

Discussion Paper Series

IZA DP No. 18913

September 2026

Moving Fathers: Workplace Interventions for Paternal Involvement in Childcare

Mari Tanaka

University of Tokyo
and Hitotsubashi University

Hiroko Okudaira

Doshisha University

Mariko Sakka

University of Tsukuba

Shintaro Yamaguchi

University of Tokyo
and IZA@LISER

The IZA Discussion Paper Series (ISSN: 2365-9793) ("Series") is the primary platform for disseminating research produced within the framework of the IZA@LISER Network, an unincorporated international network of labour economists coordinated by the Luxembourg Institute of Socio-Economic Research (LISER). The Series is operated by LISER, a Luxembourg public establishment (établissement public) registered with the Luxembourg Business Registers under number J57, with its registered office at 11, Porte des Sciences, 4366 Esch-sur-Alzette, Grand Duchy of Luxembourg.

Any opinions expressed in this Series are solely those of the author(s). LISER accepts no responsibility or liability for the content of the contributions published herein. LISER adheres to the European Code of Conduct for Research Integrity. Contributions published in this Series present preliminary work intended to foster academic debate. They may be revised, are not definitive, and should be cited accordingly. Copyright remains with the author(s) unless otherwise indicated.

Moving Fathers: Workplace Interventions for Paternal Involvement in Childcare^{*}

Abstract

We test whether employers can shift household allocation by targeting men as fathers. Among 1,225 male employees in four Japanese organizations, we randomized a two-hour work-life balance training and an information treatment correcting misperceptions of colleagues' support. Training increased fathers' weekend childcare by about one hour per day in young-child households and raised spouses' reported paid work; spouses' housework also fell. Training strengthened beliefs that paternity leave benefits workplace functioning but not leave intentions. Information modestly shifted perceived support but affected neither intentions nor information-seeking. Employer training can affect household allocation without changing formal rights or work arrangements.

JEL classification

J13, J16, J22, M54

Keywords

workplace intervention, paternity leave, household time allocation, second-order beliefs, field experiment

Corresponding author

Shintaro Yamaguchi

syamaguchi@e.u-tokyo.ac.jp

^{*} We gratefully acknowledge administrative and financial support from Evidence Studio and financial support from JSPS Grant-in-Aid for Scientific Research (No. 23H00045). We thank Marianne Bertrand, Hitoshi Shigeoka, and participants in various academic workshops and conferences for their valuable comments and suggestions. Finally, we sincerely thank the participating organizations and the collaborating nonprofit organizations for providing data and continued cooperation throughout this study. All errors are our own. The study was approved by the institutional review board of Evidence Studio and registered in the AEA RCT Registry (AEARCTR-0012315).

1 Introduction

Despite decades of progress in women’s educational attainment and labor force participation, gender disparities in household responsibilities remain persistent across developed economies. This persistence carries substantial economic costs. The unequal division of childcare and domestic work following parenthood accounts for a large share of remaining gender gaps in earnings, career advancement, and leadership representation (Cortés and Pan, 2023). More equal sharing of unpaid work may therefore matter not only for fathers’ involvement at home, but also for women’s labor-market opportunities.

Paternity-leave reforms have become a prominent policy tool for increasing men’s involvement in childcare and housework. Non-transferable leave periods reserved for fathers have substantially increased paternity-leave take-up in the Nordic countries, Germany, and Quebec (Dahl et al., 2014; Ekberg et al., 2013; Olafsson and Steingrimsdottir, 2020; Tamm, 2019; Patnaik, 2019). Their effects on household allocation, however, are mixed. In some settings, reforms increased fathers’ domestic engagement or mothers’ paid work beyond the leave period, while in others increased leave-taking produced limited broader changes (Patnaik, 2019; Farré and González, 2019; Tamm, 2019; Ekberg et al., 2013; Kluge and Tamm, 2013; Andresen and Nix, 2026). Even under supportive national policies, workplace norms can constrain fathers’ use of leave (Haas and Hwang, 2019). Formal entitlements may therefore be only one of the constraints shaping paternal involvement.

Japan provides a sharp setting in which to study this possibility. Women do more than four times as much unpaid care work as men, despite high female labor-force participation. At the same time, Japan offers unusually generous paternity-leave benefits, allowing fathers to take up to twelve months of paid leave, yet only about 30 percent of eligible fathers took any leave in 2023 (OECD, 2023; Ministry of Health, Labour and Welfare, 2021, 2024). Survey evidence identifies an unsupportive workplace atmosphere as a key deterrent to fathers’ leave-taking (Mitsubishi UFJ Research and Consulting, 2018). This suggests that the workplace itself, and not only the statutory entitlement, may be a lever on what happens at home.

This paper asks whether the employer can be that lever. Can an employer-delivered intervention change the allocation of paid and unpaid work within the household, including the paid work of the employee’s spouse, without changing statutory leave rights or employees’ formal work arrangements? A second question is how far such an intervention reaches across margins that plausibly differ in adjustment costs. We distinguish beliefs about paternal leave, flexible day-to-day household time, and the discrete, career-relevant decision to take formal leave.

We study these questions in a field experiment involving more than 1,200 male employees across 74 offices in four Japanese organizations. We evaluate two interventions that differ in content and intensity. The first is a two-hour work-life balance training,

funded by the employer and delivered during working hours to male employees and, in a parallel session, to their managers. Offices were randomized to receive the training early or later, where the later group serves as a control. The second is a light-touch information treatment, randomized at the individual level among non-managers, that corrects misperceptions about coworkers' and managers' attitudes toward paternity leave. We conducted a baseline survey and two follow-up surveys between August 2023 and March 2024, measuring beliefs, information-seeking and leave intentions, as well as respondents' childcare, housework, and paid work and, as reported by them, their spouses' time use.

A distinctive feature of both interventions is whom they target. Rather than intervening on women's labor-market behavior or on how women are treated at work, we target men as fathers and seek to change their contribution to unpaid work at home. The training is substantially broader than the information treatment. It explicitly addressed behavior at home, including fathers' tendency to overestimate their own share of childcare and housework and the division of domestic responsibilities within couples. It also discussed workplace performance during the childcare period, used male instructors as examples of working fathers, and was delivered as an employer-sponsored program to both employees and managers. The training is therefore a bundle, and our design does not isolate the effect of any one component. The information treatment instead targets pluralistic ignorance. Japanese men tend to underestimate colleagues' support for paternal involvement (Miyajima and Yamaguchi, 2017); we confirm this pattern in our baseline data and provide participants with the actual attitudes reported within their own organizations.

The training changed household outcomes beyond the treated employee. Spouses of men in treated offices increased their reported weekly paid hours by about 1.5 hours, or 7 percent of the control mean. The distributional pattern is consistent with an intensive-margin response among spouses already working rather than employment entry. Fathers' own time responded most clearly in households where the training content was immediately relevant. In an exploratory analysis of households with a child aged five or below, fathers increased their weekend childcare by about one hour per day, an effect that survives correction for multiple testing ($q = 0.044$); the full-sample estimate was smaller and not robust to the correction. In the same young-child households, spouses' housework also declined, although these estimates are less robust to multiple-testing adjustment. The training additionally strengthened men's belief that greater paternity-leave take-up can improve workplace functioning and increased subsequent information-seeking about leave policies. Stated intentions to take formal leave were unchanged.

The information intervention produced a narrower response. Correcting misperceptions modestly increased perceived support for paternity leave from coworkers and managers, but did not affect respondents' leave intentions, measured both immediately and one to three months later, or subsequent information-seeking. Thus, a light-touch correc-

tion of second-order beliefs affected the beliefs it directly targeted but did not translate into the other outcomes measured for this intervention.

Taken together, the results show that an employer can affect household outcomes without changing leave entitlements or formal work arrangements. The spouse paid-hours result holds in the full sample, while the joint pattern in households with young children—more paternal childcare and less spouse housework—is consistent with reallocation within the household. The pattern across margins is also consistent with differences in adjustment costs: beliefs and flexible household behavior respond, whereas leave intentions do not. The two interventions differ in content, messenger, delivery, and intensity, and the training bundles several components. Our design therefore does not identify which component produced the household response, whether changes in beliefs caused behavior, or whether fathers’ additional childcare caused spouses’ additional paid hours.

Because the household outcomes are self-reported, we also examine whether socially desirable reporting can account for the results. We find no evidence that treatment effects are larger among respondents with higher scores on a standard Social Desirability Scale. The pattern of effects is also selective rather than pervasive: in households with young children, spouses’ additional paid hours coincide with lower housework, while weekday time use and several other outcomes that could plausibly be subject to favorable reporting do not change. Taken together, these patterns provide little evidence that the main results reflect a general tendency to report treatment-consistent behavior.

Our first contribution is to the literature on policies that alter the division of paid and unpaid work within households. Paternity-leave reforms show that fathers’ domestic time and mothers’ market work can move together, although effects vary substantially across settings (Patnaik, 2019; Farré and González, 2019; Ekberg et al., 2013; Kluve and Tamm, 2013; Andresen and Nix, 2026). Related natural-experiment evidence shows that one partner’s work arrangements can also affect the other partner’s labor supply (Feuillade et al., 2025). Our departure is the lever. In this experiment, a two-hour program delivered by the employer changes neither leave entitlements nor formal job arrangements, yet household outcomes respond within one to two months. The results therefore show that employers can influence household allocation through a substantially lighter-touch instrument than the policy and work-arrangement changes emphasized in the existing literature.

Second, we contribute to the literature on randomized workplace interventions for gender equality. Our intervention targets a different margin from most of this literature: men’s behavior as fathers. Existing workplace experiments have primarily sought to improve women’s opportunities, evaluation, promotion, mentoring, or work arrangements, or have changed employees’ work-family conditions more generally (Bertrand and Duflo, 2017; Azmat and Boring, 2020). Some of these interventions affect employees’ own household time: flexible work arrangements increased men’s and women’s household and care

activities in an Italian firm (Angelici and Profeta, 2024), while the Work, Family, and Health Network intervention increased mothers’ time with children (Kelly et al., 2014; Davis et al., 2015). Studies that followed partners have mainly examined stress, family-time adequacy, and schedule flexibility rather than partners’ paid work (Lee et al., 2019). We instead intervene directly on men’s role at home and measure whether the response extends to the untreated spouse. To our knowledge, ours is the first employer-delivered randomized intervention designed specifically for men as fathers whose measured outcomes extend into household behavior, including the spouse’s paid work.¹

Finally, our information arm relates to the literature on misperceived norms. Bursztyn et al. (2020) show that correcting second-order beliefs about gender attitudes can change behavior. In contemporaneous and independently developed work in Japan, Bertrand et al. (2026) show that men underestimate workplace support for paternity leave and that the persistence of belief correction depends on how the information is presented. Their plain-text treatment is the closest analogue to ours. Our treatment instead uses statistics from participants’ own organizations and modestly shifts perceived support but not leave intentions or information-seeking. The muted response may reflect limited persistence of statistics-only information or less room to update given our participants’ relatively favorable baseline beliefs; our design cannot distinguish these explanations. The comparison therefore helps interpret the information result without implying that organization-specific information is inherently more or less effective.

The remainder of the paper is organized as follows. Section 2 describes the institutional background. Section 3 presents the research design and interventions. Section 4 describes the data and baseline balance. Section 5 reports the main results, and Section 6 presents robustness checks and additional analyses. Section 7 concludes.

2 Institutional Background

Japan exhibits pronounced gender disparities despite high female labor-force participation. In the labor market, women earn about 22 percent less than men and hold only around 13 percent of managerial positions. The disparities are even more pronounced within the household: women spend more than four times as much time on unpaid care work as men. Yet total hours of paid and unpaid work are similar across genders. A central feature of the Japanese setting is therefore not a large gender difference in total work, but a sharply gendered allocation of paid and unpaid work, with women bearing most

¹Nearby designs differ along important dimensions. An employer-hosted experiment at two UK banks provided male employees with information about coworkers’ support for leave but measured stated leave intentions rather than household outcomes (Davidson et al., 2021). Chu et al. (2025) study a father-targeted employer leave mandate using a difference-in-differences design and measure leave enrollment rather than household behavior. A systematic review of workplace interventions found no randomized study targeting fathers alone (Suto et al., 2022).

household responsibilities and men concentrating more heavily on market work (OECD, 2023).

This pattern persists despite generous statutory support for fathers’ leave-taking. Eligible fathers can take up to twelve months of childcare leave with substantial income replacement. Yet during our study period,² only about 30 percent of eligible fathers took paternity leave (Ministry of Health, Labour and Welfare, 2024), far below the take-up observed following the introduction of dedicated father quotas in Norway, Sweden, and Quebec (Dahl et al., 2014; Ekberg et al., 2013; Patnaik, 2019). The coexistence of generous formal rights and limited take-up suggests that barriers beyond the availability of statutory leave remain important for fathers’ use of these entitlements.

Persistent social and workplace norms are one plausible source of this gap. Traditional gender attitudes remain common in Japan: more than one quarter of survey respondents agree that maternal employment harms children (Haerpfer et al., 2020). Workplace expectations may matter more directly for fathers’ own decisions. Among fathers who forgo paternity leave, a substantial share identify an unsupportive workplace atmosphere as an important deterrent (Mitsubishi UFJ Research and Consulting, 2018). These patterns suggest an implementation gap between formal entitlements and actual behavior, in which social expectations and perceived workplace costs may continue to discourage fathers from using leave. They also make the workplace a potentially important margin for changing fathers’ behavior even when statutory entitlements are already generous.

A second feature of the Japanese context is important for understanding how changes in paternal involvement can translate into household time allocation. Formal childcare is widely available: nearly all children aged three to five attend licensed facilities, and care for younger children is also heavily subsidized (OECD, 2024). At the same time, childcare provision is concentrated in standard weekday hours. Only 6.5 percent of centers offer weekend care, and among centers offering extended hours, 83.6 percent provide less than two additional hours per day (National Council of Nursery Schools, 2022). Flexible alternatives outside formal childcare centers are also rarely used. Only 0.4 percent of two-year-old children receive care from babysitters (Ministry of Health, Labour and Welfare, 2012), and just 1.8 percent of households report using paid home-support services (Ministry of Economy, Trade and Industry, 2024). Thus, although formal childcare coverage is extensive, families have limited options for flexibly substituting outside care for parental time, leaving relatively few flexible margins for reallocating childcare within the household.

Household production beyond childcare is also intensive and strongly gendered in

²Japan’s childcare leave system is formally gender-neutral; following the literature, we use “paternity leave” to refer to fathers’ take-up of childcare leave. Our study period (August 2023 to March 2024) follows the October 2022 introduction of postnatal paternity leave (*shussei-ji ikuji kyugyo*), which raised the national salience of men’s leave-taking. Treatment and control offices operate under the same policy environment, so this backdrop does not affect the experimental comparisons.

Japan. For example, 55.5 percent of Japanese women cook and wash dishes at least three times per day, compared with 26.0 percent of women in the United States (Japan Institute for Labour Policy and Training, 2017). These patterns are consistent with demanding norms surrounding routine household production and with women continuing to bear a disproportionate share of that work.³ Greater paternal participation in domestic work therefore has the potential to relax spouses’ time constraints even when the change in fathers’ own time use is modest. Together, these institutional features leave substantial scope for changes in fathers’ involvement at home to alter the allocation of paid and unpaid work within the household.

3 Research Design

3.1 Overview

We study two randomized workplace interventions combined with a three-wave survey across four Japanese organizations. The interventions target different potential barriers to paternal involvement: a two-hour work-life balance training program and a lighter-touch information intervention. Figure 1 summarizes the experimental design and survey timeline. Our main analysis uses 1,225 male employees who completed both the baseline and first follow-up surveys (Section 3.4).

The training was randomized at the office level and offered to male employees and, in parallel sessions, to their managers. It combined content on fathers’ involvement at home and paternity leave with explicit employer sponsorship and managerial participation. The information intervention was randomized at the individual level among non-managers and corrected misperceptions about coworkers’ and managers’ attitudes toward paternity leave. The information arm therefore provides a narrower intervention that targets perceived workplace norms without the broader content and organizational framing of the training.

The experiment consisted of a baseline survey and two follow-up surveys conducted between August 2023 and March 2024. Offices assigned to the early-training group received the training between the baseline and first follow-up surveys, while control offices received it between the first and second follow-ups. The information intervention was embedded in the first follow-up survey. The first follow-up therefore provides the experimental contrast for estimating the short-run effect of the training before control offices receive it. We adopted this waitlist design because the participating organizations wanted all employees eventually to receive the training, and delayed delivery to control offices was

³A well-known illustration is Japan’s *kyaraben* culture, in which mothers are expected to prepare visually elaborate lunch boxes styled after cartoon characters rather than rely on simpler or store-bought options (Oi, 2019).

a condition for conducting the study. The resulting trade-off is that we cannot identify longer-run effects by comparing trained employees with employees who are never offered the training.

The trial was registered in the AEA RCT Registry (AEARCTR-0012315). The pre-registration specified five sets of outcomes: willingness to take paternity leave; time spent on childcare, housework, and market work; respondents' own attitudes toward paternal involvement and paternity leave; second-order beliefs about coworkers' and managers' attitudes; and working practices and relationships with coworkers. Our analysis covers all five preregistered sets.

3.2 Recruitment

Japanese firms faced growing regulatory pressure to improve gender-equality metrics, including paternity-leave take-up rates, which likely contributed to the participation of the private firms; municipalities joined because paternity leave was a salient social issue at the time. Against this backdrop, we collaborated with Evidence Studio⁴ to recruit four employers that voluntarily applied to join a program designed to promote paternal involvement in childcare and housework, including paternity-leave uptake. The participating organizations consisted of two private firms and two municipal governments.

Within each organization, employees self-selected into the program by registering through an online form. In total, 1,962 employees signed up across the four organizations.⁵ The organizations comprised 84 branch offices, and registrations came from 80 of them; the remaining four offices had no registrations.

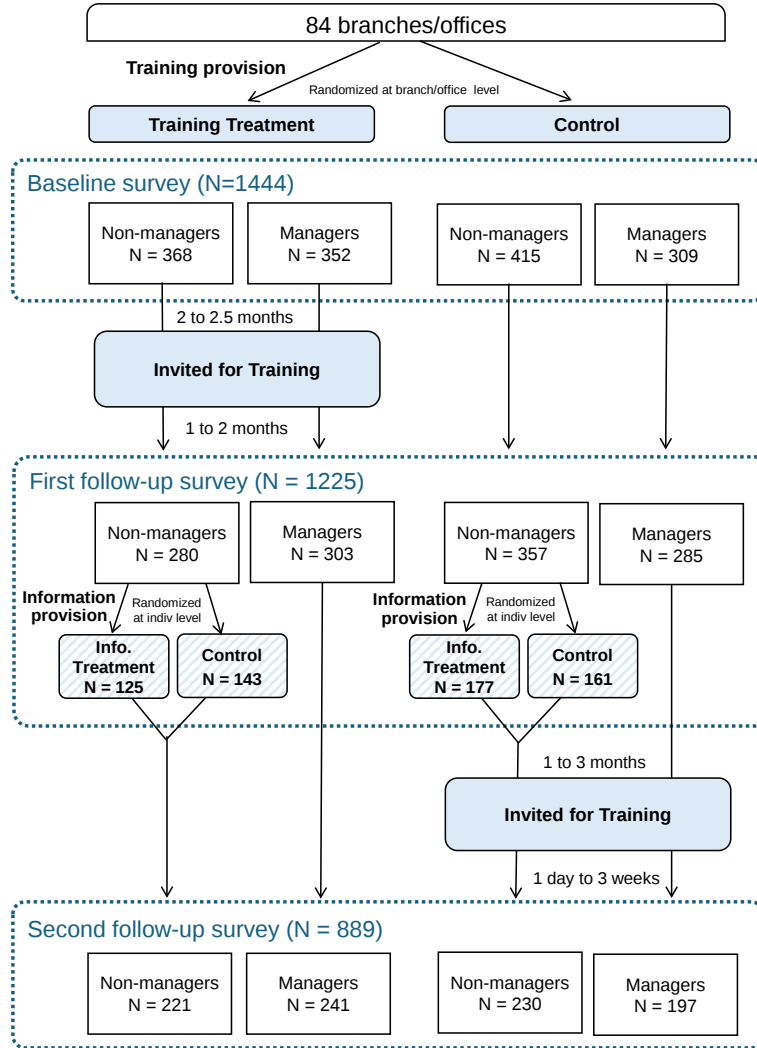
The two municipal governments are located in mid-sized municipalities with populations between 100,000 and 200,000. The two private firms are listed companies in the retail and wholesale industry, headquartered in Tokyo with branch offices distributed nationwide. Participants came from diverse occupational backgrounds and worked across a wide range of functions at both headquarters and regional branches.

The organizations span public and private employers and are not unusual on some observable gender-composition measures. Appendix Table B.1 reports pre-intervention paternity-leave take-up rates, female employee shares, and female manager shares for each organization. The share of female managers, at roughly 10 percent, is close to the national average. Paternity-leave take-up, however, varied substantially across organi-

⁴Evidence Studio (ES) was formerly known as Data for Social Transformation. Established in 2022 by executives from listed Japanese companies and researchers across disciplines, ES promotes evidence-based decision-making in industry and policy. Its programs address social issues including healthcare, welfare, poverty, and education.

⁵HR divisions circulated the invitation to all employees; the invitation letter is reproduced in Appendix A.1. The 1,962 sign-ups represent 12.4 percent of all employees in the four organizations. The invitation was sent to both male and female employees, and the sign-up rate includes female participants, although the registration form did not collect gender information.

Figure 1: Timing of surveys and interventions



Note: The figure summarizes the sequence of surveys and interventions for the training treatment and control groups and reports the number of male respondents with valid entries for the key variables at each wave. The training was randomized at the office level, whereas the information intervention was randomized at the individual level. The information intervention was embedded within the first follow-up survey (FU1) and administered only to non-manager respondents. The key variables include age, marital status, managerial status, presence of at least one child, and assignment to the training intervention. Because only employees on tenured contracts were eligible to participate in the information intervention, the counts of treated and untreated non-managers do not sum to the full non-manager sample.

zations. Among the private firms, one did not disclose its take-up rate, while the other reported a rate of about 20 percent, well below the average of 52.2 percent among publicly listed firms.

3.3 Interventions

3.3.1 WLB Training Program

Program Content The work-life balance (WLB) training program was developed and delivered by NPO Fathering Japan as a two-hour online seminar conducted during regular working hours and funded by the participating employers. Separate sessions were provided for managers and non-managers and offered in parallel. Employees who could not attend in real time were given access to a recorded version for ten days. Viewing the recording within this window counted as participation. The participant list was not visible to attendees, so employees could not observe which of their colleagues took part.

