Workplace Norms and Paternal Involvement in Childcare*

Mari Tanaka[1], Hiroko Okudaira[2], Mariko Sakka[3], and Shintaro Yamaguchi[1]

[1]Faculty of Economics, University of Tokyo
[2]Business School, Doshisha University
[3]Faculty of Medicine, University of Tsukuba

July 21, 2026

Abstract

Workplace norms may limit paternal involvement despite generous statutory leave. Among 1,225 male employees in four Japanese organizations, we randomize a two-hour work-life balance training for employees and managers, and information correcting misperceptions about colleagues' attitudes. Training strengthened men's own belief that leave benefits the workplace. Fathers' weekend childcare rose about an hour per day, and 2.3 hours where children are young; their spouses' weekly work hours rose by 3.6. Information weakly raised perceived support and moved no downstream outcome. Responsiveness declines with adjustment cost; the leave decision requires stronger organizational levers than attitudes alone.

Keywords: workplace norms, paternal involvement, paternity leave, gender equality, field experiments
JEL: J13, J16, J22, M54

*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 Hitoshi Shigeoka and the seminar participants in various academic workshops and conferences for their valuable comments and suggestions. Finally, we sincerely thank the participating organizations and the collaborating non-profit organizations for providing data and continuous 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 remarkably persistent across developed economies. This persistence carries substantial economic costs: as Cortés and Pan (2023) demonstrate, the unequal division of childcare and domestic work following parenthood explains a large share of remaining gender gaps in earnings, career advancement, and leadership representation. Beyond labor market consequences, more equitable sharing of domestic responsibilities may enhance children's developmental outcomes and increase fertility rates in low-fertility societies.[1]

To increase men's participation in childcare and housework, paternity leave reforms have emerged as a prominent policy intervention. "Daddy quotas," non-transferable leave periods reserved exclusively for fathers, have substantially increased paternity leave uptake in the Nordic countries, Germany, and Canada.[2] In some contexts, these reforms also induced lasting behavioral changes, with fathers maintaining higher levels of childcare and domestic engagement years after returning to work (Tamm, 2019; Patnaik, 2019). However, these successes in raising men's involvement at home have occurred in societies with more progressive workplace norms. In countries with traditional workplace norms, paternity leave uptake remains low despite generous policies, and fathers continue to do far less childcare and housework than mothers.

Japan illustrates the limits of policy generosity when workplace norms remain unsupportive. Women do more than four times as much unpaid care work as men, reflecting one of the world's most extreme gender gaps in household responsibilities.[3] Despite offering some of the world's most extensive paternity leave benefits, which allow up to 12 months of paid leave[4], only about 30% of eligible fathers took any leave in 2023 (Ministry of Health, Labour and Welfare, 2024). Survey evidence shows that fathers often cite workplace norms as the main barrier.[5] Even generous policies cannot overcome organizational norms that penalize men for prioritizing family over work. Promoting gender equality at home thus requires interventions that directly target workplace norms alongside economic support.

To address these workplace norm barriers, we ask not simply whether workplace interventions can raise paternal involvement, but which margins of that involvement respond. Increasing fathers' involvement requires movement at several points that differ

[1] See, for example, Sarkadi et al. (2008) on child development and Doepke and Kindermann (2019) on fertility.

[2] See Dahl et al. (2014); Ekberg et al. (2013); Olafsson and Steingrimsdottir (2020); Tamm (2019); Patnaik (2019) for evidence on increased paternity leave uptake in these settings.

[3] OECD (2023)

[4] Japanese paternity leave benefits typically provide income replacement of 67% for the first six months and 50% thereafter; see Ministry of Health, Labour and Welfare (2021).

[5] Mitsubishi UFJ Research and Consulting (2018)

in how costly they are to change: fathers must first hold supportive beliefs about leave-taking; they can then reallocate flexible, day-to-day time toward childcare and housework; and they may ultimately make the discrete, career-relevant decision to take formal leave. We order these margins by adjustment cost (from malleable beliefs, to flexible household time, to lumpy leave-taking) and ask how far up this sequence an intervention reaches, and how that depends on its intensity.

