0.057)	(0.058)	(0.078)
Reached by phone	−0.055*	−0.077*	-0.039	0.024	-0.049
	(0.025)	(0.039)	(0.039)	(0.040)	(0.052)
FOB mothers work outside†	-0.017	-0.001	-0.056	-0.057	-0.143
	(0.029)	(0.051)	(0.045)	(0.054)	(0.085)
SOB men: mothers work outside‡	0.000	0.000	0.000	0.000	0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
SOB women: mothers work outside	-0.000	0.000	-0.001	-0.001	-0.002
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
SOB partner: mothers work outside	-0.033	-0.080	-0.077	−0.113*	0.027
	(0.033)	(0.051)	(0.051)	(0.056)	(0.087)
Wife inactive	-0.001	0.034	0.064	0.062	0.114
	(0.052)	(0.075)	(0.077)	(0.074)	(0.093)
Wife employed	-0.005	0.026	0.022	0.032	0.062

Continued on next page

Table D.2 – continued from previous page

Variable	Entire Sample Sample (1)	Midline Survey Sample (2)	Endline Survey Sample (3)	Both Surveys Sample (2)∩(3)	Endline Only (No Midline) Sample (3)\(2)
	(0.052)	(0.075)	(0.077)	(0.075)	(0.094)
Husband FOB mothers work	0.019	-0.068	-0.051	-0.069	-0.003
	(0.041)	(0.058)	(0.058)	(0.063)	(0.092)
Constant	0.577***	0.619***	0.579***	0.371*	0.431
	(0.114)	(0.169)	(0.166)	(0.173)	(0.224)
Observations	3,458	1,233	1,382	1,102	576
R^2	0.008	0.020	0.028	0.046	0.056
F-stat (all baseline covariates)	0.91	0.95	1.39	2.80	1.73
p-value (all baseline covariates)	0.565	0.516	0.120	0.000	0.029
F-stat (belief variables only)	0.77	0.74	1.79	1.87	1.48
p-value (belief variables only)	0.544	0.567	0.128	0.114	0.208
F-stat (non-belief covariates)	0.96	1.03	1.18	2.93	1.54
p-value (non-belief covariates)	0.495	0.417	0.282	0.000	0.088

Notes. Each column reports OLS estimates from regressing the treatment indicator Z_i on the full set of baseline covariates, including stratification fixed effects. Standard errors in parentheses, clustered at the household level. The bottom panel reports R^2, sample size, and joint *F*-tests for (i) all baseline covariates, (ii) belief variables only, and (iii) non-belief covariates. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. †: FOB = First-Order Beliefs; ‡: SOB = Second-Order Beliefs.

Table D.3. Attrition diagnostics: maximum standardized mean differences by sample

	Pooled		Women		Men	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Midline Survey (Sample 2)	0.241	0.025	0.172	0.058	0.242	0.097
Endline Survey (Sample 3)	0.210	0.016	0.198	0.069	0.231	0.087
Both (Sample 2∩3)	0.244	0.031	0.241	0.088	0.257	0.073
Endline Only (Sample 3\2)	0.128	0.022	0.106	0.098	0.125	0.082

Notes: Entries report, for each attrition-defined sample, the maximum absolute standardized mean difference (SMD) across the baseline covariates used in the selection model. "Unweighted" SMDs compare included versus non-included units without weights; "weighted" SMDs use stabilized selection weights w_i^S targeting S=1. Columns labeled "Pooled", "Women", and "Men" are computed separately within each subsample. In all cases, post-weighting SMDs are comfortably below the usual |SMD| < 0.10 benchmark, indicating that, conditional on X_i, selection into each sample is as-good-as random.