Both sessions followed a common curriculum delivered by male instructors who were successful business professionals and actively involved fathers. The curriculum introduced Japan's paternity-leave policies as part of a broader discussion of fathers' involvement in childcare and emphasized the benefits of greater paternal engagement for families, fathers themselves, and workplaces. For families, it stressed the difficulty of one parent shouldering childcare alone and noted that fathers often overestimate their contributions to housework and childcare relative to their partners' assessments. For fathers, it presented survey evidence that most fathers who had taken leave reported no negative effect on workplace evaluations, discussed strategies for managing evaluation concerns, and emphasized the potential benefits of active childcare involvement for work-life satisfaction. For workplaces, it emphasized that paternal involvement can strengthen communication, teamwork, and shared task coverage as employees coordinate childcare responsibilities within teams. The manager session additionally covered ways to support employees with childcare responsibilities, including communication practices and evaluation systems that account for family commitments. Both sessions concluded with a group discussion. Appendix Table A.1 describes the curriculum in greater detail.

The training could plausibly operate through several channels. First, by addressing concerns about workplace evaluation and presenting paternal involvement as compatible with career commitment, it could reduce perceived career costs of taking leave or becoming more involved at home. Second, the male instructors provided visible examples of successful professionals who were also active fathers, potentially changing beliefs about what is feasible and socially acceptable (Beaman et al., 2012). Third, employer sponsorship, delivery during paid working hours, and parallel participation by managers could provide organizational endorsement for paternal involvement and reduce uncertainty about how the employer would view such behavior. This channel is related to

concerns that employees may avoid leave when they fear falling behind coworkers who remain at work (Johnsen et al., 2024). Because attendees could not observe who else participated, any such effect would operate through the content and organizational framing of the program rather than through an observable participation signal among peers. These elements were bundled in a single intervention, so our design does not separately identify the contribution of any one channel.

Randomization of Training Sessions As described in Section 3.2, the participating organizations comprised multiple offices, corresponding to branches in the private firms and municipalities in the local governments. We randomly assigned offices to either the training treatment group, which received the training earlier, or the control group, which received it later. We used stratified randomization to account for organizational differences. The two municipal governments were grouped into one stratum, all establishments of one private firm formed another stratum, and the establishments of the remaining private firm were divided into two strata to reflect its larger size and internal organizational structure. We randomized at the office level because the training was delivered to groups of employees and because coworkers within an office interact frequently, allowing potential within-office spillovers to remain within assigned units. Spillovers across offices are also possible, but the scope for such contact is limited by the organizational structure: the municipal offices are separate employers in different municipalities, while branches of the private firms are geographically dispersed nationwide. To the extent that residual cross-office spillovers occur, they would attenuate the contrast between treatment and control offices. Office-level sign-up rates were not significantly correlated with treatment assignment, and the main findings are robust to restricting the sample to the organization with the highest sign-up rate (Appendix Table B.2 and Appendix C).

3.3.2 Information Provision

Intervention Content The information intervention targeted a narrower potential barrier to paternity-leave take-up: employees’ misperceptions about how supportive their coworkers and managers were of men taking leave. Baseline responses indicated that non-managers tended to underestimate this support, a pattern documented in Section 4.1. Building on previous studies showing that information about others’ attitudes can change perceived social norms (Bursztyn et al., 2020; Bertrand et al., 2026), we provided participants with accurate information about their colleagues’ and managers’ actual attitudes toward paternity leave. The intervention was designed to correct participants’ second-order beliefs and thereby reduce the perceived social costs of taking leave.

In the baseline survey, respondents and managers reported on a six-point scale how strongly they personally agreed with the statement, “It is a good thing for men to take

leave for childcare and caregiving.”⁶ We also asked respondents how much they thought their coworkers and managers would agree with the same statement, thereby eliciting second-order beliefs about each group. In the first follow-up survey, treated non-managers were shown two charts for each target group, coworkers and managers. One chart displayed the distribution of non-managers’ beliefs about that group’s attitudes, while the other displayed the group’s actual baseline survey responses, making the discrepancy between perceived and reported attitudes visible. A distinctive feature of the intervention was that these statistics came from participants’ own colleagues and managers within the same organization rather than from society at large. The treatment screen explicitly identified this source, making the information directly relevant to the workplace in which participants would decide whether to take leave. Appendix A.4 reproduces the treatment screens.

Randomization The information intervention was implemented only among non-managers. Within each organization-level stratum, we randomly assigned half of the non-managers to receive the information and half to the control group. We randomized at the individual rather than office level because the intervention was delivered privately within the online survey and involved no group component. Although treated respondents could subsequently share the information with coworkers, the belief outcomes were measured during the same survey session, before such communication could occur. The one outcome measured later, whether participants subsequently researched paternity-leave policies, could potentially be affected by later conversations between treated and untreated employees; any such spillover would tend to attenuate the estimated treatment effect.

3.4 Survey Waves and Sample

To reduce concerns about experimenter demand and protect confidentiality, responsibilities were separated across the organizations involved in the study. The training was delivered by an external nonprofit, data collection was coordinated by the managing NPO, and all analyses were conducted independently by the researchers. HR divisions had no access to individual survey responses, while the researchers had no access to respondents’ identities. The managing NPO linked training registration records to survey responses while keeping personally identifiable information confidential. This arrangement was explained to participants at the beginning of each survey, and Appendix A.3 presents the informed-consent form.

Approximately two months after the baseline survey, the training was delivered to offices assigned to the early-treatment group, and the first follow-up survey was fielded one

⁶The wording was drawn from standardized questions used in public opinion surveys, including the Cabinet Office Opinion Poll.

to two months later. Control offices received the training one to three months after the first follow-up, followed by the second follow-up survey within three weeks. The baseline and first follow-up surveys collected measures of workplace environments, including psychological safety; behavioral outcomes, including household time allocation; respondents' own attitudes; and second-order beliefs about coworkers' and managers' attitudes. The baseline survey also collected demographic characteristics. The second follow-up provided supplementary measures, including a Japanese adaptation of the Social Desirability Scale (SDS) (Kitamura and Suzuki, 1986),⁷ which we use in Section 6.1 to assess whether the results reflect differential socially desirable reporting, as well as one later outcome for the information intervention.

Survey links were distributed to all 1,962 employees who had signed up for the program, of whom 1,771, or 90.3 percent, completed the baseline survey. Among the 1,444 male baseline respondents, 1,225, or 84.8 percent, also completed the first follow-up, implying attrition of 15.2 percent between the two waves that define our main analysis. The second follow-up was completed by 889 men, or 61.6 percent of male baseline respondents.⁸ Office-level response rates were not significantly correlated with training assignment (Appendix Table B.2), and Appendix D examines the implications of individual-level attrition for the results.

Our main analytic sample consists of male employees who completed both the baseline and first follow-up surveys and provided the demographic and treatment information required for the analysis, including age, marital status, managerial status, parental status, and training assignment. These restrictions yield 1,225 employees across 74 offices. Of the 80 offices with program registrations, five had no male respondent to the baseline survey, leaving 75 offices represented among male baseline respondents; one additional office contributed no employee to the final analytic sample.

3.5 Econometric Specification

We estimate the causal effects of the training and information interventions using specifications tailored to their distinct randomization and implementation structures. For both interventions, we use an analysis-of-covariance (ANCOVA) specification that controls for the baseline value of the outcome when available, improving statistical precision (McKenzie, 2012). Additional baseline covariates are selected using post-double-selection (PDS) LASSO (Belloni et al., 2014).

⁷Details are provided in Appendix Table C.1.

⁸Measured against all sign-ups rather than against baseline respondents, response rates were 75.6 percent at the first follow-up ($N = 1,483$) and 55.0 percent at the second ($N = 1,080$). All respondents who completed a follow-up survey also completed the baseline survey.

Training Intervention

For the training intervention, our primary estimand is the intent-to-treat (ITT) effect of assignment to the early-training group. The ITT is identified by the office-level randomization and does not depend on whether an assigned employee actually participated. It also corresponds closely to the policy choice faced by an employer: the program was offered on an opt-in basis, so the ITT measures the effect of offering employer-sponsored training under the recruitment and participation conditions studied here. We report instrumental-variable (IV) estimates based on actual participation in the appendix as a secondary estimand (Appendix Table B.3).

We estimate

$$Y_{ij} = \beta \text{Training}_j + \theta' X_{ij} + \epsilon_{ij}, \quad (1)$$

where Y_{ij} is the outcome for employee i in office j , measured in the first follow-up survey, and Training_j equals one if office j was assigned to the early-training group. The vector X_{ij} includes the baseline value of the outcome, randomization-strata fixed effects, and additional controls selected by PDS LASSO. The candidate control set consists of age, age squared, managerial status, marital status, parental status, information-treatment status, and all interactions among these variables.⁹ Standard errors are clustered at the office level.

Among male employees in early-training offices who had signed up for the program, 49 percent participated before the first follow-up survey; no employee in the later-training group received the training before that survey. Participation, defined as attending the live session or viewing the recording within the permitted window, is self-reported because the organizers did not retain individual attendance or archive-viewing records that allow administrative verification. The appendix IV estimates instrument participation with training assignment. Interpreting these estimates as the causal effect of an employee's own participation would require assignment to affect his outcomes only through his own participation. Although participant lists were not visible to coworkers, managers in early-training offices received parallel training and could have influenced subordinates who did not participate. We therefore cannot rule out office-level channels that violate this exclusion restriction. For this reason, we treat the ITT as the primary estimand and interpret the IV estimates more cautiously as the effect of the training as experienced in an assigned office, scaled by participation, rather than as the effect of an employee's own attendance in isolation.

⁹We confirm that including additional outcome variables (such as beliefs and working hours) measured in the baseline survey in the LASSO control set does not change the qualitative results.

Information Intervention

For the information intervention, which was randomized individually among non-managers within the first follow-up survey, we estimate the average treatment effect (ATE). Because assigned respondents were exposed to the information directly within the survey, compliance was effectively universal and the ATE coincides with the ITT. We estimate

$$Y_{ij} = \beta_1 \text{Information}_{ij} + \theta' X_{ij} + \epsilon_{ij}, \quad (2)$$

where Information_{ij} equals one if respondent i in office j received the information treatment. The specification includes the baseline value of the outcome, firm-group fixed effects, and controls selected using PDS LASSO. The candidate control set consists of age, age squared, marital status, parental status, training-assignment status, and all interactions among these variables. Standard errors are clustered at the office level.

Because the information was delivered within the first follow-up survey, our primary outcomes are its immediate effects on beliefs and intentions measured in that survey. We also examine whether respondents subsequently researched paternity-leave policies, measured in the second follow-up. By then, the later-training offices had also received the training, so this outcome identifies the incremental effect of the information intervention in an environment in which both training groups had been offered the WLB program. To examine whether the two interventions were complementary at the first follow-up, we also estimate a specification that adds training assignment and its interaction with information-treatment status (Appendix Table B.20).

Multiple Hypothesis Testing

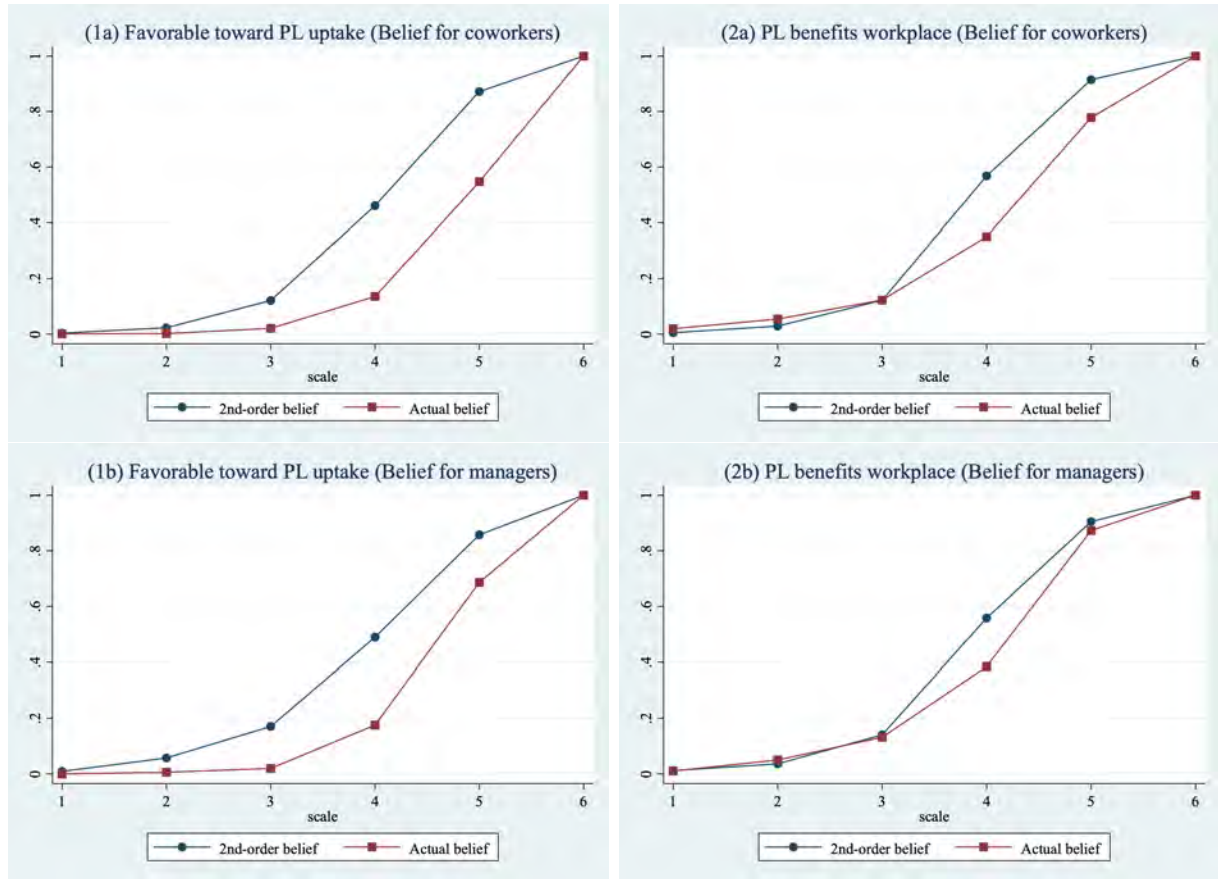
Because each intervention is evaluated on multiple preregistered outcomes, we adjust for multiple hypothesis testing using the sharpened two-stage q -values of Anderson (2008). A q -value is the false-discovery-rate analogue of a p -value: rejecting all hypotheses with q -values at or below 0.10 implies an expected false discovery rate of 10 percent among the rejected hypotheses. We compute q -values within families defined by the economic constructs being measured, distinguishing market work from home production, respondents' outcomes from spouses' outcomes, and first-order from second-order beliefs. The notes to each table specify the relevant family partition. We report q -values in brackets in tables and in parentheses in the text; no q -value is reported when a family contains only a single outcome.

4 Descriptive Statistics

4.1 Evidence for Pluralistic Ignorance

We use the baseline survey to measure both respondents' own attitudes toward paternity leave and their beliefs about the attitudes of coworkers and managers. Pluralistic ignorance arises when individuals misperceive the prevailing attitudes of others, even when their own private views are broadly similar. We focus on two statements. The first, "It is a good thing for men to take leave for childcare and caregiving," captures general support for men's paternity-leave uptake and is referred to as "Favorable toward PL uptake," where PL denotes paternity leave. The second, "Increasing men's take-up of paternity leave can improve company operations and promote a collaborative workplace culture," captures beliefs about whether paternity leave benefits the workplace. Respondents rated their own agreement with each statement on a six-point scale and also reported how strongly they believed their coworkers and managers would agree, giving us corresponding second-order beliefs. The distinction between the two statements is useful for interpreting the treatment effects in Section 5: the first concerns leave-taking in general, whereas the second explicitly links paternal leave to workplace performance.

Figure 2: Cumulative Distribution: Evidence for pluralistic ignorance in baseline survey



Notes. This figure compares cumulative distributions of second-order beliefs and corresponding actual beliefs of coworkers and managers, both measured in the baseline survey. Panels (1a) and (1b) refer to the statement “Favorable toward PL uptake”, and Panels (2a) and (2b) to “PL benefits the workplace”. In Panels (1a) and (2a), showing beliefs for coworkers, the cumulative distribution of actual beliefs is calculated using the sample of male non-managers. In Panels (1b) and (2b), showing beliefs for managers, the cumulative distribution of actual beliefs is calculated using the sample of male managers. In all panels, the cumulative distributions of second-order beliefs are calculated using all male sample.

Figure 2 compares the distributions of these second-order beliefs with the corresponding actual beliefs reported by coworkers and managers in the baseline survey. For “Favorable toward PL uptake,” the distribution of actual beliefs first-order stochastically dominates the distribution of second-order beliefs for both coworkers and managers, indicating that respondents systematically underestimate others’ support for men taking paternity leave. For “PL benefits the workplace,” the same ordering holds over the upper half of the response scale, where most observations lie, although the distributions nearly coincide and cross among the relatively few respondents in the lowest response categories.

The misperception is larger for general approval of paternity-leave uptake than for beliefs about its workplace benefits. The gaps are also broadly similar for coworkers and managers, suggesting that respondents are not systematically more pessimistic about one reference group than the other. Overall, the baseline evidence is consistent with pluralistic ignorance: employees generally underestimate how supportive their peers and supervisors are of men’s paternity leave.

Table 1: Descriptive statistics and balance test for training intervention

Variable	(1)	(2)	(3)	(4)
	Mean	Mean	Difference	No. Obs
	Control	Treatment	Treatment-Control	
Age	45.488 (10.453)	44.175 (9.820)	-1.327* (0.083)	1,225
Married	0.850 (0.357)	0.870 (0.337)	0.019 (0.366)	1,225
Manager	0.444 (0.497)	0.520 (0.500)	0.076** (0.043)	1,225
Has at least one child	0.768 (0.422)	0.768 (0.422)	0.000 (0.988)	1,225
Own belief: Favorable toward PL uptake	5.170 (0.770)	5.225 (0.771)	0.054 (0.248)	1,225
Own belief: PL benefits workplace	4.565 (1.079)	4.530 (1.082)	-0.039 (0.548)	1,225
2nd-order belief: Coworkers' favorable attitudes toward PL uptake	4.500 (0.958)	4.534 (0.901)	0.033 (0.660)	1,223
2nd-order belief: Managers' favorable attitudes toward PL uptake	4.445 (1.111)	4.432 (1.041)	-0.015 (0.851)	1,223
2nd-order belief: Coworkers' belief that PL benefits workplace	4.375 (0.893)	4.284 (0.902)	-0.093 (0.120)	1,221
2nd-order belief: Managers' belief that PL benefits workplace	4.371 (0.969)	4.270 (0.969)	-0.104 (0.208)	1,223
Weekly working hours	44.801 (15.319)	45.430 (13.566)	0.608 (0.495)	1,208
Daily childcare hours (weekday)	1.308 (1.567)	1.537 (1.696)	0.238* (0.050)	910
Daily childcare hours (weekend)	6.333 (6.007)	6.727 (5.746)	0.398 (0.339)	910
Daily housework hours (weekday)	0.945 (1.256)	0.878 (0.943)	-0.063 (0.214)	1,196
Daily housework hours (weekend)	2.117 (2.289)	1.965 (1.898)	-0.151 (0.275)	1,196
Spouse's weekly working hours	21.542 (18.075)	21.898 (17.289)	0.388 (0.706)	1,033
Spouse's daily childcare hours (weekday)	4.888 (4.691)	5.255 (4.628)	0.354 (0.314)	881
Spouse's daily childcare hours (weekend)	7.736 (6.008)	7.987 (5.779)	0.247 (0.462)	881
Spouse's daily housework hours (weekday)	4.202 (3.746)	3.989 (3.231)	-0.221 (0.398)	1,020
Spouse's daily housework hours (weekend)	4.169 (3.625)	3.895 (3.013)	-0.282 (0.230)	1,020
Researched about paternity leave	0.184 (0.388)	0.174 (0.379)	-0.011 (0.686)	1,222
Intention to take paternity leave	0.184 (0.387)	0.183 (0.387)	-0.001 (0.984)	1,184
Received information treatment	0.520 (0.500)	0.462 (0.500)	-0.056* (0.099)	589

Notes. The table reports characteristics measured in the baseline survey together with assignment to the information intervention. Columns (1) and (2) report the mean and standard deviation (in parentheses) by training treatment assignment. Column (3) reports the difference between treatment and control groups estimated from a regression of each variable on a treatment indicator, controlling for randomization-strata fixed effects. The values in parentheses in Column (3) are p-values based on standard errors clustered at the randomization-unit level. Column (4) reports the number of observations available for each variable. The last row, "Received information treatment", is defined only for non-managers, who alone were eligible for the information intervention, and is therefore based on a smaller sample.

Table 2: Descriptive statistics and balance test for information intervention

Variable	(1)	(2)	(3)	(4)
	Control	Treatment	Treatment-Control	No. Obs
Age	41.473 (11.581)	40.076 (10.901)	-1.286 (0.259)	589
Married	0.812 (0.391)	0.770 (0.422)	-0.042 (0.144)	589
Has at least one child	0.691 (0.463)	0.653 (0.477)	-0.040 (0.196)	589
Own belief: Favorable toward PL uptake	5.272 (0.814)	5.306 (0.728)	0.032 (0.641)	589
Own belief: PL benefits workplace	4.594 (1.079)	4.512 (1.224)	-0.088 (0.344)	589
2nd-order belief: Coworkers' favorable attitudes toward PL uptake	4.529 (0.926)	4.632 (1.027)	0.091 (0.274)	588
2nd-order belief: Managers' favorable attitudes toward PL uptake	4.487 (1.083)	4.414 (1.156)	-0.092 (0.318)	588
2nd-order belief: Coworkers' belief that PL benefits workplace	4.331 (0.851)	4.334 (1.013)	-0.006 (0.942)	586
2nd-order belief: Managers' belief that PL benefits workplace	4.293 (0.965)	4.275 (1.044)	-0.031 (0.680)	588
Researched about paternity leave	0.161 (0.368)	0.200 (0.401)	0.038 (0.139)	588
Intention to take paternity leave	0.251 (0.434)	0.325 (0.469)	0.074* (0.079)	570
Assigned to training intervention	0.480 (0.500)	0.423 (0.495)	-0.055* (0.093)	589

Notes. The table reports characteristics measured in the baseline survey together with assignment to the training intervention. Columns (1) and (2) report the mean and standard deviation (in parentheses) by information treatment assignment. Column (3) reports the difference between treatment and control groups estimated from a regression of each variable on a treatment indicator, controlling for firm group fixed effects. The values in parentheses in Column (3) are p-values based on standard errors clustered at the office level. Column (4) reports the number of observations available for each variable.

4.2 Balance Check

Tables 1 and 2 compare characteristics across treatment and control groups for the training and information interventions, respectively. The tables report baseline characteristics together with assignment to the other intervention. Differences between treatment and control groups are estimated using the same fixed effects as in the main specifications: randomization-strata fixed effects for the training intervention and firm-group fixed effects for the information intervention.

For the training intervention, treatment and control groups are broadly similar across most baseline characteristics, including marital and parental status, working hours, and own and second-order beliefs. Several differences are nevertheless visible. Treated employees are slightly younger on average (44.2 versus 45.5 years, $p < 0.10$), have a higher manager share (0.520 versus 0.444, $p < 0.05$), and report somewhat more weekday child-care time (1.54 versus 1.31 hours, $p < 0.10$). The manager-share difference is smaller and statistically insignificant in the full baseline sample before attrition to the first follow-up (6.5 percentage points, $p = 0.103$), indicating that part of the imbalance in the analytic sample arises through attrition. Appendix D examines this issue in detail. Among non-managers, assignment to the information intervention is also somewhat less common in the training group (0.462 versus 0.520, $p < 0.10$). Our main specification adjusts for base-

line covariates and includes information-treatment status in the PDS LASSO candidate set.

For the information intervention, treatment and control groups are again broadly similar in demographics and baseline beliefs. Two modest differences emerge: the treatment group reports a somewhat higher baseline intention to take paternity leave (0.325 versus 0.251, $p < 0.10$) and is somewhat less likely to have been assigned to the training intervention (0.423 versus 0.480, $p < 0.10$). The main specification accounts for these differences through baseline outcome controls and by including training-assignment status in the PDS LASSO candidate set. Taken together, the two tables show no broad pattern of systematic imbalance, while documenting several chance differences that are addressed in the estimation.

5 Results

This section presents the effects of the two interventions on employees’ beliefs and behaviors. Section 5.1 reports the results for the training intervention, and Section 5.2 reports those for the information intervention. We present the preregistered outcomes in the order implied by the adjustment-cost framework in Section 1, from beliefs to household time use to the leave decision, rather than in registry order.

5.1 Effects of Training Intervention

We present the effects of the training intervention in increasing order of adjustment costs, beginning with beliefs, then turning to household time allocation, information seeking and leave intentions, and finally broader workplace outcomes. Throughout this section, the primary estimand is the intent-to-treat (ITT) effect of training assignment. Corresponding instrumental-variable estimates based on training participation are reported in the appendix for reference.

5.1.1 Beliefs

Table 3 reports the effects of training assignment on respondents’ own beliefs and their second-order beliefs about coworkers’ and managers’ attitudes toward paternity leave.¹⁰ We focus on the two statements defined in Section 4.1: general approval of men taking paternity leave (“Favorable toward PL uptake”) and the belief that greater uptake can improve company operations and promote a collaborative workplace culture (“PL benefits the workplace”). The training had little effect on respondents’ already high general approval of paternity-leave uptake, increasing agreement by 0.03 points on the six-point

¹⁰Appendix Table B.4 reports estimates without PDS LASSO-selected controls.

Table 3: Effects of training on beliefs

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)	
	(1) Favorable toward PL uptake	(2) PL benefits workplace	(3) Favorable toward PL uptake	(4) PL benefits workplace	(5) Favorable toward PL uptake	(6) PL benefits workplace
Panel A: ITT						
Training assignment	0.031 (0.023)	0.182*** (0.048)	0.062 (0.060)	0.119* (0.061)	-0.036 (0.044)	0.040 (0.048)
Anderson q-value	[0.096]	[0.001]	[0.460]	[0.253]	[0.460]	[0.460]
Observations	1225	1223	1206	1208	1218	1218
Control Mean	5.265	4.591	4.594	4.547	4.585	4.567
No. clusters	74	74	74	74	74	74
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Test of distribution equality (Wilcoxon rank-sum test)						
p-value	0.122	0.005	0.101	0.030	0.456	0.845
Observations	1225	1223	1208	1211	1220	1220

Notes. This table shows the effect of the training intervention on own beliefs of the male respondents and their second-order beliefs about coworkers' and managers' attitudes. In Panel A, the coefficients are obtained by regressing the outcome variables on an indicator for assignment of the office to the early training group, and they represent the intent-to-treat (ITT) effect of training assignment. Panel B reports p -values from Wilcoxon rank-sum tests comparing the belief distributions between the training treatment and control groups. In Columns (1) and (2), the dependent variable is the respondent's level of agreement with two indicated statements regarding men's uptake of paternity leave. In Columns (3)-(6), the dependent variable is the respondent's second-order belief about coworkers (Columns (3) and (4)) or managers (Columns (5) and (6)). Anderson (2008) sharpened q -values, reported in brackets, are computed within families defined by first-order and second-order beliefs. Standard errors in parentheses are clustered at the randomization-unit level. All equations control for randomization strata fixed effects, baseline value of the outcome variable, and control variables selected by PDS LASSO.

scale (Column 1; $q = 0.096$). In contrast, it increased agreement that paternity leave benefits the workplace by 0.18 points (Column 2; $q = 0.001$). The clearest effect on respondents' own beliefs is therefore on the dimension that explicitly links paternal leave to workplace performance.

The effects on second-order beliefs are weaker. The point estimate for beliefs about coworkers' support for the workplace benefits of leave is positive at 0.12 points, but it is not statistically significant after the multiple-testing correction (Column 4; $q = 0.253$). Estimates for beliefs about managers are also small and statistically insignificant. Panel B reports Wilcoxon rank-sum tests comparing the full belief distributions across assignment groups. These tests indicate distributional differences for respondents' own belief that paternity leave benefits the workplace and for the corresponding second-order belief about coworkers, consistent with the regression results; Appendix Figure B.1 shows the cumulative distributions. If the apparent coworker-manager difference is meaningful, one possible explanation is that supervisors' views are harder to infer because interactions with managers are typically less frequent and more formal than interactions with peers.

The pattern across belief outcomes closely matches the content of the intervention. The curriculum directly emphasized how fathers' involvement and paternity leave could improve workplace functioning through communication, shared task coverage, and broader skill development within teams. General approval of leave-taking was already near the top of the scale and moved little. Nor did the training provide participants with direct

information about coworkers’ or managers’ attitudes, and second-order beliefs changed much less. Thus, the strongest belief response occurs precisely on the dimension that the curriculum addressed most directly: whether paternity leave is compatible with, and potentially beneficial for, how work gets done. Appendix Tables B.5 and B.6 show no statistically significant heterogeneity in these effects by baseline beliefs, managerial status, or parental status.

5.1.2 Household Time Allocation

We next examine whether the training changed household time allocation. Table 4 reports ITT effects for the full analytic sample, while Table 5 presents exploratory heterogeneity by the presence and age of children, distinguishing households with at least one child aged five or below, those with only older children, and those without children.¹¹ The corresponding IV estimates are reported in Appendix Tables B.7 and B.9. All measures of spouses’ time use are reported by the male respondent rather than by the spouse directly.

In the full sample, the training increased fathers’ weekend childcare time by 0.37 hours per day, although the estimate does not remain statistically significant after the multiple-testing correction (Table 4, Panel A; $q = 0.264$). The response is substantially larger among households with at least one child aged five or below, the group for whom the training content was most directly relevant. In these households, fathers increased weekend childcare by about one hour per day, roughly 10 percent of the control-group mean, and the effect remains significant after correction for multiple testing (Table 5, Panel A, Column 3; $q = 0.044$). We find no corresponding increase in weekday childcare, and working hours are also largely unchanged. The concentration of the effect on weekends is consistent with tighter weekday scheduling constraints for full-time employees, reinforced by the limited temporal flexibility of formal childcare described in Section 2.

Spouses’ paid work also responded. In the full sample, spouses’ weekly working hours increased by 1.47 hours, about 7 percent relative to the control mean (Table 4, Panel B, Column 1). Appendix Table B.11 indicates that this increase occurred primarily along the intensive margin among spouses already working: the shares working more than 10 and more than 20 hours per week increased by approximately 4 and 6 percentage points, respectively, while the shares working any positive hours or more than 30 hours were essentially unchanged. The pattern therefore points to an expansion of hours among existing workers rather than detectable labor-force entry.

Among households with young children, the point estimates also indicate sizable reductions in spouses’ housework on both weekdays and weekends, although these estimates do not survive the multiple-testing correction at the 10 percent false-discovery-rate thresh-

¹¹Appendix Table B.8 reports estimates without PDS LASSO-selected controls.

Table 4: Effects of Training on Behaviors

	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)
<i>Panel A: Own hours</i>					
Training assignment	-0.325 (0.712)	-0.045 (0.089)	0.368* (0.190)	0.004 (0.046)	0.049 (0.103)
Anderson q-value		[1.000]	[0.264]	[1.000]	[1.000]
Observations	1177	876	876	1161	1161
No. clusters	74	72	72	74	74
Control Mean	43.981	1.386	5.768	1.003	2.251
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Spouse hours</i>					
Training assignment	1.473** (0.664)	-0.002 (0.163)	0.079 (0.188)	-0.213 (0.162)	-0.035 (0.139)
Anderson q-value		[1.000]	[1.000]	[1.000]	[1.000]
Observations	994	846	846	980	980
No. clusters	74	72	72	74	74
Control Mean	20.213	4.779	7.110	4.487	4.182
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes

Notes. This table reports the effects of the training intervention on the behaviors of male respondents and their spouses. The coefficients are obtained by regressing the outcome variables on an indicator for assignment of the office to the early training group, and they represent the intent-to-treat (ITT) effect of training assignment. Panel A shows the results for male respondents, and Panel B shows the results for their spouses. Anderson (2008) sharpened q-values, reported in brackets, are computed within families defined by respondent versus spouse; following the distinction between market work and home production, working hours form a separate family from the childcare and housework outcomes. The four families are Columns (2)–(5) of Panel A, Columns (2)–(5) of Panel B, Column (1) of Panel A, and Column (1) of Panel B. The two working-hour families contain a single outcome each, so no correction applies and no q-value is reported in Column (1) of either panel. Standard errors in parentheses are clustered at the randomization-unit level. All equations control for randomization strata fixed effects, baseline value of the outcome variable, and control variables selected by PDS LASSO.

Table 5: Heterogeneous Effects of Training on Behaviors by Presence and Age of Children (exploratory subgroup analysis)

	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)
<i>Panel A: Own hours</i>					
Training assignment $\times$ Young children	-1.810 (1.234)	-0.327 (0.218)	1.031** (0.403)	-0.135 (0.0988)	-0.399 (0.331)
Anderson q-value		[0.206]	[0.044]	[0.206]	[0.206]
Training assignment $\times$ Only older children	-0.0724 (1.065)	0.101 (0.0974)	0.133 (0.234)	0.155** (0.0616)	0.180 (0.113)
Training assignment $\times$ No children	0.821 (2.664)			0.0650 (0.108)	0.258** (0.118)
Observations	1015	859	859	1001	1001
No. clusters	74	72	72	74	74
Control Mean (Young children)	43.479	2.302	10.652	1.332	3.495
Control Mean (Only older children)	45.195	0.966	3.509	0.726	1.721
Control Mean (No children)	42.000			0.846	1.564
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Spouse hours</i>					
Training assignment $\times$ Young children	1.982* (1.135)	0.140 (0.386)	-0.206 (0.541)	-0.605* (0.338)	-0.936** (0.435)
Anderson q-value		[0.559]	[0.559]	[0.144]	[0.144]
Training assignment $\times$ Only older children	0.714 (0.808)	0.0824 (0.192)	0.177 (0.311)	-0.136 (0.188)	0.259** (0.130)
Training assignment $\times$ No children	4.271 (2.843)			0.227 (0.440)	0.568 (0.372)
Observations	994	846	846	980	980
No. clusters	74	72	72	74	74
Control Mean (Young children)	16.929	8.238	11.628	5.601	5.373
Control Mean (Only older children)	21.078	3.121	4.943	4.336	3.955
Control Mean (No children)	23.297			2.760	2.638
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes

Notes. This table uses the sample of men who were married at the time of the baseline survey. We estimate heterogeneity in the intent-to-treat (ITT) effect by interacting training assignment with indicator variables for the presence and age group of children. “Young children”, “Only older children”, and “No children” are indicator variables equal to 1 if the household has at least one young child (aged 5 or below), only older children (aged 6 or above), or no children, respectively. This heterogeneity analysis is exploratory: Anderson (2008) sharpened q-values, reported in brackets, are computed within the young-children subgroup, for which the training content was directly relevant, while no correction is applied to the older-children-only and no-children subgroups. As in Table 4, working hours form a family separate from the childcare and housework outcomes, and single-outcome working-hour families are not subject to the correction, so no q-value is reported in Column (1). The corrected families are therefore Columns (2)–(5) of Panel A and Columns (2)–(5) of Panel B, in each case restricted to the young-children interaction. “Control Mean” is the mean of the dependent variable among control-group respondents in the indicated subgroup. Because childcare hours are undefined for households without children, the “No children” interaction is omitted in Columns (2) and (3). Standard errors in parentheses are clustered at the randomization-unit level. All equations control for randomization strata fixed effects, baseline value of the outcome variable, and control variables selected by PDS LASSO.

old (Table 5, Panel B; $q = 0.144$ for both). Taken together, the reductions amount to about five hours per week, more than enough to accommodate the roughly two-hour increase in spouses' paid work in this subgroup. Because both outcomes are reported by the husband, this accounting consistency does not eliminate concerns about measurement error, but it shows that the reported changes form a coherent household time budget.

Additional outcomes are consistent with a broader reallocation of household time. Appendix Table B.12 shows that the couple's total childcare time did not decline and that total housework fell substantially among households with young children. Moreover, among training participants, nearly 30 percent reported discussing household responsibilities with family members within three days of the training, and 37 percent reported engaging in housework or childcare during that period (Appendix Figure B.2). These post-training reports are available only for training participants and are therefore descriptive rather than causal, but they indicate that the intervention prompted immediate household-level responses. We cannot distinguish whether the decline in total unpaid work reflects greater efficiency or coordination, greater outsourcing, lower standards for domestic chores, or some combination of these channels. Patterns for households with only older children and those without children are mixed and are reported in Table 5; because these subgroup analyses are exploratory and the training content was less directly relevant to them, we do not emphasize the individual estimates.

5.1.3 Information Seeking and Intention to Take Paternity Leave

We next examine two outcomes more directly connected to formal paternity leave but differing in their degree of commitment: whether respondents sought information about leave policies, a relatively low-cost preparatory action, and whether they intended to take paternity leave in the future, a more consequential and career-relevant margin.

The training substantially increased information seeking. The probability of reporting having researched paternity-leave policies rose by about 6 percentage points in the full sample (Table 6, Column 1; $q = 0.015$) and by about 8 percentage points among respondents aged 50 or younger (Column 2), the age group used for the leave-intention analysis. The effect appears among both non-managers, who may use the policy themselves, and managers, who may have sought information in order to support employees with childcare responsibilities (Appendix Table B.10).

Stated intentions to take paternity leave, however, did not increase. The estimated effect is small and negative and does not survive the multiple-testing correction (Table 6, Column 3; $q = 0.151$). This contrast is informative: the training changed beliefs, increased information seeking, and shifted flexible household time-use margins, but it did not increase stated intentions to make the more discrete and potentially career-relevant choice of taking formal leave. One reason may be that this outcome is relevant only to

respondents who anticipate another child. Many respondents are in their forties and may no longer expect additional births, leaving relatively few respondents for whom a future leave decision is immediately salient.

The underlying response categories nevertheless changed in ways that help interpret the null effect on the binary intention measure. The training reduced the share saying that they would not take leave even if they had a new child, while increasing the share reporting that childrearing was complete or that they did not anticipate another child. One possible interpretation is that the training made the practical relevance of future childcare more salient without changing the binary intention measure among respondents for whom another birth was unlikely. Appendix B.2.3 reports the full category-level results.

The response also varies with workplace replaceability. In the exploratory specification in Table 6, Column 4, the estimated effect on leave intention is positive among respondents with an above-median number of coworkers able to take over their tasks and negative among those with below-median replaceability; the two coefficients differ with $p = 0.038$. This pattern is consistent with the possibility that workplace feasibility conditions whether more supportive attitudes translate into leave-taking intentions. Related evidence from Danish administrative data shows that coworkers' hours and earnings rise when a colleague takes parental leave, with adjustment costs concentrated in firms that have less scope to substitute for the absent worker (Brenøe et al., 2024).

Table 6: Effects of training on information-seeking and intention to take paternity leave

	Researched about PL		Intention to take PL	
	(1)	(2)	(3)	(4)
Training assignment	0.061*** (0.023)	0.075*** (0.022)	-0.023 (0.020)	
Anderson q-value	[0.015]		[0.151]	
Training assignment $\times$ Above-median replaceability				0.042 (0.033)
Training assignment $\times$ Below-median replaceability				-0.055** (0.027)
Observations	1211	796	775	768
No. clusters	74	69	69	69
Control Mean	0.189	0.239	0.251	0.251
PDS LASSO Control	Yes	Yes	Yes	Yes
P-value: coefficient heterogeneity				0.038

Notes. This table reports the effects of the training intervention on respondents' engagement with paternity leave. In Columns (1) and (2), the dependent variable is an indicator equal to 1 if the respondent reported having researched paternity leave policies; Column (1) uses the full sample, and Column (2) restricts the sample to respondents aged 50 or younger. In Columns (3) and (4), the dependent variable is an indicator equal to 1 if the respondent has plans or hopes to take paternity leave at some point in the future or is currently taking it, and the sample is restricted to respondents aged 50 or younger. We estimate the intent-to-treat (ITT) effect of training assignment; in Column (4), assignment is interacted with the indicated variables. "Above-median replaceability" (or "Below-median replaceability") is an indicator variable equal to 1 if the worker has more than (or fewer than or equal to) the median number of coworkers (2) available to take over their tasks within a day, measured in the baseline survey. Anderson (2008) sharpened q-values, reported in brackets, are computed within the family comprising Columns (1) and (3). Standard errors in parentheses are clustered at the randomization-unit level. All equations control for randomization strata fixed effects, baseline value of the outcome variable, and control variables selected by PDS LASSO.

5.1.4 Workplace Practices and Worker Relationships

Finally, we examine whether the intervention affected broader workplace practices and relationships. Appendix Table B.15 reports outcomes including the number of coworkers available to take over respondents' tasks, meeting time, and days worked from home, while Appendix Table B.16 reports psychological safety, coworker and supervisor relationships, and willingness to recommend the employer to a job seeker.¹² The estimates are uniformly small and statistically insignificant. We therefore find no detectable short-run effects on these broader workplace outcomes. These null results also provide useful falsification outcomes for the analysis of socially desirable responding in Section 6.1.

5.2 Effects of Information Intervention

We next examine the information intervention, which was designed to correct employees' misperceptions about coworkers' and managers' attitudes toward men taking paternity leave. Treated respondents were shown aggregate data on their colleagues' and supervisors' actual responses to the statement, "It is a good thing for men to take leave for childcare and caregiving." Relative to the broader training intervention, this light-touch treatment provides a useful benchmark for whether correcting perceived workplace norms alone, without the training's broader persuasive and organizational content, is sufficient to move downstream outcomes.

Table 7 shows that the intervention shifted second-order beliefs in the intended direction. Respondents increased their assessments of coworkers' support by 0.111 points and managers' support by 0.147 points on the six-point scale (Columns 3 and 5), with sharpened q -values of 0.126 for both estimates. The corresponding control means are 4.60 and 4.52. The rank-sum tests, which are not adjusted for multiple testing, also indicate shifts in both the coworker-belief distribution ($p = 0.014$) and the manager-belief distribution ($p = 0.072$). Appendix Figure B.4 presents the cumulative distributions. By contrast, respondents' own first-order belief about the same statement changed little. This is consistent with the intervention's design: it provided information about others' attitudes rather than arguments intended to persuade respondents directly, and respondents' own baseline attitudes were already highly favorable. Beliefs about whether paternity leave benefits the workplace also changed little, consistent with the fact that the information treatment contained no content about that statement.

The shift in second-order beliefs did not translate into changes in the downstream outcomes we measure. Among men aged 50 or younger, stated intentions to take paternity leave were essentially unchanged both immediately after the information was shown and one to three months later. The point estimates are close to zero in both waves: -0.007 relative to a control mean of 0.295 in the first follow-up and -0.010 relative to

¹²See Appendix A for variable descriptions.

0.321 in the second. This null is particularly informative because stating an intention to take leave is a low-cost socially desirable response, yet respondents did not increase their stated intentions even immediately after being shown evidence that their coworkers and managers were supportive. Appendix Table B.17 shows similarly little movement across the full set of response categories. We also find no significant effect on whether respondents subsequently researched paternity-leave policies. By the second follow-up, both early- and later-training offices had been offered the WLB training, so this outcome captures the incremental effect of the information intervention in that common training environment.

Appendix Table B.18 compares the training and information effects on information seeking directly. On an assignment basis, the information ATE is not statistically distinguishable from the training ITT ($p = 0.36$). When the training effect is instead scaled by participation using the IV estimate, the difference is statistically significant ($p = 0.03$). We treat the latter comparison only as a descriptive benchmark: the IV estimate is not the effect of an employee’s own attendance in isolation, and the two interventions differ in content, messenger, timing, and measurement wave. Appendix Table B.19 likewise shows no detectable information-treatment effects on other outcomes measured in the second follow-up, including psychological safety, coworker and supervisor relationships, and willingness to recommend the employer to a job seeker.¹³

We do not estimate effects of the information intervention on household time allocation because the survey design does not provide a valid post-treatment measure. The intervention was delivered at the beginning of the first follow-up survey, whereas the time-use questions in that survey referred to typical behavior over the preceding period. The second follow-up did not re-administer the time-use battery. Thus, for the outcomes we can measure causally, the evidence points to an effect concentrated at the belief margin, with no detectable downstream response in intentions or information seeking. In this setting, correcting misperceptions about colleagues’ and managers’ support was not sufficient on its own to generate behavioral change. This contrast also helps clarify the training results: the information treatment supplied evidence about others’ attitudes and shifted beliefs about those attitudes, whereas it contained none of the broader content concerning the workplace consequences of paternal involvement that the training emphasized. Finally, Appendix Table B.20 provides little evidence of systematic complementarity between the two interventions. Six of the seven interaction terms are small and statistically insignificant; the single exception is a negative interaction for respondents’ own belief that paternity leave benefits the workplace, significant at the 10 percent level. Given that this is one marginally significant coefficient among seven, we do not place substantive weight on it.

¹³See Appendix A for variable descriptions.