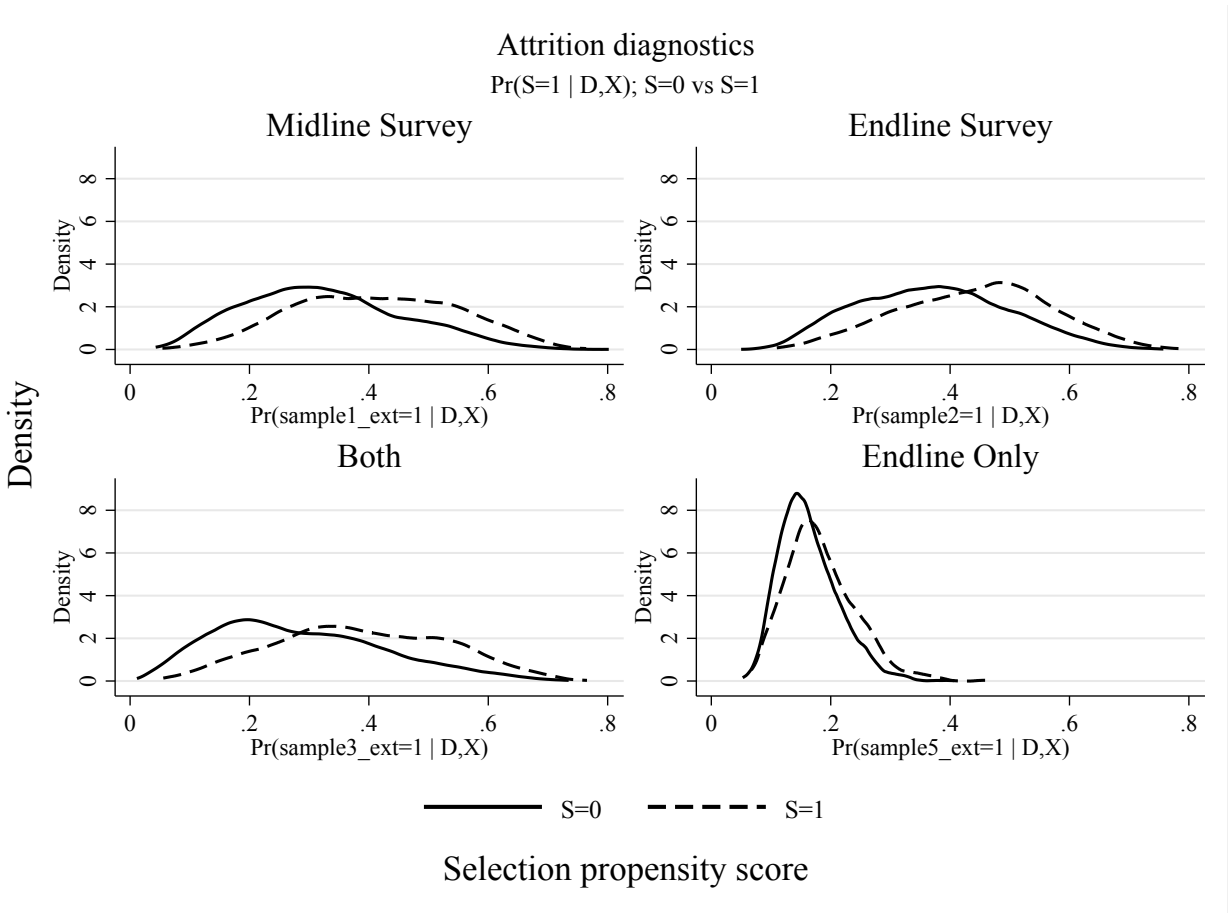


Figure D.4. Propensity Score Densities - Total

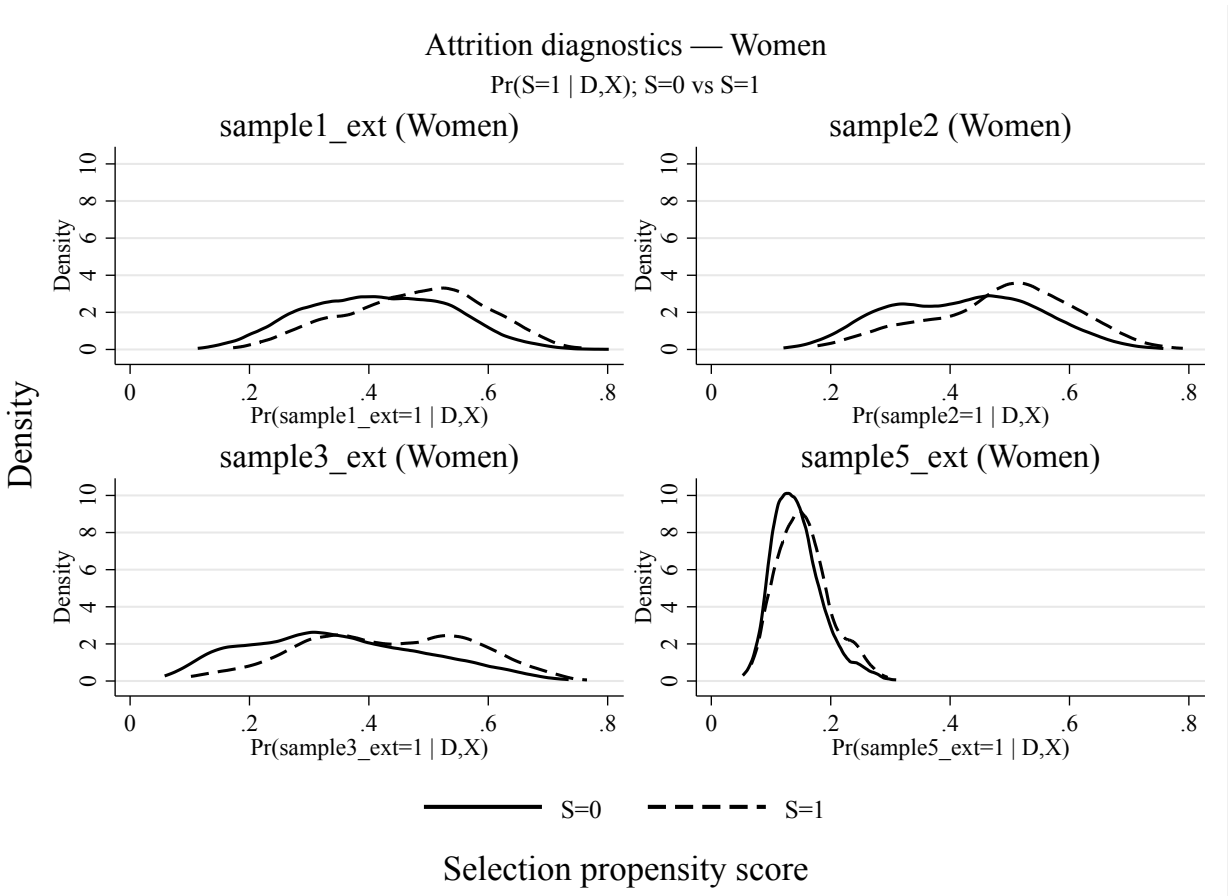


Figure D.5. Propensity Score Densities - Women

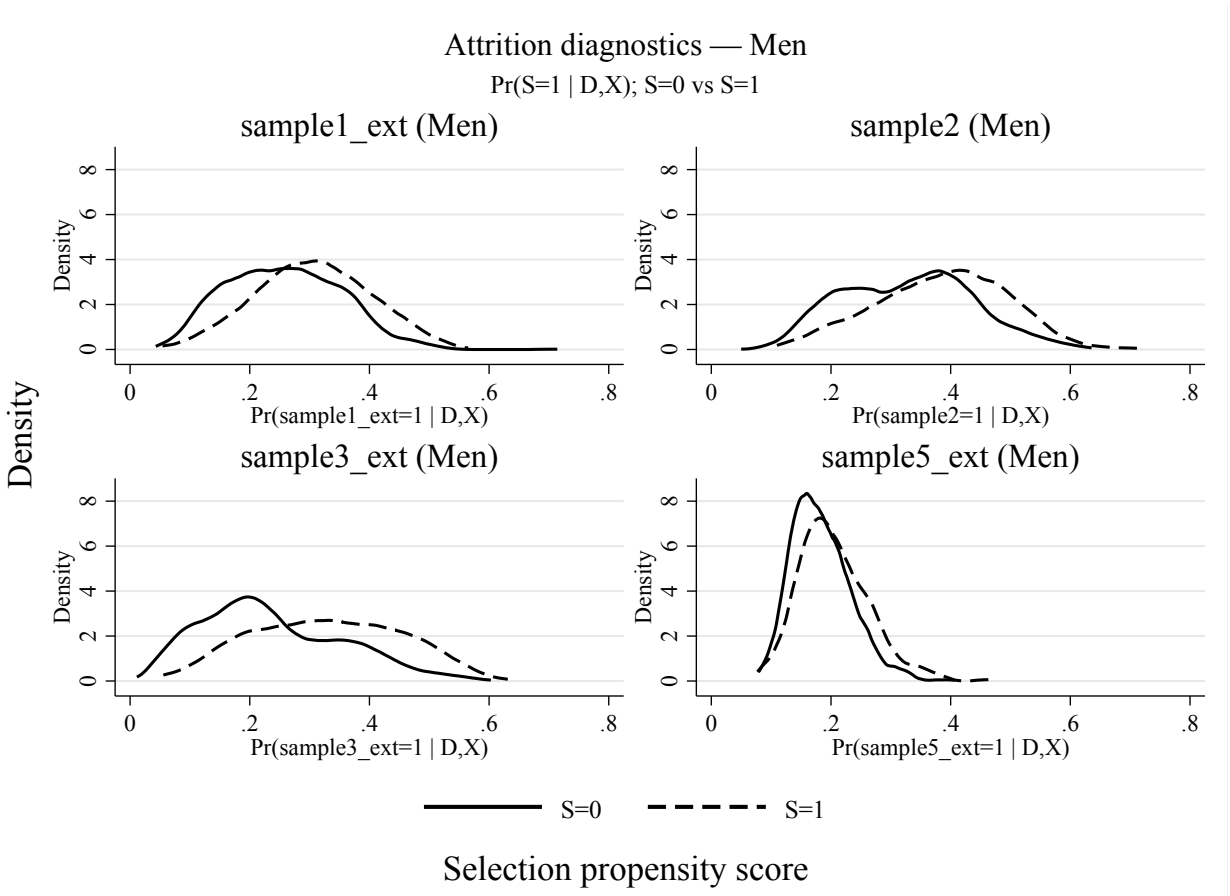


Figure D.6. Propensity Score Densities - Men

Table D.4. Baseline Characteristics: Engagers vs. Non-Engagers

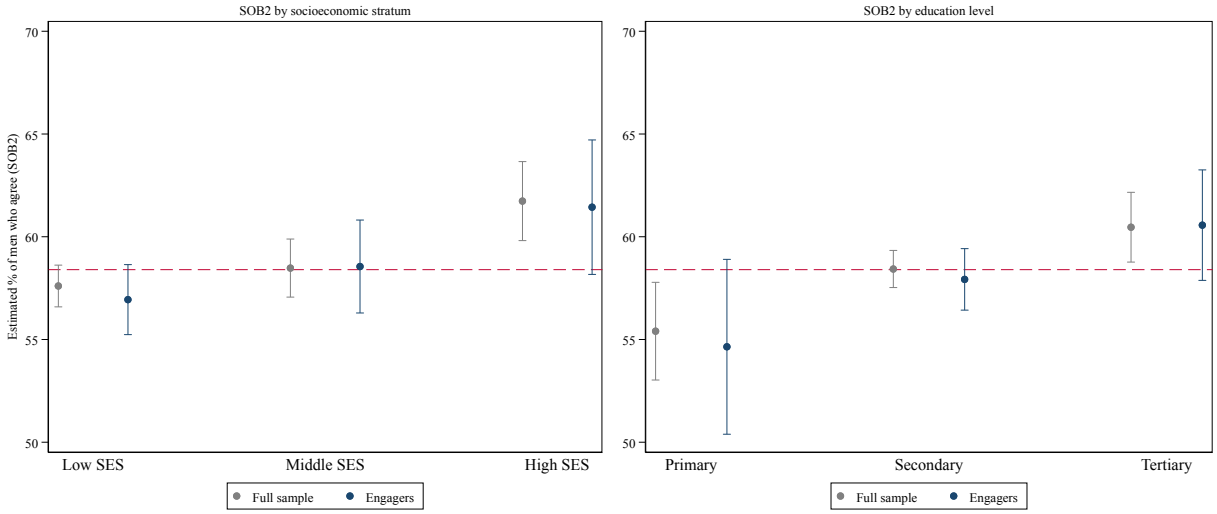
Variable	(1) Non-Engagers N	(1) Non-Engagers Mean (SE)	(2) Engagers N	(2) Engagers Mean (SE)	(1)–(2) Pairwise *t*-test N	(1)–(2) Pairwise *t*-test Difference
FOB†: Mothers of young children should work (%)	2228	0.903 (0.006)	1236	0.887 (0.009)	3464	0.016*
SOB‡: Estimated % of men who agree	2228	58.592 (0.487)	1236	58.050 (0.643)	3464	0.542
SOB: Estimated % of women who agree	2228	79.389 (0.419)	1236	78.879 (0.576)	3464	0.510
SOB: Think partner agrees (%)	2228	0.925 (0.006)	1236	0.908 (0.008)	3464	0.017
FOB: Equal care sharing (%)	2228	0.983 (0.003)	1236	0.987 (0.003)	3464	−0.004
SOB: Estimated % women agree (equal care)	2228	85.982 (0.400)	1236	87.785 (0.503)	3464	−1.803***
SOB: Think partner agrees (equal care) (%)	2228	0.945 (0.005)	1236	0.917 (0.008)	3464	0.029***
FOB: Subsidize green transport (%)	2228	0.935 (0.005)	1236	0.951 (0.006)	3464	−0.017**
Employed (%)	2228	0.752 (0.009)	1236	0.641 (0.014)	3464	0.111***
Looking for a job (%)	2228	0.057 (0.005)	1236	0.056 (0.007)	3464	0.001
Inactive (%)	2228	0.191 (0.008)	1236	0.303 (0.013)	3464	−0.112***
Age (years)	2228	33.307 (0.175)	1236	33.697 (0.224)	3464	−0.389
Primary education (%)	2228	0.141 (0.007)	1236	0.098 (0.008)	3464	0.043***
Secondary education (%)	2228	0.686 (0.010)	1236	0.733 (0.013)	3464	−0.047***
Tertiary education (%)	2228	0.173 (0.008)	1236	0.169 (0.011)	3464	0.004
Household members	2228	3.749 (0.021)	1236	3.920 (0.032)	3464	−0.171***
Number of children	2228	1.489 (0.016)	1236	1.604 (0.023)	3464	−0.116***
At least one member requiring permanent care (%)	2228	0.302 (0.010)	1236	0.353 (0.014)	3464	−0.051***

Notes. Engagers (column 2) are respondents who interacted with the WhatsApp chatbot at least through the information message (`sample1_ext`=1); non-engagers stopped before reaching that point. Column (3) reports the mean difference from a pairwise regression controlling for partner employment status and partner first-order belief (FOB2). Standard errors in parentheses. Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.10$. †: FOB = First-Order Beliefs; ‡: SOB = Second-Order Beliefs.