Table 7: Effects of information intervention

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)		Intention to take PL		Researched about PL
	(1) Favorable toward PL uptake	(2) PL benefits workplace	(3) Favorable toward PL uptake	(4) PL benefits workplace	(5) Favorable toward PL uptake	(6) PL benefits workplace	(7) Immediate effect	(8) 1-3 months later	(9)
Panel A: ATE									
Information	0.044 (0.046)	-0.007 (0.077)	0.111* (0.058)	0.036 (0.065)	0.147** (0.073)	0.099 (0.061)	-0.007 (0.031)	-0.010 (0.040)	0.027 (0.032)
Anderson q-value	[1.000]	[1.000]	[0.126]	[0.171]	[0.126]	[0.126]			
Observations	589	588	579	579	587	587	425	308	411
Control Mean	5.339	4.678	4.599	4.602	4.517	4.498	0.295	0.321	0.237
No. clusters	65	65	65	65	65	65	57	50	58
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Test of distribution equality (Wilcoxon rank-sum test)									
p-value	0.194	0.985	0.014	0.318	0.072	0.101			
Observations	589	588	580	581	588	588			

Notes. This table shows the effect of the information intervention among non-managerial employees. “first-order belief (Own belief)” is the respondent’s agreement with two indicated statements. “second-order belief” is the respondent’s assessment of how strongly coworkers or managers, respectively, agree with the indicated statement. In Columns (7)–(8), the dependent variable is an indicator for whether the respondent intends to take paternity leave in the future, measured in the first follow-up survey (Column (7)) and in the second follow-up survey (Column (8)). In these estimations, we restrict the sample to respondents aged 50 or younger. In Column (9), the dependent variable is an indicator for whether the respondent researched paternity leave policies, measured in the second follow-up survey. Panel A reports the average treatment effect (ATE) estimated by OLS, and Panel B reports Wilcoxon rank-sum tests comparing the belief distributions between the treatment and control groups. Anderson (2008) sharpened q-values, reported in brackets, are computed within families defined by first-order and second-order beliefs. Standard errors in parentheses are clustered at the level of offices. All equations control for firm group fixed effects, baseline value of the outcome variable, and control variables selected by PDS LASSO.

6 Robustness Checks

6.1 Social Desirability Bias

Because the main outcomes are self-reported, respondents may have adjusted their answers toward what they perceived as socially or organizationally desirable after an employer-sponsored intervention. We assess this concern using several features of the experimental results and survey design.

First, stated intentions to take paternity leave do not increase under either intervention (Tables 6 and 7). This is especially informative for the information treatment, where respondents were shown evidence of supportive attitudes among coworkers and managers immediately before reporting their own leave intentions. If the interventions mainly induced norm-congruent reporting, stated leave intention would be a particularly low-cost margin on which to respond. Instead, the point estimates remain close to zero.

Second, the effects we do observe are differentiated across outcomes rather than appearing as a broad increase in favorable responses. The training increases respondents’ belief that paternity leave benefits the workplace but has little effect on their general approval of leave-taking. The behavioral effects are likewise concentrated in specific household margins, most notably weekend childcare among fathers with young children, while weekday childcare and working hours remain largely unchanged. Nor do we find systematic improvements in broader outcomes that an image-conscious respondent might plausibly report more favorably after an employer-sponsored program, including psychological safety, coworker and supervisor relationships, willingness to recommend the employer, or workplace practice measures (Appendix Tables B.15 and B.16). This pattern is difficult to reconcile with a generic tendency to provide more socially desirable

answers after treatment.

Third, the survey procedures were designed to reduce incentives for image-conscious reporting. Participants were informed that individual responses would remain confidential, would not be shared with supervisors or human-resources departments, and would not be used in personnel evaluations. Data collection and analysis were conducted independently of the employers, and only anonymized and aggregated results were provided to the participating organizations. These procedures were explained in the informed-consent process described in Appendix A.3.

Finally, the second follow-up survey included a ten-item Social Desirability Scale (SDS; see Appendix Table C.1). We define an indicator for respondents with SDS scores above the sample median and re-estimate the main specifications including treatment, High SDS, and their interaction. The interaction terms are generally small and statistically insignificant, while the treatment-effect estimates remain close to the main results, albeit with somewhat larger standard errors (Appendix Tables C.2, C.3, and C.4). We therefore find little evidence that the estimated treatment effects are concentrated among respondents with a greater propensity for socially desirable responding. The SDS was collected only in the second follow-up, however, and therefore postdates the outcomes of the information intervention, so this test applies most cleanly to the training results. More generally, no survey design can fully rule out reporting bias that is itself induced by treatment. Taken together, the evidence suggests that differential socially desirable reporting is unlikely to account for the main findings, although we cannot exclude it entirely.

6.2 Attrition

Attrition from the baseline to the first follow-up survey is 15.2 percent, and assignment to the training group lowers the probability of responding by 7.7 percentage points on average. The assignment-related response gap is more pronounced among older employees. Importantly, however, we find no statistically detectable assignment-related response gap in the groups that generate our main behavioral results: among men with at least one child aged five or below, the estimated gap is -3.8 percentage points (s.e. 3.2), and among men aged 50 or younger it is -5.2 percentage points (s.e. 3.1). Comparing balance before and after attrition also shows that the manager-share difference widens from 6.5 percentage points in the full baseline sample ($p = 0.103$) to 7.6 percentage points in the analytic sample, indicating that part of the observed imbalance arises through differential attrition. Appendix D therefore re-estimates the main results using inverse-probability weights based on estimated response probabilities. The weighted estimates remain close to the corresponding unweighted estimates across the main outcomes. As an additional check, Lee (2009) bounds for the time-use outcomes in households with

young children remain bounded away from zero in the direction of the main estimates: $[0.702, 1.332]$ for fathers' weekend childcare, $[1.830, 3.076]$ for spouses' weekly working hours, and $[-1.119, -0.456]$ for spouses' weekend housework.

Attrition is also relevant for the information intervention because two outcomes are measured in the second follow-up survey. Of the 589 non-managerial men who entered the information randomization in the first follow-up, 419 responded to the second follow-up, implying attrition of 28.9 percent. In this case, however, information-treatment assignment does not significantly predict second-follow-up response, either with or without baseline covariates, and we find no systematic heterogeneity in response by treatment assignment. Reweighting respondents by the inverse of their estimated second-follow-up response probability likewise leaves the later treatment effects small and statistically insignificant: the weighted estimates are -0.022 for leave intention and 0.022 for information seeking. Taken together, these exercises provide little evidence that differential attrition explains either the training results or the later null effects of the information intervention.

6.3 Robustness to Selective Program Sign-up

The overall program sign-up rate was 12.4 percent across the participating organizations, with substantial variation across organizations; the highest organization-level sign-up rate was 44 percent. Because training assignment was randomized only after employees had signed up, selective sign-up does not by itself compromise internal validity for employees who entered the program. It does, however, shape the interpretation and external validity of the estimand: our estimates identify the effect of offering the training among employees who signed up under the recruitment conditions studied here, rather than among the entire workforce. The direction of this selection is ambiguous. Employees who sign up may already be more favorably disposed toward paternal involvement, leaving less room for the intervention to change their attitudes or behavior; alternatively, they may have greater latent demand for change and therefore respond more strongly. The resulting estimand is nevertheless directly relevant to employers considering a similar opt-in program.

Effects also need not scale proportionally under recruitment or participation strategies that induce substantially broader take-up. Employees reached only through stronger promotion, incentives, or mandatory enrollment may differ from those who signed up under the approach studied here, so we cannot infer how the treatment effect would change if coverage were substantially expanded.

As a robustness check, we replicate key analyses using the organization with the highest sign-up rate, 44 percent, where the program reached a substantially larger share of the workforce. The point estimates broadly resemble those in the full sample, although precision is lower because the subsample is considerably smaller. For example, the es-

estimated effect on fathers' weekend childcare is 0.35 hours per day, compared with 0.37 in the full sample, and the estimated increase in spouses' weekly working hours is 1.76 hours, compared with 1.47 in the full sample (Appendix Table C.6). The training effects on beliefs are somewhat larger than in the full sample (Appendix Table C.5), while the information-treatment estimates generally point in similar directions (Appendix Table C.7). These results provide no clear evidence that the main findings are driven by the relatively low sign-up rates in some participating organizations.

7 Conclusion

This study shows that an employer can affect household outcomes even without changing statutory leave rights or employees' formal work arrangements. In a randomized field experiment across four Japanese organizations, a two-hour work-life balance training raised spouses' reported paid hours in the full sample. In an exploratory analysis of households with young children, the training also increased fathers' weekend childcare and was accompanied by sizable declines in spouses' reported housework. A separate information intervention modestly shifted perceived support for paternity leave but did not affect leave intentions or subsequent information-seeking. Together, these results show that workplace interventions can reach beyond the treated employee and affect the allocation of paid and unpaid work within the household.

The findings point to a different margin for work-family policy. Most workplace interventions for gender equality seek to improve women's opportunities or treatment at work. Our intervention instead targets men as fathers and their contribution to unpaid work at home. This distinction matters because gender inequality in the labor market is partly sustained by inequality in household responsibilities. Formal leave rights remain important, but our results show that employers can influence household outcomes even when those rights and employees' job arrangements are held fixed. Because some effects accrue to the untreated spouse, employers may not fully internalize the benefits of such interventions, creating a potential rationale for public support. Our data, however, do not measure firms' private returns or the social value of these household effects.

The pattern across outcomes is consistent with differences in adjustment costs. The light-touch information treatment modestly shifted the second-order beliefs it targeted but did not change leave intentions or information-seeking. The broader training reached beliefs, information-seeking, and flexible household behavior, while stated leave intentions remained unchanged. These comparisons should not be read as identifying a mechanism: the interventions differ in content, messenger, delivery, and intensity, and the training bundles several components. We therefore cannot determine which component generated the household response, whether changes in beliefs caused changes in behavior, or whether fathers' additional childcare caused spouses' additional paid hours. Other interventions

have shown that formal leave-taking can respond to stronger organizational instruments (Chu et al., 2025; Ashraf et al., 2025), so our results do not imply that the leave margin is intrinsically unresponsive.

Two limitations are especially important. First, our follow-up covers only several months, and the crossover design prevents a long-run comparison between trained and never-trained offices. We therefore do not know whether the changes in household behavior persist or whether they ultimately affect employees' careers, spouses' longer-run labor supply, or other family outcomes. Second, the experiment was conducted among employees who voluntarily signed up for the program in four Japanese organizations. The estimated effects describe the offer under the recruitment conditions studied here; effects need not scale proportionally if employers induce substantially broader participation, and the results may differ in settings with different institutions and norms around fatherhood. Establishing the persistence and external validity of employer interventions targeted at fathers is an important direction for future work.

References

- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association* 103(484), 1481–1495.
- Andresen, M. E. and E. Nix (2026). You cannot force me into caregiving: Paternity leave and the child penalty. *The Economic Journal* 136(674), 780–797.
- Angelici, M. and P. Profeta (2024). Smart working: Work flexibility without constraints. *Management Science* 70(3), 1680–1705.
- Ashraf, N., O. Bandiera, V. Minni, and L. Zingales (2025, May). Meaning at work. NBER Working Paper 33843, National Bureau of Economic Research.
- Azmat, G. and A. Boring (2020). Gender diversity in firms. *Oxford Review of Economic Policy* 36(4), 760–782.
- Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science* 335(6068), 582–586.
- Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies* 81(2), 608–650.
- Bertrand, M., P. Cort'es, J. Pan, H. Shigeoka, and M. Yagasaki (2026). Making information stick: Evidence from correcting misperceptions toward paternity leave-taking in Japan. Unpublished manuscript.
- Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. In A. V. Banerjee and E. Duflo (Eds.), *Handbook of Economic Field Experiments*, Volume 1, pp. 309–393. North-Holland.
- Brenøe, A. A., S. Canaan, N. A. Harmon, and H. Royer (2024). Is parental leave costly for firms and coworkers? *Journal of Labor Economics* 42(4), 1135–1174.
- Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review* 110(10), 2997–3029.
- Chu, R., S. Jeon, H. S. Lee, and T. Lee (2025). Sorting of working parents into family-friendly firms. arXiv:2512.22810.

- Cortés, P. and J. Pan (2023). Children and the remaining gender gaps in the labor market. *Journal of Economic Literature* 61(4), 1359–1409.
- Crowne, D. P. and D. Marlowe (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 24, 349–54.
- Dahl, G. B., K. V. Løken, and M. Mogstad (2014). Peer effects in program participation. *American Economic Review* 104(7), 2049–74.
- Davidson, S., R. Hachohen, F. Gesiarz, T. Hardy, A. Schein, and H. Burd (2021, June). Supporting men to take longer parental leave and work flexibly. Research report, Behavioural Insights Team.
- Davis, K. D., K. M. Lawson, D. M. Almeida, E. L. Kelly, R. B. King, L. Hammer, L. M. Casper, C. A. Okechukwu, G. Hanson, and S. M. McHale (2015). Parents’ daily time with their children: A workplace intervention. *Pediatrics* 135(5), 875–882.
- Ekberg, J., R. Eriksson, and G. Friebel (2013). Parental leave — a policy evaluation of the Swedish "Daddy-month" reform. *Journal of Public Economics* 97, 131–143.
- Farré, L. and L. González (2019). Does paternity leave reduce fertility? *Journal of Public Economics* 172, 52–66.
- Feuillade, M., D. Goux, and E. Maurin (2025, July). Rise in home working and spousal labor supply. IZA Discussion Paper 17997, IZA Institute of Labor Economics.
- Haas, L. and C. P. Hwang (2019). Policy is not enough – the influence of the gendered workplace on fathers’ use of parental leave in Sweden. *Community, Work & Family* 22(1), 58–76.
- Haerpfer, C., R. Inglehart, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, B. Puranen, et al. (Eds.) (2020). *World Values Survey: Round Seven - Country-Pooled Datafile*. Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat.
- Japan Institute for Labour Policy and Training (2017). Labor situation in japan and its analysis: Detailed exposition 2016/2017. Technical report, Japan Institute for Labour Policy and Training. Chapter 7: Work-Life Balance.
- Johnsen, J. V., H. Ku, and K. G. Salvanes (2024). Competition and career advancement. *Review of Economic Studies* 91(5), 2954–2980.
- Kelly, E. L., P. Moen, J. M. Oakes, W. Fan, C. Okechukwu, K. D. Davis, L. B. Hammer, E. E. Kossek, R. B. King, G. C. Hanson, F. Mierzwa, and L. M. Casper (2014).

- Changing work and work-family conflict: Evidence from the work, family, and health network. *American Sociological Review* 79(3), 485–516.
- Kitamura, T. and C. Suzuki (1986, June). On the Japanese version of the social desirability scale (in Japanese). *Journal of Social Psychiatry (Shakai Seishin Igaku)* 9(2), 173–180. Original text in Japanese.
- Kluve, J. and M. Tamm (2013). Parental leave regulations, mothers’ labor force attachment and fathers’ childcare involvement: Evidence from a natural experiment. *Journal of Population Economics* 26(3), 983–1005.
- Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies* 76(3), 1071–1102.
- Lee, S., K. M. Lawson, and S. Damaske (2019). Crossover of resources and well-being within employee–partner dyads: Through increased schedule control. *Community, Work & Family* 22(4), 391–411.
- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics* 99(2), 210–221.
- Ministry of Economy, Trade and Industry (2024). Initiatives on the use of home support services.
- Ministry of Health, Labour and Welfare (2012). Longitudinal survey of newborns in the 21st century: Heisei 22 (2010) birth cohort, second-year follow-up.
- Ministry of Health, Labour and Welfare (2021). Reiwa-2 koyo kinto kihon chousa. Technical report, Ministry of Health, Labour and Welfare. [Basic Survey of Gender Equality in Employment Management 2020].
- Ministry of Health, Labour and Welfare (2024, July). Reiwa-5 koyo kinto kihon chousa. Technical report, Ministry of Health, Labour and Welfare. [Basic Survey of Gender Equality in Employment Management, FY2023]; male childcare leave take-up rate 30.1 percent.
- Mitsubishi UFJ Research and Consulting (2018). Danseino ikujikyugyo shutoku ni kansuru chousa kekka. Technical report, Mitsubishi UFJ Research and Consulting. [Survey results on men’s take-up of childcare leave].
- Miyajima, T. and H. Yamaguchi (2017). I want to but i won’t: Pluralistic ignorance inhibits intentions to take paternity leave in Japan. *Frontiers in Psychology* 8, 1508.

- National Council of Nursery Schools (2022). Zenkoku hoiku kyougikai kaiin no jittai chousa 2021 houkokusho [survey of member childcare centers 2021: Report]. <https://www.zenhokyo.gr.jp/>. In Japanese.
- OECD (2023). Oecd gender data portal and oecd statistics database. <https://stats.oecd.org/>.
- OECD (2024). Oecd family database. <https://www.oecd.org/els/family-database.htm>. Accessed January 2025.
- Oi, M. (2019, October). Lifting the lid on japan’s bento boxes. <https://www.bbc.com/future/article/20191013-lifting-the-lid-on-japans-amazing-bento-boxes>. BBC Future.
- Olafsson, A. and H. Steingrimsdottir (2020). How does daddy at home affect marital stability? *Economic Journal* 130(629), 1471–1500.
- Patnaik, A. (2019). Reserving time for daddy: The consequences of fathers’ quotas. *Journal of Labor Economics* 37(4), 1009–1059.
- Suto, M., O. O. Balogun, B. Dhungel, T. Kato, and K. Takehara (2022). Effectiveness of workplace interventions for improving working conditions on the health and wellbeing of fathers or parents: A systematic review. *International Journal of Environmental Research and Public Health* 19(8), 4779.
- Tamm, M. (2019). Fathers’ parental leave-taking, childcare involvement, and maternal labor market participation. *Labour Economics* 59, 184–197. Germany.
- Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics* 141(2), 1281–1301.

Supplemental Appendix

A Details of Experiments and Survey Questions

A.1 Invitation to Program

The program materials and survey instrument were administered in Japanese. This appendix reproduces them in English translation.

With the enforcement of the revised Childcare and Family Leave Act, the environment surrounding today's child-rearing generation has been undergoing significant changes. Last year, the male childcare leave take-up rate reached a record high of 17%, and society is gradually shifting toward one in which it is normal for men to take childcare leave in order to balance work and family life. Still, many employees may feel uncertain or have questions such as: how should one take childcare leave, and what should be done during the leave period?

This training program will help participants understand the meaning of men taking childcare leave, recognize the hurdles to doing so, and learn the knowledge and know-how to overcome them. Through interactive exercises, participants will not only acquire knowledge but also reflect on what it means for themselves and their families.

We will also conduct training for managers. As organizations establish supportive systems and environments and as increasingly diverse employees with different backgrounds and circumstances work together, what kind of management is required today? This program explores the concept of "Ikuboss-style management," examining how to embody it and how to lead organizational growth. *Ikuboss* refers to a manager who supports subordinates' careers and work-life balance, while also ensuring organizational performance and enjoying both work and private life. Both training programs will be conducted by experienced instructors from NPO Fathering Japan, one of the leading organizations promoting male childcare leave in Japan:

- **Training for Employees:** (*instructor name*), Representative Director of NPO Fathering Japan. (*personal description*)
- **Training for Managers:** (*instructor name*), Board Member of NPO Fathering Japan. (*personal description*)
- **About NPO Fathering Japan:** (*organizational description*)

To encourage broad participation, both training sessions will be available as an archived recording for 10 days after the event. If you cannot attend live, please make sure to watch the archive. Please register for participation using the following URLs. To track participation rates, those who wish to view the archive are also requested to register:

[*Registration link to employee training*]

[*Registration link to manager training*]

A.2 Training curriculum

The employee and manager sessions were delivered by two experienced male instructors affiliated with NPO Fathering Japan. Both had professional experience in corporate settings and personal experience taking paternity leave and combining work with active involvement in childcare. Their backgrounds were relevant to the program’s emphasis on presenting paternal involvement and leave-taking as compatible with professional careers. Table A.1 lists the blocks of each session in the order they were presented, with the share of slides each occupied and the category of message it carried. The table covers the lecture portion of each session, which is what the projection deck documents; both sessions also concluded with a group discussion, which was conducted separately from the lecture and is not reflected in the deck. We summarize the two lectures in turn.

The employee lecture opened with the instructor’s own leave history and the organization’s mission, then covered the 2022 revision of the Child Care and Family Care Leave Act: the employer’s new duty to inform employees individually and to confirm their intentions, the postpartum paternity leave window, split-taking, the wage replacement rate, the social insurance exemption, and the disclosure requirement for large firms. Its largest block, roughly a third of the slides, concerned the perinatal period and the reality of newborn care, and included a tabulation from an actual childcare diary showing that a newborn absorbs about eleven hours a day in cycles of feeding, changing, and settling that never yield more than ninety minutes of continuous free time. A shorter block on careers presented the organization’s own 2019 survey of employees who had taken leave, in which over ninety percent reported that their evaluations by supervisors and by colleagues had improved or were unchanged. The lecture closed with an individual exercise in which participants audited their own household division of tasks against survey data on how husbands and wives report that split.

The manager lecture shared the legal material and the perinatal background but re-framed the rest around the workplace, and gave the largest single block to the law. It opened with a self-check on six common assumptions about men’s leave and a warning about paternity harassment. Its second part introduced the “ikuboss” model of a supervisor who accommodates constraints on subordinates’ time, and an exercise on attribution bias in which managers compare a long-hours employee with a time-constrained one and are asked to separate observed behavior from inferred motivation. Managers were then given an operational playbook for covering an absence: notify the team, inventory the tasks, redistribute, improve or eliminate what can be, and arrange replacement staff, all framed as something they have already done for bereavement leave and long postings.

The lecture ended with a role-play on responding to a subordinate who says he does not want to take leave, and a two-minute exercise in which each manager committed to one new action.

Table A.1: Content of the two training sessions

<i>Panel A: Employee session (91 slides)</i>				
	Block	Slides	Share	Message
1	Instructor’s leave history; organization’s mission	1–10	11%	Role model
2	2022 revision of the leave law	11–18	9%	Policy
3	Environment facing young families	19–32	15%	Policy, workplace
4	Perinatal period and newborn care	33–62	33%	Household
5	Careers, evaluations, and leave-takers’ reports	64–74	12%	Workplace
6	Household division of tasks (individual exercise)	75–86	13%	Household
7	Closing	87–91	5%	Household, workplace
<i>Panel B: Manager session (79 slides)</i>				
	Block	Slides	Share	Message
1	Instructor’s leave history; organization’s mission	1–10	13%	Role model
2	Session goals; self-check on assumptions	11–13	4%	Endorsement
3	2022 revision of the leave law	14–31	23%	Policy
4	Why the postpartum period matters	32–34	4%	Household
5	Managing a diverse team; the ikuboss model	35–47	16%	Workplace, endorsement
6	Attribution bias in evaluation (exercise)	48–50	4%	Workplace
7	Workplace return to leave; covering an absence	52–59	10%	Workplace
8	Responding to a reluctant subordinate (role-play)	60–65	8%	Endorsement, workplace
9	Careers and team parenting; action commitment	66–74	11%	Workplace, household
10	Closing	75–79	6%	Endorsement

Notes. Blocks are listed in the order presented. Slide numbers refer to the projection deck used in the training, and shares are of all slides in that deck. The deck documents the lecture portion of the session; the group discussion that concluded each session is not part of it, so the shares are shares of lecture slides rather than of session time. The break slide is not assigned to a block, so the shares do not sum to one hundred. “Message” classifies the content of the block into the categories used in Section 5.1: policy information, the workplace return to leave-taking, the household division of childcare, the employer’s endorsement of leave-taking, and the instructor as a role model. The outline is the authors’ summary of the training materials provided by NPO Fathering Japan, published with its permission. It paraphrases the sessions and does not reproduce the slides, which remain the property of the organization.

A.3 Informed Consent (Baseline Survey)

This survey will be conducted by Evidence Studio and the University of Tokyo in cooperation with [ORGANIZATION NAME] for the purpose of conducting research on the promotion of men taking paternity leave. Your responses will be used as valuable data for research purposes. They will not be used for any other purpose. In particular, we will not disclose your answers in a way that personally identifies you, nor will we share them with the company, nor will they affect your personnel affairs.

Consent to Survey

[ORGANIZATION NAME] will ask you in advance whether or not you agree to participate in this survey (Research Project “Request for Participation in the Empirical Study to Promote Male Parental Leave Utilization”) in the form of an internal questionnaire at [ORGANIZATION NAME]. Please complete the survey only if you agree. In addition, the handling of response data for this survey is as follows.

Anonymity of survey results

Responses will be sent directly to Evidence Studio. No personally identifiable response data will be shared with [ORGANIZATION NAME], including the person in charge. [ORGANIZATION NAME]’s personnel involved in this project will only review statistical data that the researchers have processed into a form that does not identify you personally. Please be honest in your responses, as they will not be shared with your supervisors or affect your personal evaluations.

Use of Survey Results

The results of the research will be published at conferences and in academic journals. Your name and other personal information will not be revealed.

- (1) What is your survey ID (please check it carefully to make sure it is correct)?
- (2) Have you agreed to cooperate with this survey?

☐ Agreed.

☐ Not agreed.

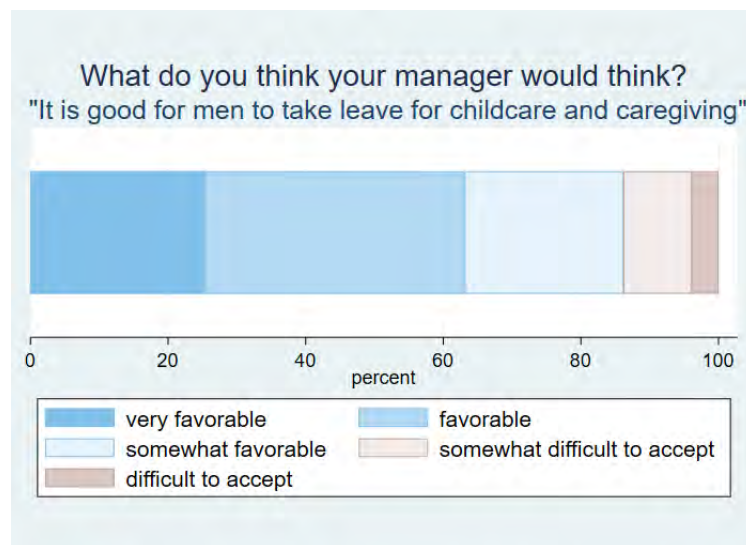
A.4 Information Provision (First Follow-up Survey)

The screens below illustrate the information shown in one participating organization. The organization name and the numerical values displayed to respondents varied across organizations based on baseline survey responses within each organization.

Image shown to treatment group (Page 1)

In the previous survey, we asked how much you believe **your direct supervisors at your workplace** would personally agree with the statement, *“It is a good thing for men to take leave for childcare and caregiving.”*

Among non-managerial employees at [ORGANIZATION NAME], **63.1%** of respondents thought that their supervisors would be “very favorable” or “favorable” toward this opinion.



On the other hand, when responses from supervisors at [ORGANIZATION NAME] were analyzed, **94%** of them indicated that they “strongly agree” or “agree” with this view.

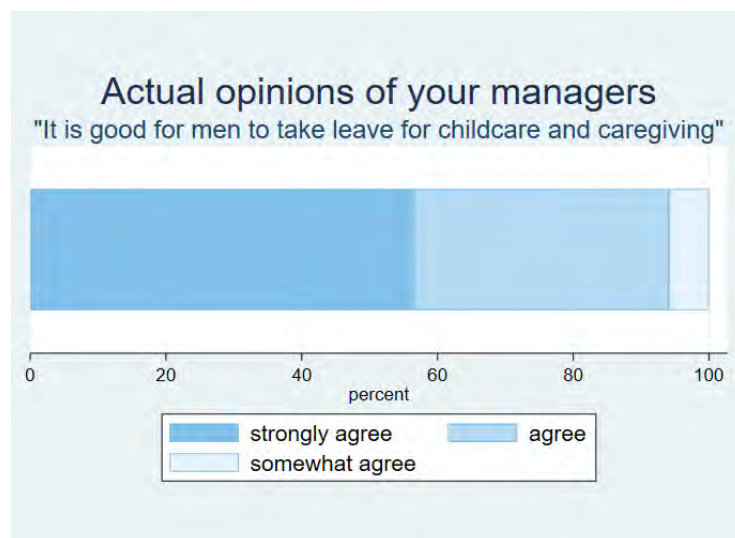
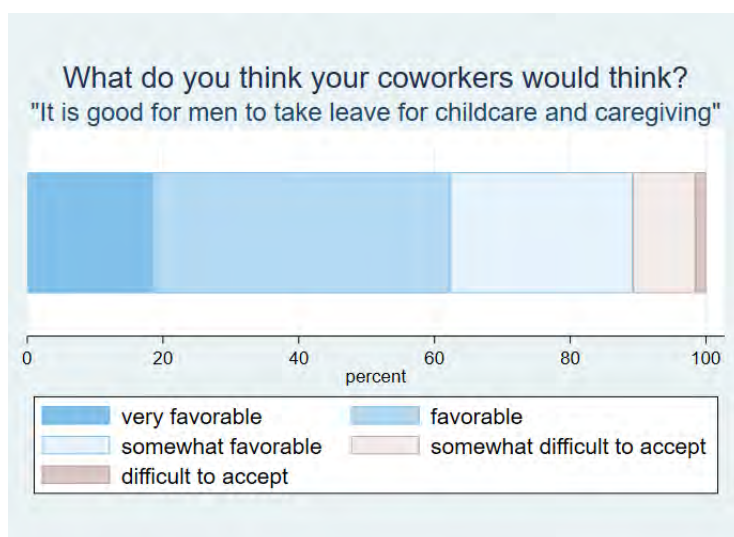


Image shown to treatment group (Page 2)

In the previous survey, we asked how much you believe **your colleagues at your workplace** would personally agree with the statement, *“It is a good thing for men to take leave for childcare and caregiving.”*

Among non-managerial employees at [ORGANIZATION NAME], **62.3%** of respondents thought that their colleagues would be “very favorable” or “favorable” toward this opinion.



On the other hand, when responses from colleagues in non-managerial positions at [ORGANIZATION NAME] were analyzed, **94.7%** of them indicated that they “strongly agree” or “agree” with this view.

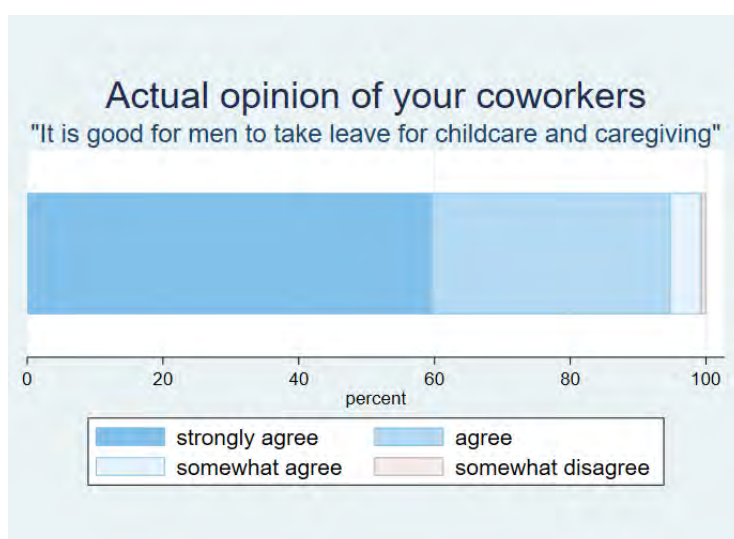


Image shown to treatment group (Page 3) and control group

On the following pages, you will find the questions for this survey. Some of the questions are the same as those in the previous survey, but please answer them based on your current thoughts and circumstances.

☐ Agreed.

A.5 Questions about Beliefs

The response options below are reproduced as they were presented to respondents, who selected the most favorable option as 1 and the least favorable as 6. Throughout the

analysis, all belief items are reverse-coded so that higher values indicate stronger agreement with the statement, or a more favorable attitude toward men’s paternity leave uptake. The control means reported in the estimation tables, and the horizontal axes of the cumulative distributions in Figure 2, are on this reverse-coded scale.

Own belief - Favorable toward PL uptake

How much do you *personally* agree with the following statement? Please select the most applicable answer.

	Response options					
	Strongly agree	Agree	Somewhat agree	Somewhat disagree	Disagree	Strongly disagree
(1) <i>It is a good thing for men to take leave for childcare and caregiving.</i>	1	2	3	4	5	6

Own belief - PL benefits workplace

How much do you *personally* agree with the following statement? Please select the most applicable answer.

	Response options					
	Strongly agree	Agree	Somewhat agree	Somewhat disagree	Disagree	Strongly disagree
(2) <i>Increasing men’s take-up of paternity leave can improve company operations and promote a collaborative workplace culture.</i>	1	2	3	4	5	6

Second-order belief about managers (colleagues) - Favorable toward PL uptake

Due to constraints in Japanese phrasing, this question omits the “It is a good thing . . .” clause used in the own-belief item. The equivalent normative nuance is instead embedded in the response options presented to respondents.

How much do you think your direct managers (colleagues) at your workplace would agree with the following statement? Please select the most applicable answer.

	Response options					
	Very favorable	Favorable	Somewhat favorable	Somewhat difficult to accept	Difficult to accept	Very difficult to accept
(1) <i>Men take leave for childcare and caregiving.</i>	1	2	3	4	5	6

Second-order belief about managers (colleagues) - PL benefits workplace

How much do you think your direct managers (colleagues) at your workplace would agree with the following statement? Please select the most applicable answer.

	Response options					
	Strongly agree	Agree	Somewhat agree	Somewhat disagree	Disagree	Strongly disagree
(2) <i>Increasing men's take-up of paternity leave can improve company operations and promote a collaborative workplace culture.</i>	1	2	3	4	5	6

A.6 Questions about Hours

At baseline, 2.6 percent of respondents left all housework fields blank, and among respondents with a child, 3.6 percent left all childcare fields blank; 98.5 percent reported their working hours. When all fields for an activity were left blank, we treat the response as missing rather than interpreting the blanks as zeros, because respondents may have found the hours difficult to estimate. In the small number of cases in which a respondent reported hours for either weekdays or weekends but left the corresponding field for the other day type blank, we code the isolated blank as zero. This affects seven respondents for housework and three for childcare.

Weekly working hours

How many hours do you typically work in a week (including overtime)? Please also answer about your spouse (including de facto partners). If your spouse is not working, please enter “0”.

- You (in hours per week): hours
- Your spouse (in hours per week): hours

Daily housework hours

How many hours per day on average do you spend on housework (excluding childcare)? Please answer separately for weekdays and weekends. Please also answer about your spouse (including de facto partners).

- You
(Weekday): hours minutes
(Weekend): hours minutes
- Your spouse
(Weekday): hours minutes
(Weekend): hours minutes

Daily childcare hours

How many hours per day on average do you spend on childcare (excluding sleeping time)? Please answer separately for weekdays and weekends. Please also answer about your spouse (including de facto partners).

- You
(Weekday): hours minutes
(Weekend): hours minutes
- Your spouse
(Weekday): hours minutes
(Weekend): hours minutes

A.7 Questions about paternity leave

Intention to take paternity leave

Do you plan or intend to take childcare leave in the future?

- ☐ I am currently taking paternity leave
- ☐ I plan or hope to take paternity leave within the next six months
- ☐ I plan or hope to take paternity leave in the future, but not sure when
- ☐ I do not plan to take paternity leave, even if I have a (new) child
- ☐ I have finished raising children, do not plan to have or raise children, or am currently raising children but do not plan to take paternity leave in the future
- ☐ Not sure
- ☐ Prefer not to answer

Researched about PL

During the past month, have you consulted someone or searched websites or documents on your own regarding the parental leave system?

- ☐ Yes
- ☐ No

A.8 Questions about Workplace Practices and Relationships

Number of workers for my task today

Imagine that one day you suddenly need to leave work early. At that time, how many coworkers would likely be able to take over your main job duties?

- persons

Number of workers for my task for a month

Imagine that three months from now, you have to take one month off from work. During that period, how many coworkers would likely be able to take over your main job duties?

- persons

Meeting hours

On average, how many hours per week do you spend in internal meetings or discussions?

- hours

Remote work days

How many days per week do you usually work remotely (telework)?

- days

Manager knows family composition

Does your direct supervisor know about your main family composition?

☐ Yes

☐ No

Psychological safety

The following questions ask about your coworkers and your direct supervisor. For each statement, please select the option that best applies to you.

[1. Always 2. Usually 3. Often 4. Sometimes 5. Rarely 6. Never]

- (1) When I make mistakes, I am often criticized.
- (2) I can point out negative issues or problems with my coworkers or direct supervisor, and they can do the same with me.
- (3) My coworkers or direct supervisor sometimes reject others because they are different.
- (4) It is safe to take risks in what I say or do with my coworkers or direct supervisor.
- (5) There is an atmosphere that makes it difficult to ask my coworkers or direct supervisor for help.
- (6) Some coworkers or my direct supervisor deliberately act in ways that undermine my work.
- (7) When working with my coworkers or direct supervisor, I feel that my talents and skills are respected and utilized.

Variable construction: responses are oriented so that higher values indicate greater psychological safety, reverse-coding the items whose wording runs in the opposite direction, and the variable is the average of the 7 items. It therefore ranges from 1 to 6 on the same scale as the individual items.

Coworker and supervisor relationships

The following questions ask about your coworkers and your direct supervisor. For each statement, please select the option that best applies to you.

[1. Always 2. Usually 3. Often 4. Sometimes 5. Rarely 6. Never]

- (1) I have opportunities to express my opinions while working.
- (2) I can rely on my coworkers to get my work done.
- (3) My direct supervisor treats me with respect.
- (4) My direct supervisor treats everyone fairly.

Variable construction: responses are oriented so that higher values indicate better relationships with coworkers and the direct supervisor, and the variable is the average of the 4 items. It therefore ranges from 1 to 6 on the same scale as the individual items.

Recommend firm to jobseeker

Would you recommend your company as a workplace to someone you know who is currently seeking a job?

- ☐ Yes
- ☐ No

B Additional Figures and Tables

B.1 Sample and Design

Table B.1: Characteristics of participating organizations before intervention

		% Paternity leave uptake	% Female employees	% Female managers
Organization 1	mid-size municipality	20	-	10-15
Organization 2	mid-size municipality	60	40	-
Organization 3	listed firm in retail/wholesale	-	20	10-15
Organization 4	listed firm in retail/wholesale	20	30	<10
Representative sample (public)	local municipality offices ^a	44.1	39.6	19
Representative sample (private)	private firms ^b	14	29.6	12.7
Representative sample (private)	listed firms ^c	52.2		-

Notes. The figures for the four organizations were rounded for confidentiality purposes. Paternity leave uptake rate is based on publicly available information from 2022. The proportion of female employees and the proportion of female managers in the four organizations are based on the most recent publicly available information prior to the interventions (2021–2023). The populations of the mid-size municipalities are between 100,000 and 200,000.

a - “Survey Results on Working Conditions in Local Governments, Survey on the Actual Conditions of Local Government Salaries (Ministry of Internal Affairs and Communications, Japan)”, “Status of Measures for Gender Equality or Women-Related Policies in Local Governments (Cabinet Office)”. All figures are from 2022, except for the proportion of female employees (2018).

b - “Basic Survey on Gender Equality in Employment Management (Ministry of Health, Labour, and Welfare, Japan)”. All figures are based on data from 2022.

c - “Survey on paternity leave uptake among 2,456 listed companies for FY2022: 52.2% uptake rate, well above official statistics (Tokyo Shoko Research., September 12, 2023)”.

Table B.2: Program sign-up and survey response rates at the office level and training treatment assignment

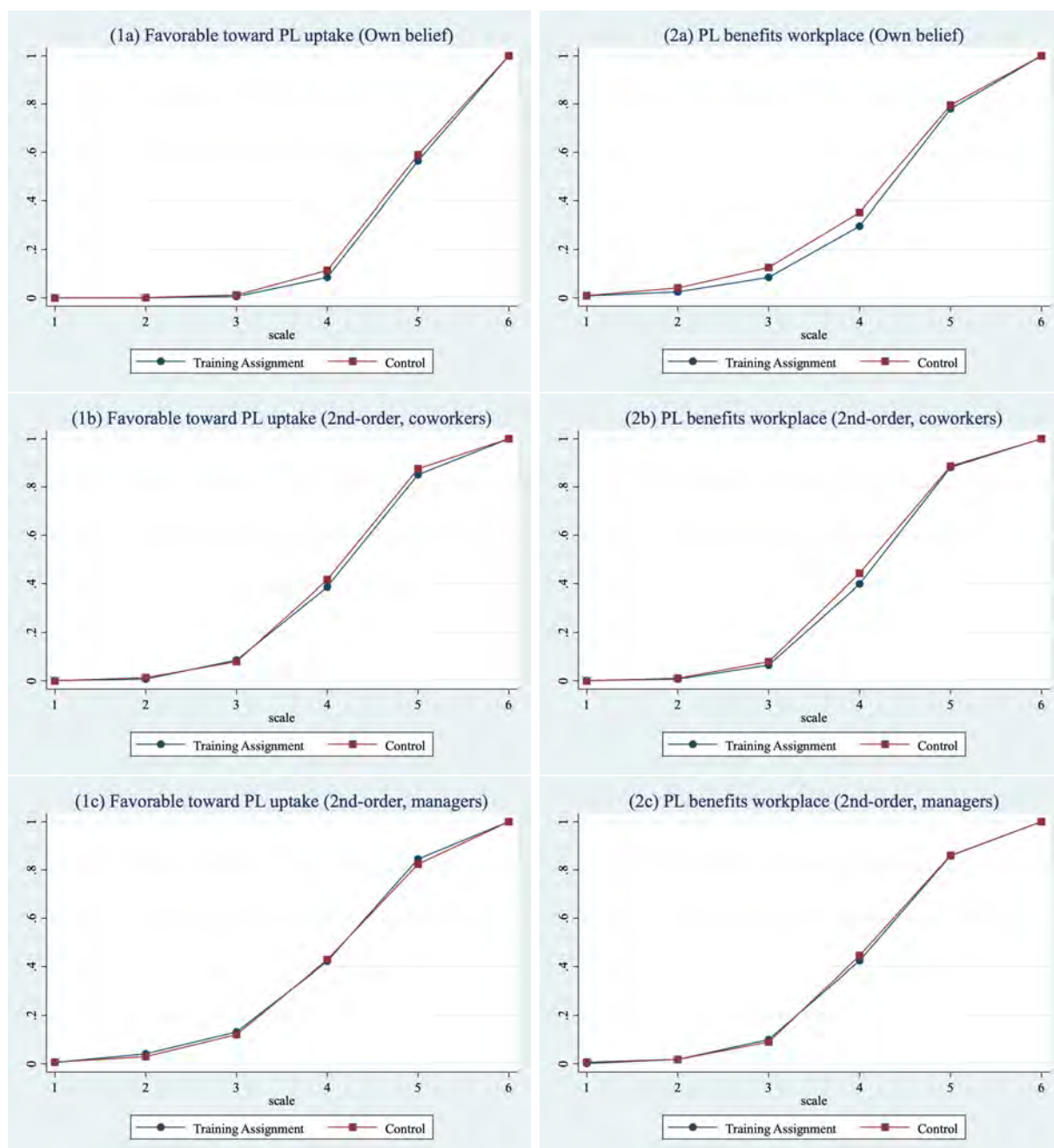
	Program Sign-up Rate		Survey Response Rate Baseline		Survey Response Rate 1st Follow-Up		Survey Response Rate 2nd Follow-up	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Training assignment	0.028 (0.076)	0.021 (0.066)	0.046 (0.038)	0.045 (0.038)	-0.034 (0.051)	-0.034 (0.052)	0.029 (0.062)	0.028 (0.063)
Observations	76	76	80	80	80	80	80	80
Strata FE	No	Yes	No	Yes	No	Yes	No	Yes
Mean Dep. Var	0.259	0.259	0.910	0.910	0.792	0.792	0.580	0.580

Notes. This table examines the relationship between the training treatment assignment and office level sign-up rate to the training program and survey response rates for the baseline, first follow-up, and second follow-up surveys. There were 1,962 employees (in 80 offices) who signed up for the training program, to whom the invitations to the baseline and follow-up surveys were sent. In Columns (1)–(2), the dependent variable is the sign-up rate for the training program among all employees in each office. The overall employee-level sign-up rate across the four organizations was 12.4 percent; the mean reported in Columns (1)–(2) is the unweighted mean of office-level sign-up rates among the 76 offices for which the number of employees is available. Because we do not have information on the total number of employees in four offices, observations for these offices are missing. In the remaining columns, the dependent variables are the response rates to the baseline survey (Columns (3)–(4)), the first follow-up survey (Columns (5)–(6)), and the second follow-up survey (Columns (7)–(8)) among male employees who signed up for the training program in each office. Columns (2), (4), (6), and (8) control for the randomization strata fixed effect. Robust standard errors are shown in the parentheses.

B.2 Results: Training Intervention

B.2.1 Beliefs

Figure B.1: Cumulative Distribution: Training Assignment vs. Control at First Follow-up



Notes. Data on first- and second-order beliefs measured in the first follow-up survey. In each panel, the horizontal axis is the six-point response scale, running from 1 (least supportive) to 6 (most supportive), and the vertical axis is the share of respondents at or below that scale point. A curve lying below and to the right of the other therefore indicates more supportive responses. Panels (1a)–(1c) report the statement “Favorable toward PL uptake” and Panels (2a)–(2c) the statement “PL benefits the workplace”; within each set, the panels report the respondent’s own belief, his second-order belief about coworkers, and his second-order belief about managers, in that order.