Table D.5. Placebo-Selection Diagnostics: Treatment Engagers vs. Control Engagers

Variable	(1) Control Engagers N	(1) Control Engagers Mean (SE)	(2) Treatment Engagers N	(2) Treatment Engagers Mean (SE)	(1)–(2) Pairwise *t*-test N	(1)–(2) Pairwise *t*-test Difference
FOB†: Mothers of young children should work (%)	616	0.893 (0.012)	620	0.881 (0.013)	1236	0.012
SOB‡: Estimated % of men who agree	616	57.609 (0.923)	620	58.489 (0.895)	1236	−0.880
SOB: Estimated % of women who agree	616	78.516 (0.819)	620	79.240 (0.811)	1236	−0.724
SOB: Think partner agrees (%)	616	0.920 (0.011)	620	0.895 (0.012)	1236	0.025
FOB: Equal care sharing (%)	616	0.985 (0.005)	620	0.989 (0.004)	1236	−0.003
SOB: Estimated % women agree (equal care)	616	87.317 (0.724)	620	88.250 (0.698)	1236	−0.933
SOB: Think partner agrees (equal care) (%)	616	0.909 (0.012)	620	0.924 (0.011)	1236	−0.015
FOB: Subsidize green transport (%)	616	0.951 (0.009)	620	0.952 (0.009)	1236	−0.000
Employed (%)	616	0.628 (0.019)	620	0.653 (0.019)	1236	−0.025
Looking for a job (%)	616	0.065 (0.010)	620	0.047 (0.008)	1236	0.018
Inactive (%)	616	0.307 (0.019)	620	0.300 (0.018)	1236	0.007
Age (years)	616	33.365 (0.304)	620	34.026 (0.329)	1236	−0.661
Primary education (%)	616	0.102 (0.012)	620	0.094 (0.012)	1236	0.009
Secondary education (%)	616	0.742 (0.018)	620	0.724 (0.018)	1236	0.018
Tertiary education (%)	616	0.156 (0.015)	620	0.182 (0.016)	1236	−0.026
Household members	616	3.888 (0.042)	620	3.952 (0.047)	1236	−0.064
Number of children	616	1.581 (0.030)	620	1.627 (0.034)	1236	−0.046
At least one member requiring permanent care (%)	616	0.338 (0.019)	620	0.368 (0.019)	1236	−0.030

Notes. Sample restricted to respondents who engaged with the WhatsApp chatbot (`sample1_ext`=1; $N = 1{,}236$: 616 control, 620 treatment). Column (3) reports the mean difference between treatment and control engagers from a pairwise regression controlling for partner employment status and partner first-order belief (FOB) about maternal employment. No difference is significant at conventional levels, supporting the validity of the placebo group. Standard errors in parentheses. Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.10$. †: FOB = First-Order Beliefs; ‡: SOB = Second-Order Beliefs.



Notes: Points = means; capped bars = 95% confidence intervals. Dashed line = city-wide average (
%4.1f
58.39896073903002
).
Full sample includes all 3,464 baseline respondents; Engagers interacted with the WhatsApp chatbot (n = 1,236).

Figure D.7. Second-Order Belief (SOB) about maternal employment by Subgroup: Full Sample vs. Engagers

Notes. Each panel shows mean baseline second-order beliefs SOB (estimated share of men in Bogotá who support maternal employment) with 95% confidence intervals, separately for the full baseline sample (gray) and chatbot engagers (navy). Left panel: by SES stratum (low = strata 1–2; middle = stratum 3; high = strata 4–6). Right panel: by education level. Dashed line = city-wide average. The narrow range across strata (4.1 pp by SES; 5.1 pp by education) indicates that the city-average norm is a valid proxy for any observable subgroup. Within each subgroup, engager and non-engager means are statistically indistinguishable (all $p > 0.15$).

Table D.6. Near-Miss Attrition: Baseline Characteristics by Endline Interview Timing

Variable	(1) Easy (November) N	(1) Mean (SE)	(2) Hard (Dec–Jan) N	(2) Mean (SE)	(1)–(2) Pairwise *t*-test N	(1)–(2) Difference
FOB: Mothers of young children should work (%)	379	0.916	492	0.878	871	0.038*
		(0.014)		(0.015)		
SOB: Estimated % of men who agree	379	58.259	492	59.417	871	−1.158
		(1.177)		(1.042)		
SOB: Estimated % of women who agree	379	80.071	492	79.524	871	0.547
		(0.992)		(0.907)		
SOB: Think partner agrees (%)	379	0.931	492	0.927	871	0.005
		(0.013)		(0.012)		
FOB: Equal care sharing (%)	379	0.997	492	0.982	871	0.016**
		(0.003)		(0.006)		
SOB: Estimated % women agree (equal care)	379	87.425	492	87.937	871	−0.512
		(0.930)		(0.798)		
Age (years)	379	34.116	492	33.896	871	0.220
		(0.395)		(0.364)		
Primary education (%)	379	0.119	492	0.132	871	−0.013
		(0.017)		(0.015)		
Secondary education (%)	379	0.715	492	0.687	871	0.028
		(0.023)		(0.021)		
Tertiary education (%)	379	0.166	492	0.181	871	−0.015
		(0.019)		(0.017)		
Employed (%)	379	0.633	492	0.632	871	0.001
		(0.025)		(0.022)		
Inactive (%)	379	0.314	492	0.303	871	0.011
		(0.024)		(0.021)		
Female (%)	379	0.644	492	0.565	871	0.079**
		(0.025)		(0.022)		

Notes. Endline respondents ($N = 871$) classified by month of interview: November 2024 (easy to reach, $n = 379$) vs. December 2024–January 2025 (hard to reach, $n = 492$). Following Behaghel et al. (2015) and Molina Millán and Macours (2025), respondents interviewed later in the fieldwork window are used as a proxy for those who were harder to contact. Column (3) reports mean differences from pairwise regressions (no additional covariates). The key second-order belief (SOB2) and all other second-order beliefs show no significant difference across interview timing groups. The marginally significant differences in FOB2 (p=0.07) and FOB3 (p=0.04) are discussed in the text; both are included as explicit covariates in the IPWRA procedure, which mitigates their impact on the main estimates. Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.10$. †: FOB = First-Order Beliefs; ‡: SOB = Second-Order Beliefs.

7.1 Appendix E: Results

Table E.1. OLS ATT Estimates for Beliefs and Misperceptions about Maternal Employment (Sample 2 ∩ 3)

	First-Order Belief	Second-Order Beliefs about		Misperceptions (D) about	
		Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)
Panel A. All respondents (OLS ATT with controls)					
Treatment	0.009	3.205**	2.886**	–0.042	–0.051*
Std. Err.	(0.018)	(1.381)	(1.308)	(0.025)	(0.032)
Fisher p-value	[0.609]	[0.019]	[0.032]	[0.111]	[0.089]
N (All)	1102	1102	1102	1102	1102
Mean Dep. Var (All)	0.906	64.369	77.803	0.813	0.525
Panel B. Men					
Treatment	0.013	2.486	3.695	–0.075*	–0.032
Std. Err.	(0.034)	(2.168)	(2.220)	(0.044)	(0.051)
Fisher p-value	[0.696]	[0.259]	[0.119]	[0.093]	[0.503]
N (Men)	453	453	453	453	453
Mean Dep. Var (Men)	0.883	66.684	77.093	0.784	0.525
Panel C. Women					
Treatment	0.003	3.759**	2.480	–0.030	–0.066*
Std. Err.	(0.021)	(1.740)	(1.615)	(0.031)	(0.040)
Fisher p-value	[0.880]	[0.026]	[0.117]	[0.323]	[0.086]
N (Women)	649	649	649	649	649
Mean Dep. Var (Women)	0.921	62.753	78.299	0.834	0.524

Notes: Standard errors in parentheses. Fisher (randomization-inference) p-values in brackets. Stars are based on Fisher p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Outcomes: (1) FOB2: first-order belief that mothers with children under 6 should be allowed to work outside the home; (2)–(3) SOB2H/SOB2M: second-order beliefs about support among men and women in Bogotá; (4)–(5) indicators equal to one if respondents underestimate true support among men or women by at least 5 percentage points. All regressions are estimated by OLS and include the same baseline covariates and stratification variables as in the main specification; only the coefficient on the treatment indicator is reported. Sample 2 ∩ 3 comprises respondents observed in both the *WhatsApp* module and the follow-up phone survey (1,102 individuals: 453 men and 649 women). This table reports the OLS counterpart to the IPWRA estimates in Table 5.

Table E.2. Weighted OLS ATT Estimates for Beliefs and Misperceptions about Maternal Employment (Sample 2 ∩ 3)

	First-Order Belief	Second-Order Beliefs about		Misperceptions (D) about	
		Men	Women	Men	Women
	(1)	(2)	(3)	(4)	(5)
Panel A. All respondents (weighted OLS ATT with controls)					
Treatment	0.002	3.448**	3.804**	–0.072**	–0.049
Std. Err.	(0.020)	(1.609)	(1.492)	(0.030)	(0.036)
p-value	[0.920]	[0.032]	[0.011]	[0.016]	[0.173]
N (All)	1102	1102	1102	1102	1102
Mean Dep. Var	0.901	64.456	77.420	0.806	0.529
Panel B. Men					
Treatment	–0.005	3.453	5.158**	–0.120**	–0.052
Std. Err.	(0.037)	(2.539)	(2.423)	(0.050)	(0.054)
p-value	[0.894]	[0.174]	[0.033]	[0.017]	[0.341]
N (Men)	453	453	453	453	453
Mean Dep. Var	0.873	66.492	76.604	0.774	0.539
Panel C. Women					
Treatment	0.002	3.556*	2.731	–0.036	–0.055
Std. Err.	(0.020)	(1.894)	(1.747)	(0.033)	(0.043)
p-value	[0.904]	[0.061]	[0.118]	[0.277]	[0.203]
N (Women)	649	649	649	649	649
Mean Dep. Var	0.928	62.516	78.198	0.835	0.519

Notes: Standard errors in parentheses. Bracketed values are conventional two-sided p-values from weighted OLS. Stars are based on these p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Outcomes: (1) FOB2: first-order belief that mothers with children under 6 should be allowed to work outside the home; (2)–(3) SOB2H/SOB2M: second-order beliefs about support among men and women in Bogotá; (4)–(5) indicators equal to one if respondents underestimate true support among men or women by at least 5 percentage points. All regressions are estimated by OLS with survey weights and include the same baseline covariates and stratification variables as in the main specification; only the coefficient on the treatment indicator is reported. Sample 2 ∩ 3 comprises respondents observed in both the *WhatsApp* module and the follow-up phone survey (1,102 individuals: 453 men and 649 women). This table reports the weighted OLS counterpart to the IPWRA estimates in Table 5.