Table B.3: Impacts of training on beliefs (IV estimates)

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Favorable toward PL uptake	PL benefits workplace	Favorable toward PL uptake	PL benefits workplace	Favorable toward PL uptake	PL benefits workplace
Training participation	0.076 (0.057)	0.447*** (0.105)	0.145 (0.145)	0.291** (0.143)	-0.108 (0.114)	0.094 (0.121)
Anderson q-value	[0.101]	[0.001]	[0.488]	[0.199]	[0.488]	[0.488]
Observations	1225	1223	1206	1208	1218	1218
Control Mean	5.265	4.591	4.594	4.547	4.585	4.567
No. clusters	74	74	74	74	74	74
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table reports instrumental-variable (IV) estimates of the effect of training participation on the belief outcomes of Table 3, instrumenting participation with training assignment. The coefficients rescale the assignment effect by the participation rate. As discussed in Section 3.5, exposure to trained managers may violate the exclusion restriction required to read them as the effect of a respondent's own attendance; they are best read as the effect of the training as experienced in a treated office, scaled by participation. Anderson (2008) sharpened q-values, reported in brackets, are computed within families defined by first-order and second-order beliefs. The rest of the specification is the same as in Table 3.

Table B.4: Impact of training on beliefs: without PDS LASSO control

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Favorable toward PL uptake	PL benefits workplace	Favorable toward PL uptake	PL benefits workplace	Favorable toward PL uptake	PL benefits workplace
Training assignment	0.041 (0.025)	0.180*** (0.050)	0.058 (0.061)	0.117* (0.061)	-0.043 (0.045)	0.038 (0.049)
Anderson q-value	[0.058]	[0.002]	[0.501]	[0.306]	[0.501]	[0.501]
Observations	1225	1223	1206	1208	1218	1218
Control Mean	5.265	4.591	4.594	4.547	4.585	4.567
No. clusters	74	74	74	74	74	74
PDS LASSO Control	No	No	No	No	No	No

Notes. The specification is the same as in Table 3 except that we do not control for the variables selected by PDS LASSO.

Table B.5: Impact of training treatment on beliefs: HTE by initial belief

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Favorable toward PL uptake	PL benefits workplace	Favorable toward PL uptake	PL benefits workplace	Favorable toward PL uptake	PL benefits workplace
Training assignment × Higher Baseline belief	0.011 (0.046)	0.156 (0.145)	0.115 (0.120)	0.119 (0.075)	-0.136 (0.129)	0.104* (0.056)
Training assignment × Lower Baseline belief	0.044* (0.026)	0.187*** (0.052)	0.055 (0.059)	0.116* (0.064)	-0.020 (0.042)	-0.012 (0.073)
Observations	1225	1223	1206	1208	1218	1218
No. clusters	74	74	74	74	74	74
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	5.265	4.591	4.594	4.547	4.585	4.567
P-val: High=Low	0.553	0.847	0.577	0.966	0.341	0.222

Notes. We estimate the heterogeneity in the treatment effect of the training intervention by the baseline value of the belief variable indicated in each column (i.e., own belief, second-order belief about coworkers, or second-order belief about managers, for the indicated statement). Specifically, for each belief variable in the column, "Higher (Lower) baseline belief" indicates that the respondent's baseline belief was above (at or below) the median among the baseline male sample. The specification follows that of Table 3, except that training assignment is interacted with these baseline belief indicators; the coefficients are intent-to-treat effects.

Table B.6: Impact of training treatment on beliefs: HTE by manager/non-manager and existence of children

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)	
	(1) Favorable toward PL uptake	(2) PL benefits workplace	(3) Favorable toward PL uptake	(4) PL benefits workplace	(5) Favorable toward PL uptake	(6) PL benefits workplace
<i>Panel A: Manager / Non-manager</i>						
Training assignment $\times$ Manager	0.048 (0.032)	0.182*** (0.056)	0.060 (0.077)	0.082 (0.075)	-0.004 (0.058)	-0.023 (0.056)
Training assignment $\times$ Non-manager	0.017 (0.030)	0.182** (0.073)	0.060 (0.065)	0.160** (0.070)	-0.075 (0.072)	0.098 (0.077)
Manager	-0.030 (0.029)	-0.034 (0.061)	-0.025 (0.031)	-0.036 (0.050)	-0.057 (0.065)	0.037 (0.074)
Observations	1225	1223	1206	1208	1218	1218
No. clusters	74	74	74	74	74	74
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	5.265	4.591	4.594	4.547	4.585	4.567
P-val: Manager=Non-manager	0.448	0.999	0.997	0.357	0.456	0.196
<i>Panel B: Have a child / No child</i>						
Training assignment $\times$ Children	0.033 (0.026)	0.186*** (0.047)	0.070 (0.063)	0.113* (0.062)	-0.012 (0.039)	0.018 (0.047)
Training assignment $\times$ No children	0.032 (0.053)	0.167 (0.128)	0.038 (0.080)	0.141 (0.093)	-0.108 (0.110)	0.121 (0.095)
Child	-0.042 (0.048)	0.026 (0.099)	-0.047 (0.063)	-0.011 (0.084)	-0.123 (0.081)	-0.062 (0.073)
Observations	1225	1223	1206	1208	1218	1218
No. clusters	74	74	74	74	74	74
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	5.265	4.591	4.594	4.547	4.585	4.567
P-val: Child=No Child	0.990	0.887	0.666	0.748	0.371	0.259

Notes. We estimate the heterogeneity of the treatment effect of the training intervention by managers and non-managers (Panel A) and by whether the respondent has at least one child or not (Panel B). The specification follows that of Table 3, except that training assignment is interacted with these baseline characteristics; the coefficients are intent-to-treat effects.

B.2.2 Behavioral Outcomes

Table B.7: Impact of Training on Behaviors (IV estimates)

	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)
<i>Panel A: Own hours</i>					
Training participation	-0.833 (1.764)	-0.113 (0.220)	0.943** (0.453)	0.022 (0.113)	0.040 (0.236)
Anderson q-value		[1.000]	[0.176]	[1.000]	[1.000]
Observations	1177	876	876	1161	1161
No. clusters	74	72	72	74	74
Control Mean	43.981	1.386	5.768	1.003	2.251
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Spouse hours</i>					
Training participation	3.578** (1.729)	-0.028 (0.385)	0.108 (0.453)	-0.495 (0.390)	-0.084 (0.335)
Anderson q-value		[1.000]	[1.000]	[1.000]	[1.000]
Observations	994	846	846	980	980
No. clusters	74	72	72	74	74
Control Mean	20.213	4.779	7.110	4.487	4.182
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes

Notes. This table reports instrumental-variable (IV) estimates of the effect of training participation on the behaviors of Table 4, instrumenting participation with training assignment. The coefficients rescale the assignment effect by the participation rate. As discussed in Section 3.5, exposure to trained managers may violate the exclusion restriction required to read them as the effect of a respondent's own attendance; they are best read as the effect of the training as experienced in a treated office, scaled by participation. Panel A shows the results for male respondents, and Panel B shows the results for their spouses. The rest of the specification, including the construction of Anderson (2008) sharpened q-values, is the same as in Table 4.

Table B.8: Impact of training on behaviors: without PDS LASSO control

	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)
<i>Panel A: Own hours</i>					
Training assignment	0.005 (0.755)	-0.031 (0.101)	0.577** (0.240)	-0.018 (0.046)	0.013 (0.096)
Anderson q-value		[1.000]	[0.082]	[1.000]	[1.000]
Observations	1177	876	876	1161	1161
No. clusters	74	72	72	74	74
Control Mean	43.981	1.386	5.768	1.003	2.251
PDS LASSO Control	No	No	No	No	No
<i>Panel B: Spouse hours</i>					
Training assignment	1.469** (0.676)	0.137 (0.187)	0.272 (0.246)	-0.204 (0.165)	-0.034 (0.145)
Anderson q-value		[1.000]	[1.000]	[1.000]	[1.000]
Observations	994	846	846	980	980
No. clusters	74	72	72	74	74
Control Mean	20.213	4.779	7.110	4.487	4.182
PDS LASSO Control	No	No	No	No	No

Notes. The specification is the same as in Table 4 except that we do not control for the variables selected by PDS LASSO.

Table B.9: Impact of Training on Behaviors by Presence and Age of Children (IV estimates)

	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)
<i>Panel A: Own hours</i>					
Training participation $\times$ Young children	-3.958 (2.670)	-0.721 (0.502)	2.286*** (0.864)	-0.290 (0.233)	-0.875 (0.757)
Anderson q-value		[0.229]	[0.034]	[0.229]	[0.229]
Training participation $\times$ Only older children	-0.166 (2.705)	0.251 (0.237)	0.331 (0.559)	0.386** (0.155)	0.442 (0.289)
Training participation $\times$ No children	2.308 (7.735)			0.182 (0.272)	0.711** (0.296)
Observations	1015	859	859	1001	1001
No. clusters	74	72	72	74	74
Control Mean (Young children)	43.479	2.302	10.652	1.332	3.495
Control Mean (Only older children)	45.195	0.966	3.509	0.726	1.721
Control Mean (No children)	42.000			0.846	1.564
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Spouse hours</i>					
Training participation $\times$ Young children	4.353* (2.475)	0.310 (0.864)	-0.451 (1.186)	-1.334* (0.719)	-2.061** (0.906)
Anderson q-value		[0.562]	[0.562]	[0.106]	[0.101]
Training participation $\times$ Only older children	1.789 (2.068)	0.204 (0.478)	0.440 (0.762)	-0.329 (0.464)	0.631* (0.338)
Training participation $\times$ No children	12.39 (8.397)			0.644 (1.330)	1.615 (1.386)
Observations	994	846	846	980	980
No. clusters	74	72	72	74	74
Control Mean (Young children)	16.929	8.238	11.628	5.601	5.373
Control Mean (Only older children)	21.078	3.121	4.943	4.336	3.955
Control Mean (No children)	23.297			2.760	2.638
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes

Notes. This table uses the sample of men who were married at the time of the baseline survey. We estimate heterogeneity in treatment effects by interacting training participation with indicator variables for the presence and age group of children and using training assignment with the same indicator variables as instrumental variables. The coefficients rescale the assignment effect by the participation rate. As discussed in Section 3.5, exposure to trained managers may violate the exclusion restriction required to read them as the effect of a respondent's own attendance; they are best read as the effect of the training as experienced in a treated office, scaled by participation. "Young children", "Only older children", and "No children" are indicator variables equal to 1 if the household has at least one young child (aged 5 or below), only older children (aged 6 or above), or no children, respectively. The rest of the specification, including the construction of Anderson (2008) sharpened q-values, is the same as in Table 5.

Table B.10: Impact of training on behaviors: HTE by manager/non-manager

	(1)	(2)	(3)	(4)	(5)	(6)
	Weekly Working Hours	Daily Childcare Hours (Weekday)	Daily Childcare Hours (Weekend)	Daily Housework Hours (Weekday)	Daily Housework Hours (Weekend)	Researched about PL
Training assignment $\times$ Manager	-0.333 (0.701)	0.033 (0.132)	0.349 (0.253)	0.081 (0.052)	0.075 (0.145)	0.073** (0.029)
Training assignment $\times$ Non-manager	-0.427 (1.148)	-0.028 (0.143)	0.477 (0.344)	-0.065 (0.080)	-0.000 (0.128)	0.055** (0.027)
Observations	1177	876	876	1161	1161	1211
No. clusters	74	72	72	74	74	74
Control Mean	43.981	1.386	5.768	1.003	2.251	0.189
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
P-values: Coefficients difference	0.944	0.757	0.771	0.152	0.694	0.578

Notes. We estimate the heterogeneity of the treatment effect of the training intervention by managers and non-managers. The specification is the same as in Table 4, Panel A, except that training assignment is interacted with indicators for managers and non-managers; the coefficients are intent-to-treat effects. In addition to the time-use outcomes of Table 4, Column (6) reports the effect on an indicator for whether the respondent researched paternity leave policies, the outcome of Table 6, Column (1).

Table B.11: Weekly working hours by thresholds

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	>0	>10	>20	>30	>40	>50	>60
<i>Panel A: Respondents</i>							
Training assignment	0.002 (0.002)	0.015 (0.013)	0.006 (0.016)	0.003 (0.016)	0.016 (0.025)	-0.013 (0.016)	-0.016 (0.011)
Observations	1177	1177	1177	1177	1177	1177	1177
No. clusters	74	74	74	74	74	74	74
Control Mean	0.995	0.912	0.896	0.896	0.630	0.209	0.047
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Spouse</i>							
Training assignment	0.005 (0.015)	0.037* (0.019)	0.061*** (0.022)	0.014 (0.019)	-0.003 (0.013)	0.003 (0.006)	0.008** (0.004)
Observations	994	994	994	994	994	994	994
No. clusters	74	74	74	74	74	74	74
Control Mean	0.697	0.607	0.436	0.322	0.098	0.021	0.002
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes

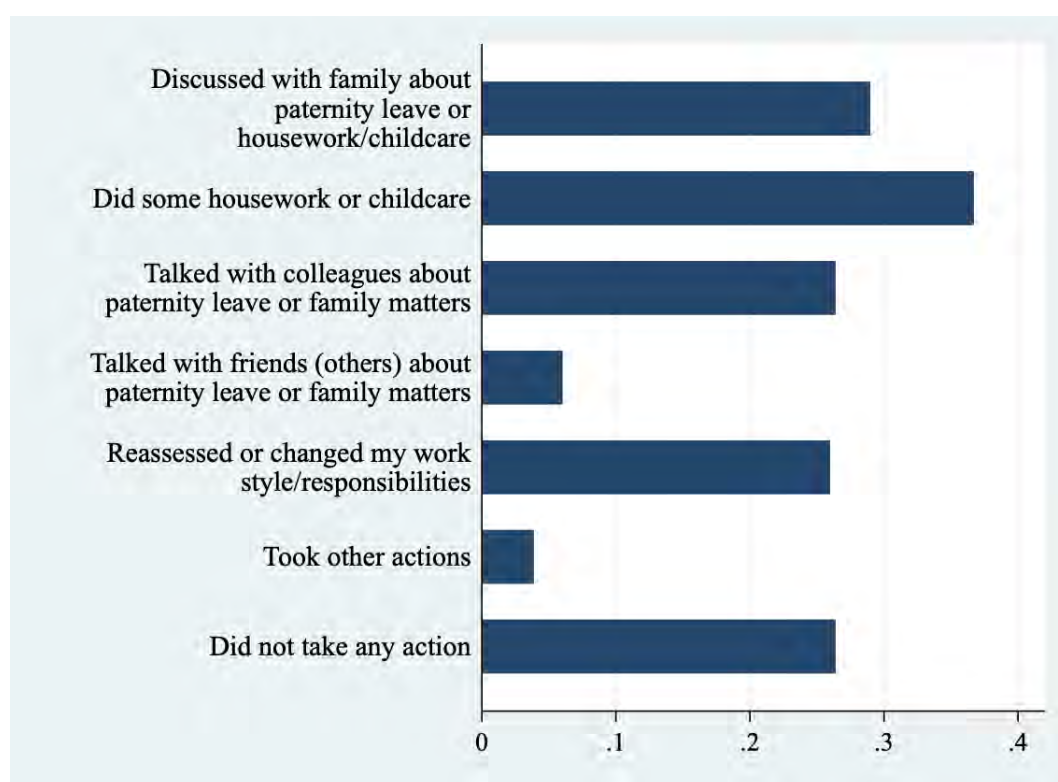
Notes. The dependent variables are indicator variables that take 1 if weekly working hours exceed the threshold indicated in each column and 0 otherwise. Panel A reports results for the respondents and Panel B for their spouses, corresponding to Panels A and B of Table 4. We control for the baseline value of the same indicator variable in all equations. The specification follows that of Table 4. The thresholds are a distributional diagnostic for the working-hours result of Table 4, Column (1), rather than a pre-specified family of outcomes: they trace where in the hours distribution the effect appears. Because the columns are overlapping indicators built from a single variable, the stars are nominal and we apply no multiplicity correction across them; they should not be read as separate tests.

Table B.12: Behavioral Changes: Combined Hours of Respondent and Spouse

	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Weekly Childcare Hours	(5) Daily Housework Hours (Weekday)	(6) Daily Housework Hours (Weekend)	(7) Weekly Housework Hours
<i>Panel A: Pooled effect (no heterogeneity)</i>							
Training assignment	1.037 (1.143)	-0.0573 (0.186)	0.501 (0.327)	0.614 (1.246)	-0.166 (0.178)	-0.0324 (0.192)	-0.830 (1.147)
Observations	994	845	845	845	979	979	979
<i>Panel B: Heterogeneity by Child Presence and Age</i>							
Training assignment $\times$ Young children	0.768 (2.128)	-0.164 (0.443)	0.841 (0.817)	0.827 (3.234)	-0.770** (0.343)	-1.485** (0.609)	-6.733** (2.705)
Training assignment $\times$ Only older children	0.887 (1.693)	0.0534 (0.230)	0.289 (0.497)	0.825 (1.842)	-0.0168 (0.201)	0.414** (0.195)	0.812 (1.212)
Training assignment $\times$ No children	4.644 (4.443)				0.290 (0.407)	0.829** (0.353)	3.178 (2.587)
Observations	994	845	845	845	979	979	979
No. clusters	74	72	72	72	74	74	74
Control Mean (Young children)	60.461	10.492	22.355	97.168	6.902	8.857	52.226
Control Mean (Only older children)	66.342	4.087	8.447	37.328	5.065	5.677	36.682
Control Mean (No children)	65.266				3.589	4.164	26.271
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table uses the sample of respondents who were married at the time of the baseline survey. “Weekly Working Hours” refers to the combined total weekly working hours of the respondent and his spouse. “Daily Childcare Hours (Weekday)” (or “Daily Childcare Hours (Weekend)”) is the total daily hours of the respondent and his spouse spent on childcare on weekdays (or weekends). Similarly, “Daily Housework Hours (Weekday)” (or “Daily Housework Hours (Weekend)”) is the couple’s daily total housework hours on weekdays (or weekends). “Weekly Childcare Hours” (or “Weekly Housework Hours”) refers to the combined total *weekly* hours that the respondent and his spouse spent on childcare (or housework). Panel A reports the pooled intent-to-treat effect of training assignment, and Panel B reports heterogeneity by the presence and age of children, using the subgroup definitions of Table 5. The number of clusters, control means, and PDS LASSO controls reported below Panel B apply to both panels. Because childcare hours are undefined for households without children, the “No children” interaction is omitted in Columns (2)–(4). The other specifications are the same as Table 5.

Figure B.2: Actions taken within 3 days after the training



Notes. In the second follow-up survey, respondents who had participated in the training were asked whether they had taken any action within three days after the training. This figure shows the fraction of respondents who took each action. Multiple choices were allowed.

B.2.3 Paternity Leave Intentions

In this appendix, which reports the detailed results supporting Section 5.1.3, we examine whether the training intervention influenced participants' intentions to take paternity leave in the future. In both the baseline and follow-up surveys, respondents were asked, "Do you plan or intend to take childcare leave in the future?"¹⁴ The outcome variable equals one if the respondent reports an intention or plan to take paternity leave, or is currently doing so. Because this question closely relates to expecting to have (another) child, changes in fertility expectations can also affect the outcome.

The analysis focuses on men aged 50 or younger, who are more likely to be of reproductive age. Consistent with this restriction, Figure B.3 shows that the average intent to take paternity leave is close to zero among older respondents.

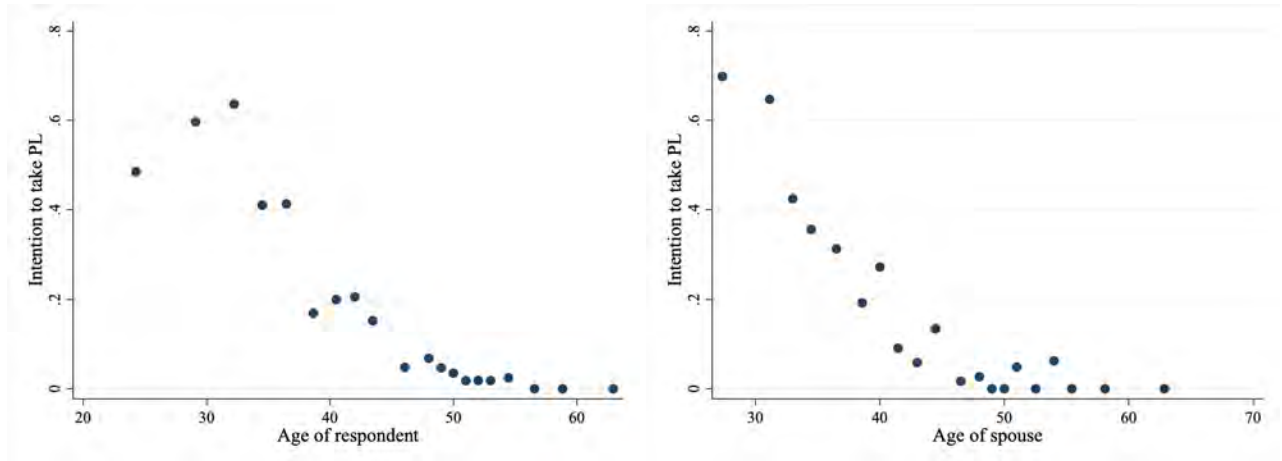
Table 6 reports the estimated ITT effects (see Table B.13 for the corresponding IV estimates). Column (3) shows a small, negative average effect of the training on reported intentions ($q = 0.151$). The outcome is closely tied to future fertility plans, and its relatively low prevalence partly reflects sample composition: many participants are in their forties and may not expect to have additional children.

The binary measure nonetheless masks a shift in how participants reasoned about future leave-taking. Table B.14 shows that the training increased the share of participants selecting the response "I have finished raising children, do not plan to have or raise children, or am currently raising children but do not plan to take paternity leave in the future." At the same time, it reduced the share choosing "I do not plan to take paternity leave, even if I have a (new) child in the future." One interpretation is that the program's emphasis on active paternal involvement made the time commitment and effort required for childcare more salient, leading some participants to reassess their parenting plans rather than reinforcing outright opposition to leave-taking.

Workplace context also appears to moderate the effects of the training. Column (4) of Table 6 stratifies respondents by the number of coworkers available to cover their tasks. The estimates show a clear divergence: the point estimate is positive among those with more potential substitutes and negative and significant among those with fewer; the two interaction coefficients differ at $p = 0.038$ ($p = 0.043$ in the corresponding IV specification of Table B.13). This pattern indicates that even when attitudes toward paternity leave become more positive, the perceived feasibility of taking leave continues to depend on practical constraints such as staffing flexibility and task coverage.

¹⁴Appendix A provides the full question wording and response options.

Figure B.3: Age and intention to take paternity leave



Notes. This figure shows the relationship between the respondent's intention to take paternity leave and the respondent's age (left) and spouse's age (right). Using data from the baseline survey, we plot binned scatter plots, where each point represents the share of respondents intending to take paternity leave within one of 20 equally sized age bins.