Table E.3. OLS ATT Estimates for Beliefs and Misperceptions about Spouse's Attitudes (Sample 2 ∩ 3)

	Working Mothers			Equal Task Sharing		
	First-order (1)	2nd-order (Spouse) (2)	Misperception (D) (3)	First-order (4)	2nd-order (Spouse) (5)	Misperception (D) (6)
Panel A. All respondents (OLS ATT with controls)						
Treatment	0.009	0.054***	–0.027	0.018**	0.032**	–0.023
Std. Err.	(0.018)	(0.016)	(0.030)	(0.009)	(0.016)	(0.019)
Fisher p-value	[0.612]	[0.000]	[0.381]	[0.041]	[0.046]	[0.198]
N	1102	1102	746	1102	1102	746
Mean Dep. Var (All)	0.906	0.916	0.160	0.976	0.915	0.076
Panel B. Men						
Treatment	0.013	0.061**	–0.057	0.004	–0.009	–0.016
Std. Err.	(0.034)	(0.023)	(0.037)	(0.020)	(0.013)	(0.017)
Fisher p-value	[0.700]	[0.024]	[0.154]	[0.861]	[0.554]	[0.404]
N (Men)	453	453	341	453	453	341
Mean Dep. Var	0.883	0.929	0.120	0.962	0.987	0.026
Panel C. Women						
Treatment	0.003	0.051**	–0.007	0.025***	0.055**	–0.032
Std. Err.	(0.021)	(0.023)	(0.040)	(0.009)	(0.025)	(0.033)
Fisher p-value	[0.878]	[0.032]	[0.876]	[0.003]	[0.024]	[0.324]
N (Women)	649	649	405	649	649	405
Mean Dep. Var	0.921	0.906	0.193	0.986	0.864	0.119

Notes: Standard errors in parentheses. Fisher (randomization-inference) p-values in brackets. Stars are based on Fisher p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Working-mothers outcomes: (1) FOB2: first-order belief that mothers with children under 6 should be allowed to work outside the home; (2) SOB2P: second-order belief about the partner's support for working mothers; (3) misperception indicator equal to one if the respondent misperceives the partner's true support for working mothers. Equal-task-sharing outcomes: (4) FOB3: first-order belief that fathers should share care and tasks equally; (5) SOB3P: second-order belief about the partner's support for equal task sharing; (6) misperception indicator equal to one if the respondent misperceives the spouse's true support for equal task sharing. All regressions are estimated by OLS with the same baseline covariates and stratification variables as in the main specification; only the coefficient on the treatment indicator is reported. Sample 2 ∩ 3 comprises respondents observed in both the *WhatsApp* module and the follow-up phone survey (1,102 individuals: 453 men and 649 women). This table reports the OLS counterpart to the IPW estimates in Table 6.

Table E.4. Weighted OLS ATT Estimates for Beliefs and Misperceptions about Spouse's Attitudes (Sample 2 ∩ 3)

	Working Mothers			Equal Task Sharing		
	First-order (1)	2nd-order (Spouse) (2)	Misperception (D) (3)	First-order (4)	2nd-order (Spouse) (5)	Misperception (D) (6)
Panel A. All respondents (weighted OLS ATT with controls)						
Treatment	.010	.057***	–.021	.019*	.035**	–.016
Std. Err.	(.019)	(.016)	(.031)	(.010)	(.015)	(.020)
p-value						
N (All)	1102	1102	746	1102	1102	746
Mean Dep. Var	.904	.924	.152	.976	.926	.075
Panel B. Men						
Treatment	.006	.057**	–.039	.014	–.003	–.011
Std. Err.	(.035)	(.024)	(.041)	(.020)	(.015)	(.020)
p-value	[.855]	[.016]	[.336]	[.477]	[.830]	[.570]
N (Men)	453	453	341	453	453	341
Mean Dep. Var	.880	.938	.117	.963	.983	.030
Panel C. Women						
Treatment	.007	.058**	–.007	.023***	.069***	–.031
Std. Err.	(.020)	(.023)	(.041)	(.008)	(.025)	(.035)
p-value	[.724]	[.012]	[.874]	[.003]	[.006]	[.386]
N (Women)	649	649	405	649	649	405
Mean Dep. Var	.926	.911	.193	.989	.871	.126

Notes: Standard errors in parentheses. Bracketed values are conventional two-sided p-values from weighted OLS. Stars are based on these p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Working-mothers outcomes: (1) indicator that spouses disagree on the working-mothers statement; (2) indicator that the respondent expects the spouse to hold the opposite view; (3) indicator that the respondent misperceives the spouse's beliefs about working mothers. Equal-task-sharing outcomes: (4) indicator that spouses disagree on the equal-task-sharing statement; (5) indicator that the respondent expects the spouse to hold the opposite view; (6) indicator that the respondent misperceives the spouse's beliefs about equal task sharing. All regressions are estimated by OLS with survey weights and include the same baseline covariates and stratification variables as in the main specification; only the coefficient on the treatment indicator is reported. Sample sizes differ by column because disagreement and misperception outcomes are defined only for couples. Sample 2 ∩ 3 comprises respondents observed in both the *WhatsApp* module and the follow-up phone survey.

Table E.5. Weighted OLS ATT Estimates for Disagreement and Misperceptions about Spouse's Attitudes (Sample 2 ∩ 3)

	Working Mothers			Equal Task Sharing		
	First-order (1)	2nd-order (Spouse) (2)	Misperception (D) (3)	First-order (4)	2nd-order (Spouse) (5)	Misperception (D) (6)
Panel A. All respondents (weighted OLS ATT with controls)						
Treatment	0.025	–0.015	–0.021	–0.009	–0.041**	–0.016
Std. Err.	(0.041)	(0.020)	(0.031)	(0.023)	(0.017)	(0.020)
p-value	[0.542]	[0.453]	[0.498]	[0.696]	[0.016]	[0.424]
N (All)	746	1102	746	746	1102	746
Mean Dep. Var	0.199	0.114	0.152	0.054	0.087	0.075
Panel B. Men						
Treatment	–0.003	0.001	–0.039	–0.017	0.007	–0.011
Std. Err.	(0.048)	(0.032)	(0.041)	(0.026)	(0.020)	(0.020)
p-value	[0.957]	[0.984]	[0.336]	[0.523]	[0.741]	[0.570]
N (Men)	341	453	341	341	453	341
Mean Dep. Var	0.187	0.109	0.117	0.052	0.037	0.030
Panel C. Women						
Treatment	0.044	–0.022	–0.007	0.001	–0.088***	–0.031
Std. Err.	(0.043)	(0.027)	(0.041)	(0.025)	(0.025)	(0.035)
p-value	[0.308]	[0.417]	[0.874]	[0.957]	[0.001]	[0.386]
N (Women)	405	649	405	405	649	405
Mean Dep. Var	0.212	0.119	0.193	0.056	0.135	0.126

Notes: Standard errors in parentheses. Bracketed values are conventional two-sided p-values from weighted OLS. Stars are based on these p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Working-mothers outcomes: (1) indicator that spouses disagree on the working-mothers statement; (2) indicator that the respondent expects the spouse to hold the opposite view; (3) indicator that the respondent misperceives the spouse's beliefs about working mothers. Equal-task-sharing outcomes: (4) indicator that spouses disagree on the equal-task-sharing statement; (5) indicator that the respondent expects the spouse to hold the opposite view; (6) indicator that the respondent misperceives the spouse's beliefs about equal task sharing. All regressions are estimated by OLS with survey weights and include the same baseline covariates and stratification variables as in the main specification; only the coefficient on the treatment indicator is reported. Sample sizes differ by column because disagreement and misperception outcomes are defined only for couples. Sample 2 ∩ 3 comprises respondents observed in both the *WhatsApp* module and the follow-up phone survey.

Table E.6. Changes in interest about wife's LFP – Baseline and Weighted-IPW OLS

Dependent Variable	Wife Should Attend Course	Self: Interested	Partner: Believed Interested
Panel A: All respondents (Baseline OLS)			
Treatment	0.067	-0.031	0.012
(Std. err)	(0.051)	(0.048)	(0.051)
p-value	[0.192]	[0.513]	[0.822]
{Fisher p-value}	{0.011}	{0.238}	{0.721}
Women	0.412	0.096	-0.239
(Std.Err.)	(0.041)	(0.038)	(0.045)
p-value	[0.000]	[0.011]	[0.000]
Treatment × **Women**	-0.077	0.033	-0.015
(Std.Err.)	(0.059)	(0.055)	(0.063)
p-value	[0.192]	[0.551]	[0.808]
{Fisher p-value}	{0.000}	{0.000}	{0.000}
N	1017	984	973
Mean Dep. Var (All)	0.705	0.789	0.459
Panel B: All respondents (IPW-weighted)			
Treatment	0.074	-0.033	-0.004
(Std. Err.)	(0.053)	(0.050)	(0.055)
p-value	[0.168]	[0.513]	[0.937]
Women	0.430	0.095	-0.213
(Std. Err.)	(0.044)	(0.040)	(0.048)
p-value	[0.000]	[0.017]	[0.000]
Treatment *times* **Women**	-0.084	0.036	-0.015
(Std. Err.)	(0.063)	(0.058)	(0.067)
p-value	[0.179]	[0.535]	[0.827]
N (All, weighted)	1017	984	973
Mean Dep. Var (All, weighted)	0.704	0.784	0.447

Notes: Regressions weighted by inverse probability of being in the WhatsApp sample (selection IPW). Coefficients with clustered standard errors at the household level in parentheses ( ). Clustered *p*-values in brackets [ ]. Fisher exact p-value in {} for *Treatment* and *Treatment* × *Women* coefficients were calculated based on 10,000 replications.