Table B.13: Impacts of training on information-seeking and intention to take paternity leave (IV estimates)

	Researched about PL		Intention to take PL	
	(1)	(2)	(3)	(4)
Training participation	0.155*** (0.052)	0.173*** (0.049)	-0.052 (0.048)	
Anderson q-value	[0.006]		[0.163]	
Training participation $\times$ Above-median replaceability				0.079 (0.069)
Training participation $\times$ Below-median replaceability				-0.126* (0.070)
Observations	1211	796	775	768
No. clusters	74	69	69	69
Control Mean	0.189	0.239	0.251	0.251
PDS LASSO Control	Yes	Yes	Yes	Yes
P-value: coefficient heterogeneity				0.043

Notes. The specification follows that of Table 6, except that here we report instrumental-variable (IV) estimates of the effect of training participation, instrumenting participation (and, in Column (4), its interactions with the indicated variables) with training assignment (and its interactions). The coefficients rescale the assignment effect by the participation rate. As discussed in Section 3.5, exposure to trained managers may violate the exclusion restriction required to read them as the effect of a respondent's own attendance; they are best read as the effect of the training as experienced in a treated office, scaled by participation.

Table B.14: Impacts of training on intention to take paternity leave: variables spanning full choice set for the question

	(1)	(2)	(3)	(4)
	Have an intention to take PL	Will not take PL even if I have a (new) child	Finished raising or not planning to raise children or take PL	Unsure
Training assignment	-0.023 (0.020)	-0.029** (0.012)	0.069** (0.030)	-0.016 (0.029)
Observations	775	775	775	775
No. clusters	69	69	69	69
Control Mean	0.251	0.045	0.456	0.248
PDS LASSO Control	Yes	Yes	Yes	Yes

Notes. The dependent variables are constructed based on the 1st follow-up survey question asking “Do you plan or intend to take childcare leave in the future? Please select one that applies: 1. I am currently taking paternity leave; 2. I plan or hope to take paternity leave within the next six months; 3. I plan or hope to take paternity leave in the future, but not sure when; 4. I do not plan to take paternity leave, even if I have a (new) child in the future; 5. I have finished raising children, do not plan to have or raise children, or am currently raising children but do not plan to take paternity leave in the future; 6. Not sure; 7. Prefer not to answer.” Column (1) is an indicator variable equal to 1 if the respondent selected option 1, 2, or 3. Column (2) is an indicator variable equal to 1 if the respondent selected option 4. Column (3) is an indicator variable equal to 1 if the respondent selected option 5. Column (4) is an indicator variable equal to 1 if the respondent selected option 6. We exclude respondents who selected option 7. The rest of the specification is the same as Table 6. As in Column (3) of Table 6, we focus on the sample of respondents who are aged 50 or younger. This decomposition is exploratory. The four outcomes are mutually exclusive categories from a single survey item, and the significance stars report nominal p -values without a separate multiple-testing adjustment across categories.

B.2.4 Workplace Outcomes

Table B.15: Workplace practices

	(1) Num. workers for my task today	(2) Num. workers for my task during one-month absence	(3) Meeting hours	(4) Days work from home
Training assignment	-0.0215 (0.0780)	-0.0089 (0.0871)	-0.1434 (0.4887)	-0.0274 (0.0178)
Observations	1095	1095	1196	1198
No. clusters	72	72	74	74
Control Mean	2.060	2.024	4.865	0.240
PDS LASSO	Yes	Yes	Yes	Yes

Notes. This table reports the effect of the training intervention on workplace practice outcomes. Column (1) is the number of coworkers available to cover the respondent's tasks if he suddenly had to leave work early on a given day, and Column (2) is the number available to cover them throughout a one-month absence beginning three months later. Column (3) is weekly meeting hours. Column (4) is the number of days per week working from home. We estimate the intent-to-treat (ITT) effect of training assignment. Standard errors in parentheses are clustered at the randomization-unit level. All equations control for randomization strata fixed effects, baseline value of the outcome variable, and control variables selected by PDS LASSO.

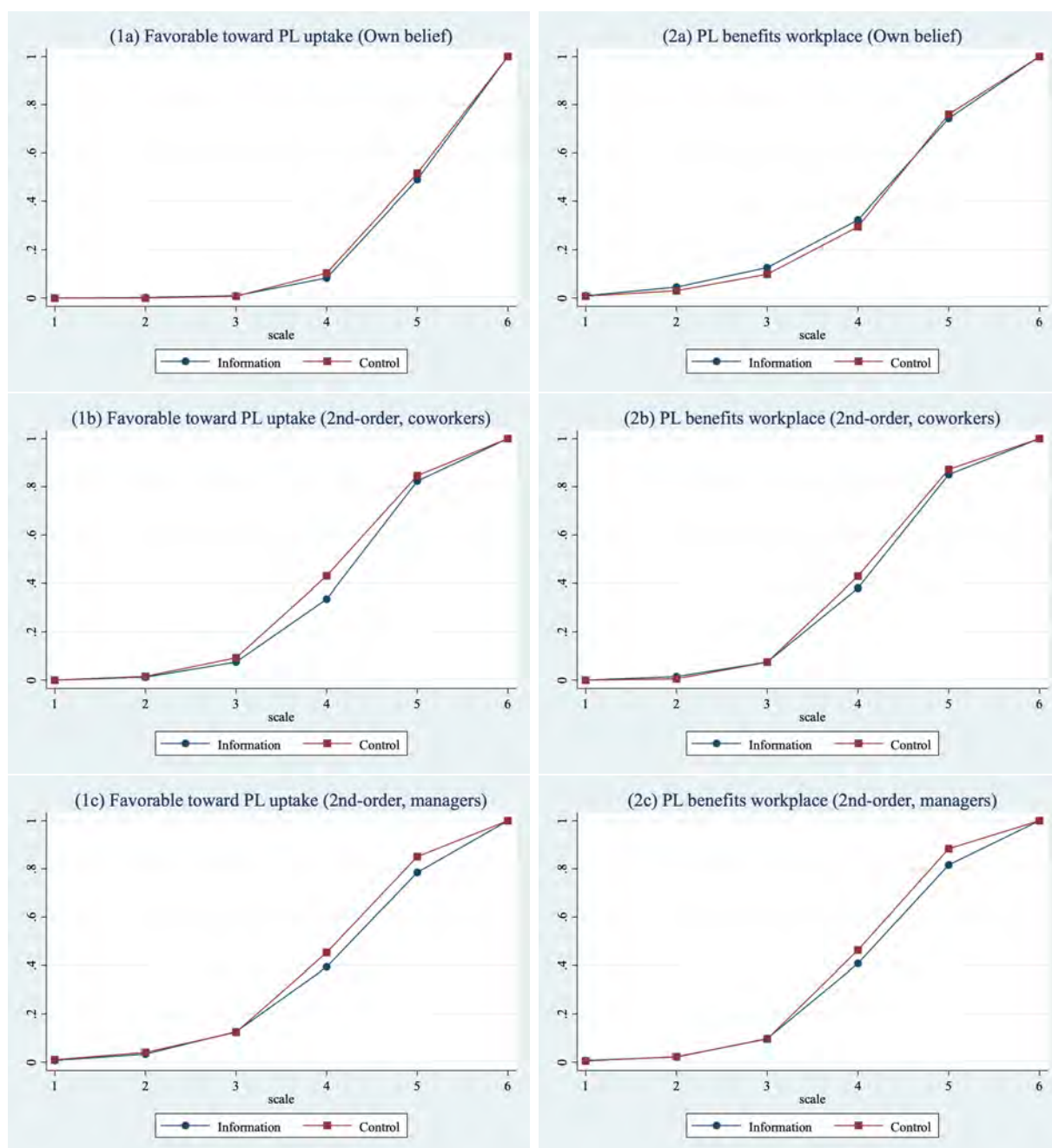
Table B.16: Workplace relationships

	(1) Manager knows family composition	(2) Psychological safety	(3) Coworker and supervisor relationships	(4) Recommend firm to jobseeker
Training assignment	-0.002 (0.014)	0.004 (0.056)	0.029 (0.066)	0.007 (0.025)
Observations	1194	1203	1202	1194
No. clusters	74	74	74	74
Control Mean	0.783	4.022	3.787	0.495
PDS LASSO	Yes	Yes	Yes	Yes

Notes. This table reports the effect of the training intervention on workplace relationship outcomes. Column (1) is an indicator variable equal to 1 if the respondent reports that their manager is aware of their family composition. Column (2) is the respondent's perceived psychological safety in the workplace, measured on a scale. Column (3) is the respondent's assessment of the quality of relationships with coworkers and the direct supervisor, measured on a scale. Column (4) is an indicator variable equal to 1 if the respondent would recommend the firm to a job seeker. We estimate the intent-to-treat (ITT) effect of training assignment. Standard errors in parentheses are clustered at the randomization-unit level. All equations control for randomization strata fixed effects, baseline value of the outcome variable, and control variables selected by PDS LASSO.

B.3 Results: Information Intervention

Figure B.4: Cumulative Distribution: Information Treatment vs. Control



Notes. Data on first- and second-order beliefs measured in the first follow-up survey, in the same session in which the information was shown. Axes and panel ordering follow Figure B.1: the horizontal axis is the six-point response scale, running from 1 (least supportive) to 6 (most supportive), the vertical axis is the share of respondents at or below that scale point, and a curve lying below and to the right of the other indicates more supportive responses.

Table B.17: Impacts of information on intention to take paternity leave: variables spanning full choice set for the question

	(1)	(2)	(3)	(4)
	Have an intention to take PL	Will not take PL even if I have a (new) child	Finished raising or not planning to raise children or take PL	Unsure
Information	-0.007 (0.031)	0.004 (0.013)	0.036 (0.037)	-0.021 (0.037)
Observations	425	425	425	425
No. clusters	57	57	57	57
Control Mean	0.309	0.055	0.313	0.323
PDS LASSO Control	Yes	Yes	Yes	Yes

Notes. The dependent variables are constructed based on the first follow-up survey question asking “Do you plan or intend to take childcare leave in the future? Please select one that applies: 1. I am currently taking paternity leave; 2. I plan or hope to take paternity leave within the next six months; 3. I plan or hope to take paternity leave in the future, but not sure when; 4. I do not plan to take paternity leave, even if I have a (new) child in the future; 5. I have finished raising children, do not plan to have or raise children, or am currently raising children but do not plan to take paternity leave in the future; 6. Not sure; 7. Prefer not to answer.” Column (1) is an indicator variable equal to 1 if the respondent selected option 1, 2, or 3. Column (2) is an indicator variable equal to 1 if the respondent selected option 4. Column (3) is an indicator variable equal to 1 if the respondent selected option 5. Column (4) is an indicator variable equal to 1 if the respondent selected option 6. We exclude respondents who selected option 7. The rest of the specification is the same as Column (7) in Table 7. As in Column (7) in Table 7, we focus on the sample of respondents who are aged 50 or younger. This decomposition is exploratory. The four outcomes are mutually exclusive categories from a single survey item, and the significance stars report nominal p -values without a separate multiple-testing adjustment across categories.

Table B.18: Cross-Equation Test of Training and Information Treatment Effects on Information Seeking Behavior

	Training (1)	Information (2)	Difference (1)–(2)
<i>Panel A: Training ITT vs Information ATE</i>			
Coefficient	0.0614	0.0266	0.0349
Standard error	(0.0230)	(0.0322)	(0.0381)
p-value (diff.)			0.3605
Observations	1,211	411	
Clusters	74	58	
<i>Panel B: Training IV vs Information ATE</i>			
Coefficient	0.1553	0.0266	0.1288
Standard error	(0.0529)	(0.0322)	(0.0581)
p-value (diff.)			0.0268
Observations	1,211	411	
Clusters	74	58	

Notes: This table reports cross-equation tests of the null hypothesis that the training and information treatments have equal effects on the probability that a respondent researches about paternity leave. Because the two outcomes are observed at different survey waves (training → first follow-up; information → second follow-up), each equation is estimated on its own sample, and the joint covariance matrix is obtained by seemingly unrelated estimation (**suest**) with standard errors clustered at the office level to account for units appearing in both wave samples.

Panel A compares the training intent-to-treat (ITT) effect with the information average treatment effect (ATE). Panel B replaces the training ITT with the corresponding IV estimate, obtained by instrumenting actual training participation with training assignment, which rescales the assignment effect by the participation rate; as discussed in Section 3.5, it is not the effect of a respondent’s own attendance in isolation, so Panel B is a descriptive benchmark rather than a comparison of treatment strength per person treated. To make the two-stage-least-squares equation compatible with **suest**, we implement 2SLS via the control-function representation.

PDS LASSO selected controls from the main specification are used as control variables. Standard errors are in parentheses. The p -value in the “Difference” column corresponds to the Wald test of equality of the two coefficients based on the **suest** cluster-robust variance matrix. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.19: Impact of information treatment on outcomes measured in second follow-up survey

	(1) Psychological Safety	(2) Coworker and supervisor relationships	(3) Recommend firm to jobseekers
Information	0.0777 (0.0632)	0.148 (0.0985)	-0.00735 (0.0391)
Observations	410	410	405
Control Mean	3.979	3.782	0.539
PDS LASSO Control	Yes	Yes	Yes

Notes. This table reports the effect of the information intervention on workplace outcomes measured in the second follow-up survey. Column (1) is the respondent's perceived psychological safety in the workplace, measured on a scale. Column (2) is the respondent's assessment of the quality of relationships with coworkers and the direct supervisor, measured on a scale. Column (3) is an indicator variable equal to 1 if the respondent would recommend the firm to a job seeker. The coefficients represent the Average Treatment Effect (ATE) estimated by OLS. Standard errors in parentheses are clustered at the level of offices. All equations control for firm group fixed effects, baseline value of the outcome variable, and control variables selected by PDS LASSO.

Table B.20: Interaction of information and training on beliefs and intention

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)		Intention to take PL
	(1) Favorable toward PL uptake	(2) PL benefits workplace	(3) Favorable toward PL uptake	(4) PL benefits workplace	(5) Favorable toward PL uptake	(6) PL benefits workplace	(7)
Training assignment	-0.046 (0.047)	0.295*** (0.090)	0.102 (0.089)	0.118 (0.108)	-0.131* (0.075)	0.116 (0.104)	0.038 (0.036)
Information	0.002 (0.072)	0.108 (0.086)	0.159** (0.068)	0.009 (0.087)	0.089 (0.091)	0.127 (0.088)	0.016 (0.042)
Training assignment × Information	0.094 (0.081)	-0.232* (0.136)	-0.100 (0.118)	0.076 (0.119)	0.121 (0.139)	-0.051 (0.118)	-0.045 (0.062)
Observations	589	588	579	579	587	587	425
Control Mean	5.329	4.523	4.516	4.539	4.542	4.439	0.295
No. clusters	65	65	65	65	65	65	57
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table examines interactions between the training and information interventions on outcomes measured in the first follow-up survey. The specification is the same as Table 7 except that the training-assignment indicator and its interaction term with the information treatment indicator are added.

C Details of Robustness Checks

C.1 Details of Social Desirability Scale

The SDS was included in the second follow-up survey and consisted of ten dichotomous (“Yes” or “No”) questions drawn from a Japanese adaptation of the Crowne and Marlowe (1960) instrument developed by Kitamura and Suzuki (1986). This version was refined to reflect culturally accepted but uncommon behaviors in Japan and excludes items related to mental health symptoms, helping focus the measure on the inclination to present oneself favorably rather than on psychological well-being. We calculated a Social Desirability Score by summing the number of socially desirable responses (higher scores indicate a stronger tendency toward socially desirable reporting) and defined a binary indicator, High SDS, equal to one if the score exceeds the sample median. Appendix Table C.1 lists the ten items used.

Table C.1: Japanese Module of Social Desirability Scale

If I could enter a movie theater without paying and no one saw me, I would probably do it.	Yes = 0, No = 1
Sometimes, I resent authority figures (bosses, teachers, parents, etc.), even when I know they are right.	Yes = 0, No = 1
I have pretended to be ill.	Yes = 0, No = 1
I don't mind admitting when I don't know something.	Yes = 1, No = 0
I always behave politely, even to people I don't like.	Yes = 1, No = 0
I have never considered shifting the blame for something I've done onto someone else.	Yes = 1, No = 0
I have never felt annoyed when someone has done me a favor and made me feel indebted.	Yes = 1, No = 0
I have never found it troublesome when someone has a different way of thinking.	Yes = 1, No = 0
I have seldom wanted to scold someone harshly.	Yes = 1, No = 0
I sometimes feel irritated by requests from others who take my goodwill for granted.	Yes = 0, No = 1

Notes. Kitamura and Suzuki (1986)'s short-form module developed from Crowne and Marlowe (1960).

Table C.2: Impact of training on beliefs: interaction with SDS

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)	
	(1) Favorable toward PL uptake	(2) PL benefits workplace	(3) Favorable toward PL uptake	(4) PL benefits workplace	(5) Favorable toward PL uptake	(6) PL benefits workplace
Training assignment	0.050 (0.037)	0.187** (0.085)	0.034 (0.064)	0.115 (0.079)	-0.087 (0.071)	0.012 (0.060)
Training assignment × High SDS	-0.045 (0.070)	-0.102 (0.127)	-0.026 (0.109)	-0.020 (0.101)	-0.071 (0.103)	0.005 (0.077)
High SDS	0.041 (0.046)	0.152* (0.078)	0.003 (0.076)	0.058 (0.077)	0.110* (0.058)	0.089 (0.056)
Observations	882	880	872	872	880	879
No. clusters	69	69	69	69	69	69
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	5.294	4.660	4.629	4.550	4.616	4.598

Notes. This table shows a robustness check for the impact of the training intervention on beliefs to address concerns about social desirability bias. Using the Social Desirability Score (SDS), which measures the respondent's tendency to provide socially favorable answers, we construct an indicator variable, High SDS, which equals one if the respondent's SDS is above the sample median, and zero otherwise. As a robustness check, we modify the main specification in Table 3 by adding the High SDS indicator and its interaction with training assignment; the coefficients are intent-to-treat effects.

Table C.3: Impact of information on beliefs: interaction with SDS

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Favorable toward PL uptake	PL benefits workplace	Favorable toward PL uptake	PL benefits workplace	Favorable toward PL uptake	PL benefits workplace
Information	0.247*** (0.068)	0.074 (0.146)	0.117 (0.093)	0.105 (0.091)	0.190** (0.094)	0.246*** (0.079)
Information $\times$ High SDS	-0.338*** (0.087)	-0.058 (0.185)	-0.027 (0.127)	-0.094 (0.095)	0.080 (0.154)	-0.090 (0.145)
High SDS	0.224*** (0.075)	0.096 (0.118)	0.013 (0.091)	0.122 (0.095)	0.030 (0.085)	0.208** (0.088)
Observations	415	414	409	408	415	414
No. clusters	58	58	58	58	58	58
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	5.332	4.692	4.624	4.595	4.466	4.464

Notes. This table shows a robustness check for the impact of the information intervention on beliefs to address concerns about social desirability bias. Using the Social Desirability Score (SDS), which measures the respondent's tendency to provide socially favorable answers, we construct an indicator variable, High SDS, which equals one if the respondent's SDS is above the sample median, and zero otherwise. As a robustness check, we modify the main specification in Table 7 by including the information treatment indicator, the High SDS indicator, and their interaction term. Because the SDS was measured after the information intervention, this interaction should be interpreted descriptively rather than as heterogeneity by a predetermined characteristic.

Table C.4: Behavioral changes: interaction with Social Desirability Score

<i>Panel A: Own behaviors</i>						
	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)	(6) Researched about PL
Training assignment	0.0874 (1.188)	-0.166 (0.120)	0.467 (0.333)	-0.0487 (0.0786)	-0.132 (0.168)	0.0607 (0.0399)
Training assignment $\times$ High SDS	-0.313 (1.505)	0.196 (0.201)	-0.267 (0.494)	0.0815 (0.119)	0.277 (0.301)	0.00192 (0.0522)
High SDS	0.338 (1.258)	0.198 (0.151)	0.0807 (0.380)	-0.00964 (0.0851)	-0.0664 (0.180)	0.00925 (0.0238)
Observations	859	639	639	852	852	874
No. clusters	69	66	66	69	69	69
Control Mean	43.492	1.394	5.771	1.021	2.277	0.208
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Spouse's behaviors</i>						
	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)	
Training assignment	1.549 (1.100)	0.185 (0.286)	-0.111 (0.421)	-0.267 (0.303)	-0.00860 (0.217)	
Training assignment $\times$ High SDS	-0.420 (1.601)	0.0323 (0.461)	0.399 (0.681)	0.142 (0.399)	0.0800 (0.352)	
High SDS	0.694 (1.406)	-0.119 (0.323)	-0.334 (0.568)	-0.391 (0.257)	0.115 (0.225)	
Observations	724	616	616	718	718	
No. clusters	68	65	65	67	67	
Control Mean	19.757	4.661	7.117	4.457	4.062	
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	

Notes. This table shows a robustness check for the impact of the training intervention on behaviors to address concerns about social desirability bias. Using the Social Desirability Score (SDS), which measures the respondent's tendency to provide socially favorable answers, we construct an indicator variable, High SDS, which equals one if the respondent's SDS is above the sample median, and zero otherwise. As a robustness check, we modify the main specification in Table 4 by adding the High SDS indicator and its interaction with training assignment; the coefficients are intent-to-treat effects. In addition to the time-use outcomes of Table 4, Column (6) of Panel A reports the effect on an indicator for whether the respondent researched paternity leave policies, the outcome of Table 6, Column (1). Because this outcome has no spouse counterpart, Panel B reports five columns.

C.2 Subsample Analysis of the Organization with the Highest Program Sign-up Rate

Table C.5: Impact of Training on Beliefs: Subsample of the Organization with the Highest Program Sign-up Rate

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)	
	(1) Favorable toward PL uptake	(2) PL benefits workplace	(3) Favorable toward PL uptake	(4) PL benefits workplace	(5) Favorable toward PL uptake	(6) PL benefits workplace
Training assignment	0.067*** (0.023)	0.193** (0.076)	0.117 (0.088)	0.193*** (0.070)	-0.086 (0.069)	0.016 (0.061)
Observations	606	604	596	595	600	601
No. clusters	34	34	34	34	34	34
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	5.110	4.456	4.410	4.386	4.302	4.314

Notes. In this table, we use the sample of the participating organization with the highest program sign-up rate (44 percent). The rest of the specification is the same as that of Table 3.

Table C.6: Impact of Training on Behaviors: Subsample of the Organization with the Highest Program Sign-up Rate

<i>Panel A: Own behaviors</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
	Weekly Working Hours	Daily Childcare Hours (Weekday)	Daily Childcare Hours (Weekend)	Daily Housework Hours (Weekday)	Daily Housework Hours (Weekend)	Researched about PL
Training assignment	0.32 (1.17)	0.06 (0.15)	0.35 (0.33)	-0.10 (0.07)	-0.12 (0.12)	0.03 (0.03)
Observations	580	406	406	574	574	599
No. clusters	34	33	33	34	34	34
Control Mean	43.814	1.169	4.998	0.978	2.170	0.121
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
<i>Panel B: Spouse's behaviors</i>						
	(1)	(2)	(3)	(4)	(5)	
	Weekly Working Hours	Daily Childcare Hours (Weekday)	Daily Childcare Hours (Weekend)	Daily Housework Hours (Weekday)	Daily Housework Hours (Weekend)	
Training assignment	1.76* (0.98)	0.07 (0.22)	-0.07 (0.30)	-0.32 (0.24)	0.06 (0.20)	
Observations	463	389	389	459	459	
No. clusters	34	33	33	34	34	
Control Mean	20.891	3.956	6.312	4.399	4.020	
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	

Notes. In this table, we use the sample of the participating organization with the highest program sign-up rate (44 percent). The rest of the specification is the same as that of Table 4. In addition to the time-use outcomes of Table 4, Column (6) of Panel A reports the effect on an indicator for whether the respondent researched paternity leave policies, the outcome of Table 6, Column (1). Because this outcome has no spouse counterpart, Panel B reports five columns.