Table E.7. OLS ATT Estimates for Job Search and Labor-Market Aspirations

	Started to Change Job			Aspires Better LM Situation		Wants Work–Family Balance	
	(1) Sample 2∩3	(2) Direct/Ind. Exp. (S3)	(3) Placebo	(4) Sample 2∩3	(5) Sample 3	(6) Sample 2∩3	(7) Sample 3
Panel A. All respondents (OLS ATT with controls)							
Treatment	0.059**	0.041	–0.060	0.016	0.005	0.048*	0.036
Std. Err.	(0.028)	(0.027)	(0.048)	(0.032)	(0.027)	(0.031)	(0.026)
Fisher p-value	[0.042]	[0.111]	[0.229]	[0.588]	[0.855]	[0.093]	[0.179]
N (All)	1102	1233	445	1102	1382	1102	1382
Mean Dep. Var (All)	0.731	0.726	0.692	0.505	0.499	0.368	0.366
Panel B. Men							
Treatment	0.007	0.000	–0.042	–0.053	–0.022	0.113**	0.067*
Std. Err.	(0.047)	(0.040)	(0.077)	(0.053)	(0.041)	(0.051)	(0.040)
Fisher p-value	[0.891]	[0.994]	[0.592]	[0.284]	[0.569]	[0.025]	[0.087]
N (Men)	453	549	229	453	603	453	603
Mean Dep. Var (Men)	0.689	0.689	0.655	0.481	0.486	0.358	0.353
Panel C. Women							
Treatment	0.094***	0.076**	–0.098	0.059	0.025	0.004	0.012
Std. Err.	(0.033)	(0.032)	(0.061)	(0.041)	(0.036)	(0.040)	(0.035)
Fisher p-value	[0.003]	[0.018]	[0.120]	[0.140]	[0.503]	[0.913]	[0.749]
N (Women)	649	684	216	649	779	649	779
Mean Dep. Var (Women)	0.761	0.756	0.731	0.521	0.510	0.374	0.376

Notes: Standard errors in parentheses. Fisher (randomization-inference) p-values in brackets. Stars are based on Fisher p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Columns (1)–(3) refer to indicators of having started to change jobs since baseline: main specification using Sample 2 ∩ 3, an alternative coding based on direct/indirect exposure at endline (Sample 3), and a placebo baseline outcome, respectively. Columns (4)–(5) indicate whether the respondent aspires to a better labor-market situation; columns (6)–(7) indicate whether they prioritize balancing work and family life, each measured for Sample 2 ∩ 3 and for the full endline (Sample 3) sample, respectively. All regressions are estimated by OLS and include the same baseline covariates and stratification variables as in the main belief specifications; only the coefficient on the treatment indicator is reported. Sample 2 ∩ 3 comprises respondents observed in both the *WhatsApp* module and the follow-up phone survey (1,102 individuals: 453 men and 649 women). This table reports the OLS counterpart to the IPWRA estimates in Table 8.

Table E.8. Weighted OLS (IPW) Estimates for Job Search and Labor-Market Aspirations

	Started to Change Job			Aspires Better LM Situation		Wants Work–Family Balance	
	(1) Sample 2∩3	(2) Direct/Ind. Exp. (S3)	(3) Placebo	(4) Sample 2∩3	(5) Sample 3	(6) Sample 2∩3	(7) Sample 3
Panel A. All respondents (Weighted OLS with IPW)							
Treatment	0.050	0.032	–0.060	0.010	0.005	0.051	0.036
Std. Err.	(0.031)	(0.028)	(0.048)	(0.036)	(0.027)	(0.034)	(0.026)
Analytical p-value	[0.107]	[0.253]	[0.211]	[0.781]	[0.853]	[0.134]	[0.166]
N (All)	1102	1233	445	1102	1382	1102	1382
Mean Dep. Var (All, weighted)	0.734	0.725	0.692	0.503	0.499	0.363	0.366
Panel B. Men (Weighted OLS with IPW)							
Treatment	0.020	0.005	–0.042	–0.051	–0.022	0.119**	0.067*
Std. Err.	(0.049)	(0.043)	(0.077)	(0.057)	(0.041)	(0.054)	(0.040)
Analytical p-value	[0.687]	[0.914]	[0.582]	[0.368]	[0.590]	[0.027]	[0.090]
N (Men)	453	549	229	453	603	453	603
Mean Dep. Var (Men, weighted)	0.683	0.680	0.655	0.473	0.486	0.364	0.353
Panel C. Women (Weighted OLS with IPW)							
Treatment	0.089***	0.068**	–0.098	0.065	0.025	–0.012	0.012
Std. Err.	(0.034)	(0.033)	(0.061)	(0.043)	(0.036)	(0.041)	(0.035)
Analytical p-value	[0.008]	[0.039]	[0.109]	[0.134]	[0.494]	[0.771]	[0.736]
N (Women)	649	684	216	649	779	649	779
Mean Dep. Var (Women, weighted)	0.783	0.770	0.731	0.531	0.510	0.362	0.376

Notes: Standard errors in parentheses. Analytical p-values in brackets. Stars denote significance based on analytical p-values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Columns (1)–(3) refer to indicators of having started to change jobs since baseline: main specification using Sample 2 ∩ 3, an alternative coding based on direct/indirect exposure at endline (Sample 3), and a placebo baseline outcome, respectively. Columns (4)–(5) capture whether the respondent aspires to a better labor-market situation, and columns (6)–(7) whether they prioritize balancing work and family, each measured for Sample 2 ∩ 3 and for the endline (Sample 3), respectively. All regressions are estimated by OLS using stabilized inverse-probability weights for selection into the relevant analysis sample and for treatment assignment, and include the same baseline covariates and stratification blocks as in the main belief specifications. Sample sizes match those in Table E.7; weighted means of the dependent variable are computed using the same IPW weights. This table reports the weighted-OLS counterpart to the IPWRA estimates in Table 8.

7.2 Appendix F: Heterogeneity results by baseline labor-market attachment of the wife

Table F.1. Belief Updating by Wife Baseline Inactivity: OLS and IPW-Weighted OLS

	First-order Beliefs		Second-order beliefs about Men		about Women		about Spouse	
	OLS	**IPW**	**OLS**	**IPW**	**OLS**	**IPW**	**OLS**	**IPW**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Panel A: All								
Treatment	0.031*	0.032*	2.593*	2.664	2.721**	3.308*	0.029*	0.031
(Std. Err.)	(0.021)	(0.019)	(1.823)	(2.030)	(1.707)	(1.827)	(0.020)	(0.022)
[Fisher p-val]	[0.062]	—	[0.056]	—	[0.044]	—	[0.083]	—
Wife inactive	-0.046	-0.061**	0.668***	0.216	0.413	0.401	-0.036	-0.039
(Std. Err.)	(0.024)	(0.026)	(1.783)	(1.863)	(1.777)	(1.945)	(0.023)	(0.025)
[Fisher p-val]	[1.000]	—	[<0.001]	—	[0.953]	—	[1.000]	—
Treatment × Wife inactive	-0.040***	-0.045	0.701	1.131	0.569	0.081	0.050	0.053
(Std. Err.)	(0.038)	(0.041)	(2.738)	(3.005)	(2.686)	(2.922)	(0.033)	(0.034)
[Fisher p-val]	[<0.001]	—	[1.000]	—	[1.000]	—	[1.000]	—
N	1102	1102	1102	1102	1102	1102	1102	1102
Control mean	0.906	0.905	64.369	64.266	77.803	77.639	0.916	0.921
Panel B: Men								
Treatment	0.045	0.044	4.004*	5.332*	2.245	3.753	0.063**	0.058**
(Std. Err.)	(0.034)	(0.034)	(2.879)	(3.164)	(3.052)	(3.226)	(0.026)	(0.025)
[Fisher p-val]	[0.181]	—	[0.075]	—	[0.343]	—	[0.025]	—
Wife inactive	-0.063	-0.081*	1.218***	1.254	-1.048	-1.001	-0.018	-0.031
(Std. Err.)	(0.040)	(0.044)	(2.514)	(2.615)	(2.785)	(3.063)	(0.033)	(0.035)
[Fisher p-val]	[1.000]	—	[<0.001]	—	[1.000]	—	[1.000]	—
Treatment × Wife inactive	-0.064	-0.082	-2.737***	-4.050	2.719	1.916	-0.001	0.007
(Std. Err.)	(0.069)	(0.076)	(4.362)	(4.929)	(4.717)	(5.086)	(0.046)	(0.049)
[Fisher p-val]	[1.000]	—	[<0.001]	—	[1.000]	—	[1.000]	—
N	453	453	453	453	453	453	453	453
Control mean	0.883	0.875	66.684	66.735	77.093	76.750	0.929	0.936
Panel C: Women								
Treatment	0.014	0.016	2.546	1.986	2.465	2.556	0.012	0.023
(Std. Err.)	(0.025)	(0.023)	(2.355)	(2.549)	(2.107)	(2.259)	(0.029)	(0.031)
[Fisher p-val]	[0.505]	—	[0.147]	—	[0.130]	—	[0.587]	—
Wife inactive	-0.033	-0.046	0.395	-0.607	1.693***	1.569	-0.052	-0.048
(Std. Err.)	(0.029)	(0.031)	(2.494)	(2.586)	(2.304)	(2.488)	(0.032)	(0.035)
[Fisher p-val]	[0.998]	—	[0.976]	—	[<0.001]	—	[0.998]	—
Treatment × Wife inactive	-0.023***	-0.023	2.710	4.031	0.030	-0.668	0.087	0.080*
(Std. Err.)	(0.045)	(0.047)	(3.500)	(3.662)	(3.324)	(3.603)	(0.047)	(0.048)
[Fisher p-val]	[<0.001]	—	[1.000]	—	[1.000]	—	[1.000]	—
N	649	649	649	649	649	649	649	649
Control mean	0.921	0.925	62.753	62.689	78.299	78.206	0.906	0.911

Notes: OLS columns report unweighted regressions; IPW columns apply stabilized inverse-probability weights for selection into Sample 2 ∩ 3, as described in Section 4.3. Robust standard errors clustered at the household level are reported in parentheses. In OLS columns, Fisher randomization-inference p-values are reported in brackets and are used to assign significance stars in those columns; in IPW columns, significance stars are based on the household-clustered standard errors. The dependent variable in columns (1)–(2) is an indicator for agreeing that mothers with children under 6 should be free to work outside the home (first-order belief). Columns (3)–(4) and (5)–(6) report perceived shares (0–100) of, respectively, men and women in Bogotá who agree with the same statement (community second-order beliefs). Columns (7)–(8) report perceived approval of maternal employment by the respondent's spouse/partner (spousal second-order belief).

Table F.2. Labour market related outcome changes by working status of the wife - Baseline OLS and IPW weighted OLS

	Midline outcome		**Endline outcomes**					
	Wife priority for the course		Job search Effort (D)		Aspires better LM Situation (D)		Wants Work-Family Balance (D)	
	OLS	**IPW**	**OLS**	**IPW**	**OLS**	**IPW**	**OLS**	**IPW**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Panel A: All								
Treatment	0.061**	0.065	0.063**	0.060	0.085***	0.097**	-0.002	-0.022
(Std. Err.)	(0.040)	(0.041)	(0.038)	(0.040)	(0.043)	(0.045)	(0.041)	(0.043)
[Fisher p-val]	[0.036]		[0.020]		[0.007]		[0.940]	
Wife inactive	0.077***	0.064	0.023***	0.022	0.034***	0.040	0.014	0.008
(Std. Err.)	(0.041)	(0.043)	(0.035)	(0.036)	(0.040)	(0.042)	(0.038)	(0.039)
[Fisher p-val]	[0.000]		[0.000]		[0.000]		[0.820]	
Treatment × Wife inactive	-0.095***	-0.112*	0.001	0.000	-0.138***	-0.160**	0.112*	0.133**
(Std. Err.)	(0.058)	(0.061)	(0.054)	(0.056)	(0.064)	(0.068)	(0.062)	(0.066)
[Fisher p-val]	[0.000]		[1.000]		[0.000]		[0.069]	
N	1017	1017	1102	1102	1102	1102	1102	1102
Mean Dep. Var	0.705	0.704	0.731	0.745	0.505	0.507	0.368	0.365
Panel B: Men								
Treatment	0.114**	0.128*	-0.033	-0.040	0.038	0.039	0.044	0.047
(Std. Err.)	(0.069)	(0.070)	(0.064)	(0.067)	(0.073)	(0.078)	(0.072)	(0.075)
[Fisher p-val]	[0.034]		[0.465]		[0.460]		[0.365]	
Wife inactive	0.109***	0.111	-0.030***	-0.032	0.040***	0.051	-0.006	-0.023
(Std. Err.)	(0.076)	(0.079)	(0.054)	(0.056)	(0.060)	(0.062)	(0.055)	(0.057)
[Fisher p-val]	[0.000]		[0.005]		[0.002]		[0.999]	
Treatment × Wife inactive	-0.092***	-0.137	0.078***	0.092	-0.185***	-0.193*	0.140	0.129
(Std. Err.)	(0.105)	(0.110)	(0.094)	(0.097)	(0.107)	(0.112)	(0.102)	(0.105)
[Fisher p-val]	[0.000]		[0.005]		[0.002]		[0.999]	
N	373	373	453	453	453	453	453	453
Mean Dep. Var	0.466	0.461	0.689	0.681	0.481	0.468	0.358	0.365
Panel C: Women								
Treatment	0.023	0.027	0.112***	0.108**	0.098**	0.106*	-0.028	-0.048
(Std. Err.)	(0.040)	(0.041)	(0.047)	(0.047)	(0.055)	(0.057)	(0.053)	(0.054)
[Fisher p-val]	[0.408]		[0.000]		[0.015]		[0.472]	
Wife inactive	0.040***	0.021	0.081*	0.074	0.041	0.037	0.028	0.036
(Std. Err.)	(0.041)	(0.044)	(0.046)	(0.045)	(0.054)	(0.057)	(0.052)	(0.055)
[Fisher p-val]	[0.000]		[0.000]		[0.000]		[1.000]	
Treatment × Wife inactive	-0.072***	-0.081	-0.042	-0.047	-0.087***	-0.102	0.071***	0.095
(Std. Err.)	(0.059)	(0.064)	(0.066)	(0.066)	(0.082)	(0.086)	(0.080)	(0.084)
[Fisher p-val]	[0.000]		[1.000]		[0.000]		[0.000]	
N	644	644	649	649	649	649	649	649
Mean Dep. Var	0.843	0.838	0.761	0.785	0.521	0.532	0.374	0.365

Notes: Odd columns report unweighted OLS; even columns report OLS weighted by stabilized inverse-probability weights for inclusion in the relevant analysis sample (IPW). Standard errors are clustered at the household level. In the OLS columns, Fisher randomization-inference (RI) p-values are reported in brackets for the treatment and interaction coefficients; significance stars in the OLS columns correspond to these RI p-values. In the IPW columns, significance stars correspond to conventional two-sided p-values computed from household-clustered standard errors. Stars denote: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

7.3 Appendix G: Heterogeneity results by intrahousehold exposure

Table G.1. Beliefs changes by gender and intra-household exposure - *Baseline OLS and IPW weighted OLS*

	First-order Beliefs		Second-order beliefs about Men		about Women		about Spouse	
	OLS	**IPW**	**OLS**	**IPW**	**OLS**	**IPW**	**OLS**	**IPW**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Panel A: Direct exposure								
Treatment	0.009	0.003	2.985**	3.764	4.205***	5.283**	0.065***	0.062***
(Std. Err.)	(0.033)	(0.036)	(2.122)	(2.489)	(2.169)	(2.425)	(0.023)	(0.024)
Fisher p-val	[0.581]		[0.028]		[0.001]		[0.000]	
Female	0.034***	0.043*	-3.148	-3.182*	1.786***	2.686	-0.012	-0.020
(Std. Err.)	(0.024)	(0.026)	(1.717)	(1.809)	(1.722)	(1.894)	(0.023)	(0.024)
Fisher p-val	[0.000]		[1.000]		[0.000]		[1.000]	
Treatment × Female	-0.004	0.006	0.473	-0.340	-2.121*	-3.353	-0.018	-0.008
(Std. Err.)	(0.039)	(0.042)	(2.730)	(3.093)	(2.693)	(2.973)	(0.032)	(0.033)
Fisher p-val	[0.980]		[1.000]		[0.065]		[1.000]	
N	1102	1102	1102	1102	1102	1102	1102	1102
Mean Dep. Var	0.906	0.905	64.369	64.266	77.803	77.639	0.916	0.921
Panel B: Indirect Treatment								
Treatment	-0.055**	-0.045	-1.891	-0.652	1.225	2.754	-0.030	-0.047
(Std. Err.)	(0.043)	(0.041)	(2.585)	(2.674)	(2.696)	(2.696)	(0.037)	(0.042)
Fisher p-val	[0.041]		[0.347]		[0.531]		[0.237]	
Female	0.037	0.046*	-2.592***	-3.040	1.767***	2.152	-0.009	-0.013
(Std. Err.)	(0.024)	(0.027)	(1.741)	(1.910)	(1.738)	(1.955)	(0.023)	(0.025)
Fisher p-val	[1.000]		[0.000]		[0.000]		[0.000]	
Treatment × Female	0.105	0.063	7.586***	7.606*	-4.203***	-2.605	0.077***	0.092
(Std. Err.)	(0.050)	(0.054)	(4.262)	(3.910)	(4.790)	(4.064)	(0.054)	(0.061)
Fisher p-val	[1.000]		[0.000]		[0.000]		[0.000]	
N	834	834	834	834	834	834	834	834
Mean Dep. Var	0.896	0.865	63.448	63.615	76.628	76.805	0.898	0.871
Panel C: Both Spouses Treated								
Treatment	0.018	0.018	3.133	5.203**	6.622***	9.525***	0.084***	0.089***
(Std. Err.)	(0.037)	(0.039)	(2.308)	(2.567)	(2.382)	(2.313)	(0.024)	(0.023)
Fisher p-val	[0.474]		[0.104]		[0.000]		[0.000]	
Female	0.036	0.047*	-2.963	-2.841	1.555***	2.418	-0.014	-0.021
(Std. Err.)	(0.024)	(0.026)	(1.731)	(1.811)	(1.726)	(1.888)	(0.023)	(0.023)
Fisher p-val	[0.000]		[1.000]		[0.000]		[1.000]	
Treatment × Female	-0.043***	-0.047	3.271	2.618	-5.747***	-7.829**	-0.026	-0.020
(Std. Err.)	(0.049)	(0.053)	(3.170)	(3.386)	(3.187)	(3.181)	(0.036)	(0.035)
Fisher p-val	[0.000]		[1.000]		[0.000]		[1.000]	
N	901	901	901	901	901	901	901	901
Mean Dep. Var	0.900	0.895	64.549	64.989	77.536	77.675	0.913	0.918

Table G.2. Labor-market related outcomes by gender and intra-household exposure - *Baseline OLS and IPW weighted OLS*

	Midline outcome Wife priority for the course		**Endline outcomes** Job search Effort (D)	
	OLS (1)	**IPW** (2)	**OLS** (3)	**IPW** (4)
Panel A: All				
Treatment	0.073**	0.081	0.009	0.005
(Std. Err.)	(0.051)	(0.054)	(0.047)	(0.051)
Fisher p-val	[0.011]		[0.734]	
Female	0.430***	0.437***	0.016***	0.041
(Std. Err.)	(0.042)	(0.043)	(0.034)	(0.036)
Fisher p-val	[0.000]		[0.002]	
Treatment × Female	-0.086***	-0.092	0.079	0.076
(Std. Err.)	(0.059)	(0.062)	(0.057)	(0.060)
Fisher p-val	[0.000]		[1.000]	
N	1017	1017	1102	1102
Mean Dep. Var	0.705	0.702	0.731	0.745
Panel B: Alone (Col. 1 & 2) – Indirect through spouse (Col. 3 & 4)				
Treatment	0.030	0.046	0.014	0.004
(Std. Err.)	(0.068)	(0.070)	(0.053)	(0.055)
Fisher p-val	[0.291]		[0.735]	
Female	0.431***	0.438***	-0.003	0.014
(Std. Err.)	(0.042)	(0.044)	(0.035)	(0.038)
Fisher p-val	[0.000]		[0.995]	
Treatment × Female	-0.010***	-0.022	-0.065	-0.052
(Std. Err.)	(0.075)	(0.078)	(0.099)	(0.098)
Fisher p-val	[0.000]		[0.995]	
N	789	789	834	834
Mean Dep. Var	0.716	0.714	0.706	0.702
Panel C: Both Spouses Treated				
Treatment	0.095**	0.100	0.004	-0.024
(Std. Err.)	(0.060)	(0.063)	(0.054)	(0.060)
Fisher p-val	[0.015]		[0.917]	
Female	0.428***	0.433***	0.008*	0.035
(Std. Err.)	(0.042)	(0.044)	(0.034)	(0.036)
Fisher p-val	[0.000]		[0.069]	
Treatment × Female	-0.164***	-0.179**	0.129	0.136*
(Std. Err.)	(0.074)	(0.078)	(0.067)	(0.071)
Fisher p-val	[0.000]		[0.983]	
N	746	746	901	901
Mean Dep. Var	0.688	0.683	0.724	0.730

7.4 Appendix H: Mediation Analysis

Empirical setup. We use a simple IV–mediation framework to explore whether changes in perceived societal support for maternal employment help explain the (small) effects of the informational treatment on labor-related outcomes. For each of the three follow-up outcomes—(i) job search at follow-up, (ii) strengthening labor-market goals, and (iii) work–family balance goals—we consider two mediators measured at follow-up: perceived societal support from men and from women in Bogotá, denoted $M \in \{\text{perceived societal support from men}, \text{perceived societal support from women}\}$.