Table C.7: Impact of information on beliefs: Subsample of the Organization with the Highest Program Sign-up Rate

	1st-order belief (Own belief)		2nd-order belief (Coworkers)		2nd-order belief (Managers)	
	(1) Favorable toward PL uptake	(2) PL benefits workplace	(3) Favorable toward PL uptake	(4) PL benefits workplace	(5) Favorable toward PL uptake	(6) PL benefits workplace
Information	0.056 (0.069)	0.017 (0.085)	0.111 (0.086)	0.008 (0.078)	0.132 (0.125)	0.123* (0.074)
Observations	234	233	229	229	234	234
No. clusters	29	29	29	29	29	29
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean	5.221	4.639	4.575	4.583	4.369	4.352

Notes. In this table, we use the sample of the participating organization with the highest program sign-up rate (44 percent). The rest of the specification is the same as that of Table 7.

D Attrition

Our main analytic sample comprises the 1,225 male employees who answered both the baseline and first follow-up surveys, out of 1,444 men who answered the baseline. Attrition between these two waves is thus 15.2 percent. Because the intent-to-treat estimand is identified by comparing offices as assigned, the response margin that could threaten internal validity is selective attrition, not selection into sign-up or into participation: if the decision to answer the follow-up depends on training assignment, the ITT can be biased even though assignment is random. This appendix documents the extent of such attrition and shows that it does not account for our results. We first describe who leaves the sample, then present inverse-probability-weighted (IPW) estimates as our primary correction, and finally report Lee (2009) bounds, which dispense with the selection-on-observables assumption at the cost of wider intervals. Section D.4 turns to the second response margin, attrition between the first and second follow-up surveys, which is the margin relevant for the information experiment.

D.1 Who attrits

Two facts locate the attrition. First, comparing the baseline-sample balance (Table D.1) with the analytic-sample balance (Table 1) isolates what attrition adds. On the full baseline sample of 1,444 men, the treatment–control difference in the manager share is 6.5 percentage points and statistically insignificant ($p = 0.103$); in the analytic sample it widens to 7.6 points and becomes significant ($p < 0.05$). The imbalance that our estimation controls for is therefore in substantial part a product of differential attrition rather than of randomization. The remaining baseline characteristics (age, marital status, parental status, working hours, and own and second-order beliefs) are balanced on the baseline sample, as they are on the analytic sample.

Second, we regress an indicator for answering the first follow-up on training assignment and baseline covariates (Table D.2). Assignment to training lowers the response rate by 7.7 points on average (Column 1), but the shortfall is far from uniform. Interacting assignment with age (Column 4) shows that it is concentrated among employees over fifty: the main training coefficient is -0.247 , while the interactions with the younger age group (age 50 or below) offset most of it ($+0.106$). Managers respond at higher rates regardless of assignment. The pattern is one in which employees for whom the training was most personally relevant were the least likely to drop out.

This matters because our behavioral results are concentrated in exactly the groups that do not exhibit differential attrition. Restricting the response regression to men with at least one young child (aged five or below), the households that drive the childcare and spousal results, leaves no detectable assignment gap in response (-0.038 , s.e. 0.032; Table D.3, Panel A). The same holds for men aged fifty or younger (-0.052 , s.e. 0.031;

Panel B). We find no statistically detectable assignment-related response gap in either the young-child or age-50-or-younger subsample.

D.2 Inverse probability weighting

Our primary correction reweights each respondent by the inverse of his estimated probability of responding, so that respondents of a type that answered at low rates count for more, and the weighted respondent sample reproduces the full randomized sample (Wooldridge, 2007). We focus on six outcomes highlighted in the main analysis: beliefs regarding the benefits of paternity leave on the workplace (the respondent’s own belief and second-order beliefs about coworkers), whether the respondent researched paternity leave, and time-use variables (the respondent’s hours spent on childcare on weekends, and the spouse’s working hours and daily housework hours). For the time-use variables, we focus on the sample restricted to respondents with at least one young child (aged 0–5).

We estimate the response probability separately within the training and control groups from the covariate specification of Table D.2 (Columns 5–6) and Table D.3 (Panel A, Columns 4–5), and reestimate the main ITT specification weighting each observation by the inverse of its fitted probability. Because weights explode as fitted probabilities approach zero, we winsorize the fitted probabilities at the first and ninety-ninth percentiles before inverting them, so that a handful of low-probability respondents do not dominate the weighted estimate. All other elements of the specification are retained from the main analysis: the PDS LASSO-selected controls, the baseline outcome, strata fixed effects, and office-level clustering.

Table D.4 reports the results. Across all six outcomes the weighted estimates are close to their unweighted counterparts in both magnitude and significance: for the first-order belief that leave benefits the workplace, 0.175 against a main estimate of 0.182; for the second-order belief about coworkers, 0.110 against 0.119; for having researched leave policies, 0.062 against 0.061; and, among households with young children, 0.920 against 1.031 for weekend childcare, 2.140 against 1.982 for spouses’ weekly hours, and -0.972 against -0.936 for spouses’ weekend housework. Reweighting to undo the observable component of selective response leaves the estimates essentially unchanged, which is direct evidence that the attrition documented above is not driving the results.

D.3 Lee bounds

As a check that does not rest on selection being explained by observables, we compute Lee (2009) bounds. These trim the less-attrited assignment group to equalize response rates and report the range of treatment effects consistent with any values the trimmed observations might have taken, so they are worst-case by construction and correspondingly wide. Because Lee bounds are defined for a treatment indicator without additional

covariates, we apply them to the outcome after residualizing on the controls of our main specification, so that the bounds are computed on the same covariate-adjusted variation that the main estimates use.¹⁵ Monotonicity, the assumption that assignment moves response in one direction, is consistent with the uniformly non-positive assignment effect on response documented above.

The bounds behave as the attrition pattern predicts (Table D.5). For the three time-use outcomes, restricting to the young-child sample, where response does not depend on assignment, yields bounds that exclude zero: $[0.702, 1.332]$ for weekend childcare, $[1.830, 3.076]$ for spouses' weekly hours, and $[-1.119, -0.456]$ for spouses' weekend housework, each preserving the sign of the corresponding point estimate, though the bounds are estimated imprecisely. The bounds for the information-seeking outcome also exclude zero: $[0.027, 0.077]$. For the belief outcomes the bounds are wider: for both the first-order and the second-order belief that leave benefits the workplace, the worst-case lower bound includes zero. We do not read this as evidence against those effects: the IPW estimates, which use the response model rather than discarding it, leave them intact. The second-order belief result in particular is one we already treat with caution in the main text, and the bounds do not sharpen it.

D.4 Attrition in the second follow-up survey

The information experiment was embedded in the first follow-up survey, and a subset of its outcomes (the intention to take leave and whether the respondent researched leave policies) are observed in the second follow-up. The response margin relevant for those estimates is therefore attrition between the first and the second follow-up, among the 589 non-managerial men who answered the first follow-up and entered the information experiment's randomization. Of these, 419 answered the second follow-up, so attrition on this margin is 28.9 percent, larger than on the first-follow-up margin. As above, what would threaten internal validity is not the level of attrition but its dependence on assignment.

Table D.6 gives no evidence of such dependence. Assignment to the information treatment leaves the probability of answering the second follow-up essentially unchanged, with or without baseline covariates (Columns 1–3), and interacting assignment with the demographic covariates (Column 4) uncovers no heterogeneity of the kind found for training. What does predict response is unrelated to assignment: younger men and men with shorter baseline working hours answer at higher rates (Columns 2–3). The sample contains no managers by construction, since the information treatment was administered only to non-managerial respondents.

¹⁵We partial out the main specification's controls from the outcome, holding the PDS LASSO-selected set fixed, and compute the bounds on the resulting residual. Standard errors are obtained by bootstrap with 1,000 replications.

Because response on this margin does not appear to depend on assignment, the case for correcting is weaker than on the training margin, but we repeat the reweighting exercise of Section D.2. We estimate the response probability separately within the information and no-information groups from the covariate specification of Table D.6 (Columns 5–6), winsorize the fitted probabilities at the first and ninety-ninth percentiles, and reweight each observation by the inverse of its winsorized probability, retaining all other elements of the main specification. Table D.7 reports the results: the weighted estimates for both outcomes remain small and statistically insignificant, close to their unweighted counterparts. The absence of a detectable effect of the information treatment is thus not an artifact of who chose to answer the second survey.

We do not report Lee (2009) bounds for this margin. Those bounds are informative when a point estimate is significant and the question is whether worst-case selection could overturn it; here the estimates to be defended are nulls, which worst-case trimming can only widen further. The trimming proportion would in any case be governed by an assignment–response difference that is itself statistically indistinguishable from zero.

Table D.1: Baseline balance on the full baseline sample

Variable	(1)	(2)	(3)	(4)
	Mean	Mean	Difference	No. Obs
	Control	Treatment	Treatment-Control	
Age	45.362 (10.607)	44.603 (10.240)	-0.817 (0.286)	1,444
Married	0.849 (0.358)	0.857 (0.350)	0.009 (0.698)	1,444
Manager	0.427 (0.495)	0.489 (0.500)	0.065 (0.103)	1,444
Has at least one child	0.764 (0.425)	0.753 (0.432)	-0.009 (0.693)	1,444
Own belief: Favorable toward paternity leave uptake	5.182 (0.760)	5.200 (0.778)	0.023 (0.619)	1,444
Own belief: Paternity leave benefits the workplace	4.575 (1.094)	4.567 (1.059)	-0.009 (0.875)	1,444
2nd-order belief: Coworkers' favorable attitudes toward PL uptake	4.499 (0.958)	4.522 (0.894)	0.029 (0.650)	1,442
2nd-order belief: Managers' favorable attitudes toward PL uptake	4.417 (1.119)	4.428 (1.032)	0.022 (0.759)	1,442
2nd-order belief: Coworkers' belief that PL benefits workplace	4.376 (0.881)	4.308 (0.902)	-0.068 (0.196)	1,439
2nd-order belief: Managers' belief that PL benefits workplace	4.373 (0.949)	4.291 (0.957)	-0.079 (0.278)	1,442
Weekly working hours	44.886 (15.341)	44.820 (13.608)	-0.108 (0.898)	1,423
Daily childcare hours (weekday)	1.337 (1.571)	1.511 (1.696)	0.197* (0.061)	1,056
Daily childcare hours (weekend)	6.373 (5.916)	6.479 (5.867)	0.166 (0.672)	1,056
Daily housework hours (weekday)	0.955 (1.263)	0.906 (1.033)	-0.039 (0.473)	1,407
Daily housework hours (weekend)	2.126 (2.240)	1.981 (1.896)	-0.139 (0.290)	1,407
Spouse's weekly working hours	21.776 (18.263)	22.102 (17.335)	0.419 (0.639)	1,207
Spouse's daily childcare hours (weekday)	5.007 (4.745)	5.054 (4.663)	0.067 (0.844)	1,024
Spouse's daily childcare hours (weekend)	7.822 (5.971)	7.763 (5.841)	-0.011 (0.972)	1,024
Spouse's daily housework hours (weekday)	4.191 (3.716)	3.946 (3.176)	-0.261 (0.230)	1,192
Spouse's daily housework hours (weekend)	4.144 (3.579)	3.870 (2.987)	-0.289 (0.125)	1,192
Researched about paternity leave	0.183 (0.387)	0.166 (0.372)	-0.015 (0.557)	1,439
Intention to take paternity leave	0.192 (0.394)	0.179 (0.384)	-0.011 (0.612)	1,390

Notes. The sample is all male employees who answered the baseline survey with a non-missing office identifier ($N = 1,444$), that is, before attrition to the first follow-up. Columns (1) and (2) report the mean and standard deviation (in parentheses) by training assignment. Column (3) reports the treatment-control difference estimated from a regression of each variable on an assignment indicator, controlling for randomization-strata fixed effects, with the p -value (in parentheses) based on standard errors clustered at the randomization-unit level. Column (4) reports the number of observations. This table is the pre-attrition counterpart of Table 1.

Table D.2: Determinants of first follow-up response by training assignment

	Answer 1st Follow-Up					
	(1)	(2)	(3)	(4)	(5) Training Sample	(6) No Training Sample
Training assignment	-0.077** (0.033)	-0.086** (0.033)	-0.086** (0.033)	-0.247*** (0.074)		
Age ≤ 50		0.075*** (0.023)	0.060*** (0.020)	0.006 (0.016)	0.104*** (0.032)	0.021 (0.015)
Manager		0.076** (0.030)	0.070** (0.031)	0.057** (0.026)	0.082 (0.050)	0.054* (0.029)
Married		0.011 (0.035)	0.004 (0.036)	-0.032 (0.046)	0.039 (0.056)	-0.025 (0.048)
Has at least one child		0.030 (0.021)	0.026 (0.025)	0.011 (0.029)	0.038 (0.039)	0.007 (0.031)
Intention to take paternity leave			-0.012 (0.027)	-0.011 (0.026)	0.019 (0.024)	-0.036 (0.041)
Researched about paternity leave			0.027 (0.022)	0.031 (0.023)	0.015 (0.033)	0.041 (0.032)
Favorable toward PL			0.022* (0.013)	0.023* (0.014)	0.054*** (0.019)	-0.009 (0.011)
PL benefits workplace			-0.020* (0.011)	-0.020* (0.012)	-0.041** (0.017)	-0.002 (0.013)
Working hours			0.001 (0.001)	0.001 (0.001)	0.002* (0.001)	-0.000 (0.001)
Age $\leq 50 \times$ Training				0.106** (0.043)		
Manager $\times$ Training				0.028 (0.057)		
Married $\times$ Training				0.069 (0.070)		
Has at least one child $\times$ Training				0.025 (0.049)		
Observations	1444	1444	1368	1368	685	683
No. clusters	75	75	75	75	38	37
Mean	0.849	0.849	0.854	0.854	0.815	0.893

Notes. The sample is the full baseline sample of male employees ($N = 1,444$). The dependent variable is an indicator equal to one if the respondent answered the first follow-up survey. Column (1) regresses the indicator on training assignment alone; Columns (2) and (3) add baseline covariates. Column (4) adds interactions of training assignment with the demographic covariates. Columns (5) and (6) restrict the baseline sample to the training treatment sample and control sample, respectively. All specifications include randomization-strata fixed effects and standard errors clustered at the randomization-unit level. “Mean” is the mean of the dependent variable.

Table D.3: First follow-up response in the target subsamples*Panel A: Men with at least one young child (aged 5 or below)*

	Answer 1st Follow-Up				
	(1)	(2)	(3)	(4) Training Sample	(5) No Training Sample
Training assignment	-0.028 (0.032)	-0.030 (0.032)	-0.038 (0.032)		
Age ≤ 50		-0.007 (0.080)	0.001 (0.088)	-0.117*** (0.037)	0.066 (0.115)
Manager		0.018 (0.023)	0.017 (0.027)	-0.018 (0.043)	0.050 (0.035)
Intention to take paternity leave			-0.011 (0.033)	0.051 (0.048)	-0.069 (0.042)
Researched about paternity leave			-0.016 (0.029)	-0.066 (0.050)	0.003 (0.043)
Favorable toward PL			0.012 (0.028)	0.017 (0.046)	0.012 (0.033)
PL benefits workplace			-0.021* (0.012)	-0.039** (0.018)	-0.009 (0.014)
Working hours			-0.000 (0.001)	0.001 (0.002)	-0.001 (0.001)
Observations	338	338	330	168	162
No. clusters	53	53	53	25	28
Mean	0.902	0.902	0.903	0.887	0.920

Panel B: Men aged 50 or younger

	Answer 1st Follow-Up		
	(1)	(2)	(3)
Training assignment	-0.044 (0.029)	-0.046 (0.030)	-0.052 (0.031)
Manager		0.020 (0.030)	0.023 (0.032)
Married		-0.040 (0.035)	-0.046 (0.035)
Has at least one child		0.078*** (0.026)	0.080*** (0.028)
Intention to take paternity leave			-0.012 (0.025)
Researched about paternity leave			0.028 (0.025)
Favorable toward PL			0.017 (0.011)
PL benefits workplace			-0.026*** (0.010)
Working hours			0.000 (0.001)
Observations	931	931	896
No. clusters	70	70	70
Mean	0.866	0.866	0.866

Notes. Each panel restricts the baseline sample to the indicated subsample. The dependent variable is an indicator equal to one if the respondent answered the first follow-up survey. Column (1) regresses the indicator on training assignment alone; Columns (2) and (3) add baseline covariates. In Panel A, Columns (4) and (5) restrict the subsample to the training treatment sample and control sample, respectively. All specifications include randomization-strata fixed effects and standard errors clustered at the randomization-unit level. “Mean” is the mean of the dependent variable.

Table D.4: Inverse-probability-weighted ITT estimates for training treatment

	(1) PL benefits workplace 1st-order belief (Own belief)	(2) PL benefits workplace 2nd-order belief (Coworkers)	(3) Researched about PL	(4) Daily Childcare Hours (Weekend)	(5) Weekly Working Hours (Spouse)	(6) Daily Housework Hours (Spouse,Weekend)
Training assignment	0.175*** (0.055)	0.110* (0.062)	0.062** (0.024)	0.920** (0.383)	2.140* (1.084)	-0.972** (0.434)
Observations	1165	1153	1156	287	285	282
N clusters	74	74	74	52	51	50

Notes. Each column reports the intent-to-treat effect of training assignment, weighting each observation by the inverse of its estimated probability of answering the first follow-up survey. Response probabilities are estimated separately within the training and control groups from the covariate specification of Table D.2 (Columns (5)–(6)) for the belief and information-seeking outcomes and its counterpart including baseline outcomes for the time-use outcomes, and are winsorized at the first and ninety-ninth percentiles before inversion. The time-use outcomes (Columns 4–6) use the sample of married men with at least one young child. All other elements of the specification match the corresponding main-table estimates, including PDS LASSO-selected controls, the baseline outcome, strata fixed effects, and office-level clustering.

Table D.5: Lee bounds for training treatment

	(1) PL benefits workplace 1st-order belief (Own belief)	(2) PL benefits workplace 2nd-order belief (Coworkers)	(3) Researched about PL	(4) Daily Childcare Hours (Weekend)	(5) Weekly Working Hours (Spouse)	(6) Daily Housework Hours (Spouse,Weekend)
Training assignment						
Lower bound	-0.040 (0.066)	-0.029 (0.045)	0.027 (0.023)	0.702 (0.530)	1.830 (1.850)	-1.119** (0.505)
Upper bound	0.319*** (0.065)	0.246*** (0.049)	0.077 (0.050)	1.332*** (0.510)	3.076* (1.698)	-0.456 (0.524)
Baseline observations	1441	1426	1429	325	320	318
Observations with outcome	1225	1225	1225	303	303	303

Notes. Lee (2009) bounds on the ITT effect of training assignment, applied to the residual of the outcome after partialling out the main specification's controls (with the LASSO-selected set held fixed). Each column reports the lower and upper bound for the indicated outcome. Columns (1)–(3) use the male baseline sample; the time-use outcomes in Columns (4)–(6) use the sample of married men with at least one young child (aged 5 or below). "Baseline observations" is the number of baseline respondents in the estimation sample, and "Observations with outcome" is the number among them who answered the first follow-up. Bootstrap standard errors (1,000 replications) are in parentheses.

Table D.6: Determinants of second follow-up response by information treatment

	Answer 2nd Follow-Up					
	(1)	(2)	(3)	(4)	(5) Information Sample	(6) No Information Sample
Information	0.023 (0.037)	0.018 (0.039)	0.020 (0.035)	0.084 (0.103)		
Age ≤ 50		0.084** (0.037)	0.070** (0.034)	0.115** (0.055)	0.004 (0.050)	0.129** (0.057)
Married		-0.049 (0.045)	-0.059 (0.051)	-0.129* (0.067)	0.011 (0.070)	-0.130* (0.074)
Has at least one child		-0.011 (0.062)	0.019 (0.068)	0.098 (0.096)	-0.068 (0.079)	0.115 (0.101)
Intention to take paternity leave			0.076 (0.054)	0.076 (0.054)	0.087 (0.076)	0.058 (0.069)
Researched about paternity leave			0.016 (0.041)	0.017 (0.041)	-0.010 (0.061)	0.054 (0.060)
Favorable toward PL			0.012 (0.020)	0.011 (0.020)	0.030 (0.038)	-0.005 (0.033)
PL benefits workplace			0.008 (0.017)	0.005 (0.016)	-0.007 (0.024)	0.020 (0.021)
Working hours			-0.003** (0.001)	-0.003** (0.001)	-0.002 (0.002)	-0.004** (0.002)
Age $\leq 50 \times$ Information				-0.092 (0.080)		
Married $\times$ Information				0.148 (0.090)		
Has at least one child $\times$ Information				-0.166 (0.103)		
Observations	589	589	558	558	271	287
No. clusters	65	65	65	65	50	55
Mean	0.711	0.711	0.717	0.717	0.727	0.707

Notes. The sample is the non-managerial male respondents to first follow-up survey ($N = 589$). The dependent variable is an indicator equal to one if the respondent answered the second follow-up survey. Column (1) regresses the indicator on information treatment assignment alone; Columns (2) and (3) add baseline covariates. Column (4) adds interactions of information treatment assignment with the demographic covariates. Columns (5) and (6) restrict the baseline sample to the information treatment sample and control sample, respectively. All specifications include firm-group fixed effects and standard errors clustered at the office level. “Mean” is the mean of the dependent variable.

Table D.7: Inverse-probability-weighted ATE estimates for information treatment

	(1) Intention to take PL (1-3 months later)	(2) Researched about PL
Information	-0.022 (0.042)	0.022 (0.035)
Observations	302	392
N clusters	50	58

Notes. Each column reports the average treatment effect of information treatment on outcomes measured in the second follow-up survey, weighting each observation by the inverse of its estimated probability of answering the second follow-up survey. Response probabilities are estimated separately within the information treatment and control groups from the covariate specification of Table D.6 (Columns (5)–(6)), and are winsorized at the first and ninety-ninth percentiles before inversion. All other elements of the specification match the corresponding main-table estimates in Table 7 (Columns (8) and (9)), including the restriction to respondents aged 50 or younger in Column (1), PDS LASSO-selected controls, the baseline outcome, firm-group fixed effects, and office-level clustering.