For each outcome–mediator pair, we estimate the following system:

$$M_i = \pi D_i + X_i'\gamma + u_i, \tag{2}$$

$$Y_i = \alpha M_i + X_i'\delta + v_i, \tag{3}$$

where D_i is the randomized information treatment, Y_i is one of the three labor outcomes, and X_i is the same vector of baseline controls and stratification variables used in the main IPW specifications. We estimate (3) by 2SLS, instrumenting the mediator M_i with D_i, and we report: (i) the reduced-form effect of D on Y, $\hat{\beta}^{RF}$, from a regression of Y_i on D_i and X_i; (ii) the IV coefficient $\hat{\alpha}$ of M on Y; and (iii) the first-stage F-statistic for D in (2). All models are IPW-weighted and include the same absorbed strata fixed effects as in the main analysis. We present results for the full sample and, separately, for women and men.

With a single endogenous mediator and a single instrument, the standard IV–mediation identity implies

$$\hat{\beta}^{RF} \approx \hat{\alpha}\hat{\pi},$$

where $\hat{\pi}$ is the first-stage effect of D on M. Thus, the "mediated component" $\hat{\alpha}\hat{\pi}$ coincides with the reduced-form effect by construction, and the implied mediated share, $\hat{\alpha}\hat{\pi}/\hat{\beta}^{RF}$, is equal to one. We retain these columns in Tables H.1–H.3 for expositional completeness, but the substantive content is in the signs and magnitudes of $\hat{\beta}^{RF}$, $\hat{\alpha}$, and the strength of the first stage.

Results. Tables H.1–H.3 report the IV–mediation estimates for the full sample, women only, and men only. Across all specifications, point estimates are small and most coefficients are statistically indistinguishable from zero. First-stage F-statistics range from roughly 1.6 to 6.5, indicating that the instrument is, at best, moderately strong for the mediators, and that the IV estimates should be interpreted as descriptive and highly

imprecise.

For the *full sample* (Table H.1), the reduced-form effects of the information treatment on job search and work–family balance goals are positive and on the order of 0.05, while the effect on strengthening labor-market goals is close to zero. The corresponding IV coefficients $\hat{\alpha}$ on perceived support from men and women are also positive but small in magnitude. Together, these patterns are directionally consistent with the idea that higher perceived societal support for working mothers is associated with slightly more active job search and a somewhat stronger emphasis on work–family balance, but the estimates are very imprecise.

For *women* (Table H.2), the reduced-form effect of treatment on job search is larger, around 0.09, and statistically significant at conventional levels. The associated IV coefficients on the mediators are positive (e.g., $\hat{\alpha} \approx 0.025$ for perceived support from men) and of the expected sign, suggesting that women who perceive greater societal support are, if anything, more likely to report active job search at follow-up. For strengthening labor-market goals, the point estimates remain positive but imprecise, while for work–family balance goals the reduced-form effects are slightly negative and close to zero. Given the wide confidence intervals and modest first-stage strength, these results are best read as suggestive rather than as evidence of a well-identified mediation channel.

For *men* (Table H.3), the reduced-form effects on job search are very small, and the estimated IV coefficients are close to zero. The most notable point estimates appear for work–family balance goals, where the reduced-form effect is about 0.12 and the mediator coefficients $\hat{\alpha}$ are positive, again consistent with the notion that higher perceived societal support moves respondents toward placing more weight on balancing work and family. However, the first-stage F-statistics for perceived support from men are around 1.8, and even the stronger first stages for perceived support from women remain modest, so these patterns should not be overinterpreted.

Overall, the mediation exercise does not uncover a strong or statistically robust channel linking the information treatment to labor outcomes via updated second-order beliefs. Nevertheless, the signs of the coefficients are generally aligned with our theory of change: when the point estimates are non-trivial, they tend to indicate that higher perceived societal support for maternal employment is associated with more active job search and a greater stated interest in work–family balance, particularly for women and for men's work–family preferences. In light of the weak first stages and the limited statistical power, we view these patterns as suggestive consistency checks rather than as additional core results.

Table H.1. IV Mediation of Information Treatment on Labor Outcomes (Full Sample)

Outcome, Mediator	RF: $D \to Y$ $\hat{\beta}^{RF}$	IV: $M \to Y$ $\hat{\alpha}$	Mediated $\hat{\alpha}\hat{\pi}$	Share $\hat{\alpha}\hat{\pi}/\hat{\beta}^{RF}$	First-stage F (D $\to$ M)
Panel A. Job search at follow-up					
Perceived societal support from men	0.050	0.014	0.050	1.00	4.59
	(0.031)	(0.011)			
Perceived societal support from women	0.050	0.013	0.050	1.00	6.51
	(0.031)	(0.010)			
Panel B. Strengthening labor-market goals					
Perceived societal support from men	0.010	0.003	0.010	1.00	4.59
	(0.036)	(0.010)			
Perceived societal support from women	0.010	0.003	0.010	1.00	6.50
	(0.036)	(0.009)			
Panel C. Work–family balance goals					
Perceived societal support from men	0.051	0.015	0.051	1.00	4.59
	(0.034)	(0.012)			
Perceived societal support from women	0.051	0.013	0.051	1.00	6.50
	(0.034)	(0.011)			

Notes: Sample includes all respondents interviewed in the follow-up survey with non-missing values for the outcome and mediators (the number of observations ranges from $N = 841$ to $N = 1{,}102$ depending on the outcome). All regressions are IPW-weighted and include the same baseline controls and belief covariates as the main specifications, as well as the same absorbed strata fixed effects. "RF: $D \to Y$" reports the reduced-form effect of the information treatment on the outcome. "IV: $M \to Y$" reports the 2SLS coefficient of the mediator M on Y, instrumented with D. The mediated component is computed as $\hat{\alpha}\hat{\pi}$, where $\hat{\pi}$ is the first-stage effect of D on the mediator; the share divides this component by the reduced-form effect. First-stage F refers to the cluster-robust F test of D in the mediator regression. Perceived societal support from men (sob2_w3) and from women (sob2m_w3) are measured at follow-up. Standard errors, in parentheses, are clustered at the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. With a single endogenous mediator and a single instrument, $\hat{\alpha}\hat{\pi}$ coincides with the reduced-form effect, so the mediated share equals 1 by construction; we retain the column for expositional clarity. Given first-stage F-statistics around 4.6–6.5, the IV results should be interpreted as suggestive rather than precise.

Table H.2. IV Mediation of Information Treatment on Women's Labor Outcomes

Outcome, Mediator	RF: $D \to Y$ $\hat{\beta}^{RF}$	IV: $M \to Y$ $\hat{\alpha}$	Mediated $\hat{\alpha}\hat{\pi}$	Share $\hat{\alpha}\hat{\pi}/\hat{\beta}^{RF}$	First-stage F (D $\to$ M)
Panel A. Job search at follow-up					
Perceived societal support from men	0.089**	0.025*	0.089	1.00	3.53
	(0.034)	(0.0156)			
Perceived societal support from women	0.089**	0.033	0.089	1.00	2.44
	(0.034)	(0.0227)			
Panel B. Strengthening labor-market goals					
Perceived societal support from men	0.065	0.018	0.065	1.00	3.52
	(0.043)	(0.0154)			
Perceived societal support from women	0.065	0.024	0.065	1.00	2.44
	(0.043)	(0.0208)			
Panel C. Work–family balance goals					
Perceived societal support from men	−0.012	−0.003	−0.012	1.00	3.52
	(0.041)	(0.0115)			
Perceived societal support from women	−0.012	−0.004	−0.012	1.00	2.44
	(0.041)	(0.0149)			

Notes: Sample restricted to female respondents interviewed in the follow-up survey with non-missing values for the outcome and mediators (the number of observations ranges from $N = 412$ to $N = 553$ depending on the outcome). All regressions are IPW-weighted and include the same baseline controls and belief covariates as the main specifications, as well as the same absorbed strata fixed effects. "RF: $D \to Y$" reports the reduced-form effect of the information treatment on the outcome. "IV: $M \to Y$" reports the 2SLS coefficient of the mediator M on Y, instrumented with D. The mediated component is computed as $\hat{\alpha}\hat{\pi}$, where $\hat{\pi}$ is the first-stage effect of D on the mediator; the share divides this component by the reduced-form effect. First-stage F refers to the cluster-robust F test of D in the mediator regression. Perceived societal support from men (sob2_w3) and from women (sob2m_w3) are measured at follow-up. Standard errors, in parentheses, are clustered at the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. With a single endogenous mediator and a single instrument, $\hat{\alpha}\hat{\pi}$ coincides with the reduced-form effect, so the mediated share equals 1 by construction; we retain the column for expositional clarity. Given first-stage F-statistics around 2.4–3.5, the IV results should be interpreted as suggestive rather than precise.

Table H.3. IV Mediation of Information Treatment on Men's Labor Outcomes

Outcome, Mediator	RF: $D \to Y$ $\hat{\beta}^{RF}$	IV: $M \to Y$ $\hat{\alpha}$	Mediated $\hat{\alpha}\hat{\pi}$	Share $\hat{\alpha}\hat{\pi}/\hat{\beta}^{RF}$	First-stage F (D $\to$ M)
Panel A. Job search at follow-up					
Perceived societal support from men	0.020	0.006	0.020	1.00	1.63
	(0.049)	(0.0154)			
Perceived societal support from women	0.020	0.004	0.020	1.00	4.20
	(0.049)	(0.0100)			
Panel B. Strengthening labor-market goals					
Perceived societal support from men	-0.051	-0.015	-0.051	1.00	1.85
	(0.057)	(0.0191)			
Perceived societal support from women	-0.051	-0.010	-0.051	1.00	4.53
	(0.057)	(0.0121)			
Panel C. Work–family balance goals					
Perceived societal support from men	0.119	0.035	0.119	1.00	1.85
	(0.054)	(0.1324)			
Perceived societal support from women	0.119	0.023	0.119	1.00	4.53
	(0.054)	(0.0149)			

Notes: Sample restricted to male respondents interviewed in the follow-up survey with non-missing values for the outcome and mediators. All regressions are IPW-weighted and include the same baseline controls and belief covariates as the main specifications, as well as the same absorbed strata fixed effects. "RF: $D \to Y$" reports the reduced-form effect of the information treatment on the outcome. "IV: $M \to Y$" reports the 2SLS coefficient of the mediator M on Y, instrumented with D. The mediated component is computed as $\hat{\alpha}\hat{\pi}$, where $\hat{\pi}$ is the first-stage effect of D on the mediator; the share divides this component by the reduced-form effect. First-stage F refers to the cluster-robust F test of D in the mediator regression. Perceived societal support from men (sob2_w3) and from women (sob2m_w3) are measured at follow-up. Standard errors, in parentheses, are clustered at the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. With a single endogenous mediator and a single instrument, $\hat{\alpha}\hat{\pi}$ coincides with the reduced-form effect, so the mediated share equals 1 by construction; we retain the column for expositional clarity. Given first-stage F-statistics around 1.6–4.5, the IV results should be interpreted as suggestive rather than precise.

Table H.4. Romano-Wolf Step-Down p-values by Outcome Family

	All OLS p / RW p	**Men** OLS p / RW p	**Women** OLS p / RW p
Family F1: Community beliefs (Table 6)			
First-order belief (FOB)	0.624 / 0.980	0.338 / 0.824	0.126 / 0.445
Perceived men's support (SOB)	0.806 / 0.990	0.476 / 0.894	0.093 / 0.401
Perceived women's support (SOB)	0.950 / 0.998	0.855 / 0.972	0.385 / 0.750
Underestimate men's support (Dmis)	0.955 / 0.998	0.749 / 0.972	0.664 / 0.776
Underestimate women's support (Dmis)	0.481 / 0.939	0.740 / 0.972	0.522 / 0.776
Family F2: Spousal beliefs (Table 7)			
Perceived spousal support (working mothers)	0.801 / 0.824	0.289 / 0.524	0.299 / 0.734
Misperception — working mothers†		—	
First-order belief (equal task sharing)	0.561 / 0.824	0.610 / 0.615	0.964 / 0.956
Perceived spousal support (equal tasks)	0.199 / 0.498	0.234 / 0.524	0.426 / 0.734
Misperception — equal tasks†		—	
Family F3: Course allocation (Table 8)			
Wife prioritized for course (y_1)	0.476 / 0.842	0.151 / 0.383	0.754 / 0.981
Own interest in course (y_2)	0.617 / 0.864	0.537 / 0.794	0.971 / 0.996
Partner's interest in course (y_3)	0.926 / 0.936	0.556 / 0.794	0.954 / 0.996
Family F4: Labor-market outcomes (Table 9)			
Job mobility	0.031 / 0.094*	0.998 / 0.998	0.003 / 0.013**
Labor-market aspirations	0.611 / 0.627	0.319 / 0.521	0.145 / 0.234
Work–family balance preference	0.119 / 0.197	0.027 / 0.077*	0.922 / 0.927

Notes: Each column reports OLS p-value / Romano-Wolf step-down p-value for the treatment indicator D. Regressions use unweighted OLS with controls identical to the main IPWRA specifications (stratification fixed effects, baseline lags, demographic controls, clustered SEs at household level). Point estimates are virtually identical to the IPWRA estimates in Tables 5–8. Romano-Wolf p-values (Romano and Wolf, 2005) use 1,000 permutations (seed 12345) within each outcome family. Families are defined by outcome domain: F1 = community beliefs (Table 5), F2 = spousal beliefs (Table 6), F3 = course allocation (Table 7), F4 = labor-market outcomes (Table 8). †Misperception outcomes excluded from Family F2 because they are measured only on the matched-couple subsample ($N = 578$) while other F2 outcomes have $N = 834$; Romano-Wolf is not valid across outcomes with different sample sizes within a family. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$ based on Romano-Wolf p-values.

Table H.5. Lee (2009) Sharp Bounds on Treatment Effects

	Lower	95% CI	Upper	95% CI
Panel A: Course allocation (Table 7); selection = WhatsApp completion; trim ≤ 3%				
All respondents				
Wife prioritized (y1)	0.001	[−0.069, 0.072]	0.023	[−0.040, 0.086]
Own interest (y2)	−0.029	[−0.098, 0.041]	−0.007	[−0.061, 0.046]
Partner interest (y3)	−0.005	[−0.075, 0.066]	0.010	[−0.067, 0.087]
Men (trim = 0.8% for y1)				
Wife prioritized (y1)	0.062	[−0.062, 0.187]	0.071	[−0.054, 0.195]
Own interest (y2)	−0.042	[−0.184, 0.101]	−0.028	[−0.123, 0.067]
Partner interest (y3)	−0.003	[−0.148, 0.142]	0.020	[−0.098, 0.137]
Women (trim = 2.9% for y1)				
Wife prioritized (y1)	−0.036	[−0.133, 0.062]	−0.005	[−0.066, 0.055]
Own interest (y2)	−0.023	[−0.120, 0.074]	0.005	[−0.061, 0.070]
Partner interest (y3)	−0.006	[−0.099, 0.087]	0.008	[−0.085, 0.101]
Panel C: Community beliefs (Table 5); selection = endline completion; trim = 3.2%				
All respondents (pooled)				
FOB: maternal employment norm	−0.038	[−0.101, 0.026]	−0.005	[−0.036, 0.026]
Misperception: community support for men	−0.045	[−0.117, 0.027]	−0.012	[−0.051, 0.027]
Misperception: community support for women	−0.037	[−0.100, 0.026]	−0.004	[−0.065, 0.057]
Panel B: Labor-market outcomes (Table 8); selection = endline completion				
All respondents (trim = 3.2%)				
Job mobility	−0.004	[−0.075, 0.066]	0.029	[−0.023, 0.080]
Labor aspirations	−0.011	[−0.071, 0.049]	0.022	[−0.044, 0.088]
Work-family balance	0.021	[−0.037, 0.080]	0.054	[−0.007, 0.116]
Women (trim = 2.5%)				
Job mobility	0.038	[−0.032, 0.109]	0.064	[−0.024, 0.153]
Labor aspirations	0.009	[−0.078, 0.097]	0.035	[−0.053, 0.123]
Work-family balance	−0.007	[−0.099, 0.086]	0.019	[−0.062, 0.100]
Men (trim = 10.2%)				
Job mobility	−0.096	[−0.225, 0.033]	0.017	[−0.060, 0.093]
Labor aspirations	−0.072	[−0.180, 0.035]	0.041	[−0.065, 0.147]
Work-family balance	0.025	[−0.062, 0.112]	0.139*	[0.016, 0.261]

Notes: Lee (2009) sharp bounds on the average treatment effect, robust to differential attrition under the monotonicity assumption (treatment cannot induce selection out of follow-up). "Lower" and "Upper" are point estimates of β^L and β^U; 95% CIs are bootstrapped (2,000 replications, seed 12345) using the `leebounds` command (Tauchmann). Bootstrapping resamples at the individual level; couple-level clustering is not supported, so CIs are slightly liberal. Selection variable for Panel A is `sample1_ext` (completed WhatsApp module); for Panels B–C it is `sample2` (completed endline phone survey). The treatment effect on both selection variables is negligible and statistically insignificant (+0.004 and −0.011, respectively), implying very small trimming fractions and hence tight bounds for most outcomes. The exception is Panel B men, where differential compliance yields a 10.2% trimming fraction; even so, both bounds for work-family balance are positive. The key result for women's job mobility (Panel B) has both bounds positive, consistent with a true positive treatment effect; for men's work-family balance the upper bound is statistically positive ($p = 0.026$). $^*p < 0.05$ based on 95% bootstrapped CI.

Table H.6. IPWRA Sensitivity to Attrition Weight Specification

	Attrition weight specification			
	(1) Main probit PS	(2) Winsorised cap at $\hat{w}_{95}$	(3) Trimmed drop $\hat{p} < 0.10$	(4) Logit PS logit PS
Panel A: Wife-prioritized course (men, Table 7, Panel B col. 1)				
ATT coefficient	0.093	0.094	0.091	0.094
p-value	0.115	0.086	0.108	0.105
Observations	373	373	368	373
Panel B: Job-search intensity (women, Table 8, Panel C col. 1)				
ATT coefficient	0.096	0.091	0.096	0.095
p-value	0.006	0.009	0.006	0.007
Observations	649	649	649	649
Panel C: Work–family balance (men, Table 8, Panel B col. 3)				
ATT coefficient	0.110	0.101	0.072	0.109
p-value	0.054	0.057	0.198	0.051
Observations	453	453	442	453

Notes: Each cell reports the IPWRA ATT estimate and its two-sided p-value under a different attrition weight specification. Column (1) is the baseline specification used throughout the paper: attrition propensity scores estimated via probit on the full baseline covariate set, with inverse-probability weights $w_i = 1/\hat{p}_i^S$ applied to the selected sample. Column (2) winsorises weights at the 95th percentile ($\hat{w}_{95} = 6.34$ for sample3_ext) to bound the influence of extreme observations while retaining all observations. Column (3) drops the 1% of observations with estimated propensity scores below 0.10 ($N = 11$), which receive the largest weights; the trimmed sample has $N = 442$ for men. Column (4) replaces the probit with a logit specification for the attrition model. All columns use the same IPWRA outcome and treatment models as the main tables, with stratification fixed effects and cluster-robust standard errors at the household level. The p-values are derived from the normal approximation to the IPWRA sampling distribution. Minor differences from the main tables reflect a parsimonious outcome equation (stratification variables only) used here to isolate the sensitivity to weight choice.