We conduct a field experiment involving over 1,200 male employees across 74 offices within four organizations to evaluate two interventions of differing intensity: a light-touch information treatment that corrects misperceptions about colleagues' attitudes, and a more immersive two-hour work-life balance training. We target men directly rather than women, departing from existing gender equality interventions that focus on changing women's behavior.

The training curriculum directly addressed behavior at home: among other things, it confronted fathers' tendency to overestimate their own contribution to housework and childcare and encouraged couples to discuss the division of domestic labor, so that household time use was an explicitly targeted outcome rather than an incidental one. Beyond this content, the training may affect paternal involvement through three channels, which we order by the margin of involvement they most directly target. First, the curriculum emphasizes how to maintain workplace performance during the childcare period, speaking directly to participants' own belief that paternity leave can benefit the workplace. Second, the sessions are led by male instructors who serve as credible role models of working fathers, demonstrating that day-to-day involvement in childcare is compatible with career success. Third, participation in the training can provide a coordination device for the leave decision itself. As shown by Johnsen et al. (2024), men avoid leave when they fear falling behind peers who remain at work. Sitting through a session that the employer funded and scheduled during paid working hours, with aligned training running for managers, tells a participant where his organization stands, which is the belief that governs that risk. Because the sessions were online and participant lists were hidden, this channel operates through the endorsement conveyed to those who attended rather than through observing which coworkers took part. This third channel, however, operates on the costliest margin, and its prediction is conditional on leave being practically feasible, a point on which the heterogeneity by staffing replaceability in Section 5.1.3 bears directly.

Our information intervention targets norm misperception through correction of pluralistic ignorance, where individuals may privately support paternity leave but incorrectly believe colleagues and managers disapprove. As Bursztyn et al. (2020) demonstrate in other contexts, information provision can effectively change behavior when individuals hold private views that differ from their perceptions of others' attitudes. Miyajima and Yamaguchi (2017) provide direct evidence of this phenomenon in Japan, where men systematically underestimate colleagues' support for paternal involvement, which we confirm

in our own data. We employ a light-touch approach, revealing actual colleague attitudes toward paternity leave within participants' own organizations to correct misperceptions about workplace norms.

To our knowledge, our study provides the first experimental evidence on workplace interventions specifically designed to target men in order to promote greater paternal involvement in childcare and housework. While existing interventions have largely focused on changing women's behavior or improving women's workplace accommodations, we directly address the constraints that limit men's participation in childcare and domestic work. This distinction matters: rather than asking women to manage both career and family demands more efficiently, we encourage men to take on a larger share of unpaid work at home. We show that an employer-sponsored program can generate significant behavioral changes in household time allocation, increasing fathers' childcare involvement and enabling greater maternal labor supply.

Our experimental design employs a cluster-randomized controlled trial with a baseline survey and two follow-up surveys conducted between August 2023 and March 2024. The training intervention is randomized at the office level using an early versus delayed treatment design, while the information intervention is randomized at the individual level among non-managers. Our outcome measures include attitudes toward paternity leave, time allocation across childcare, housework, and work hours, and spouse labor supply.

Our findings show that workplace interventions can promote paternal involvement. The training significantly strengthened participants' own conviction that paternity leave benefits the workplace, although its effect on their perceptions of coworkers' support was weaker. The training also generated substantial behavioral change. Fathers increased their weekend childcare time by about one hour per day on average across all participants, a 16 percent rise relative to the control group mean, with effects concentrated among those with young children. This additional paternal involvement did not crowd out mothers' childcare: the training added paternal hours rather than reallocating existing childcare between parents. Spouses of treated fathers increased their weekly work hours by 3.6 hours, and in households with young children they also reduced their housework time. It also increased the share of participants who reported researching paternity leave policies. An employer-sponsored intervention can thus reach household time allocation, with effects that extend beyond the employee to the spouse's labor supply.

What the training moved robustly, however, is a first-order belief rather than a perceived norm: participants' own conviction that leave benefits the workplace, not their reading of what colleagues think. The information treatment, which moved perceptions of colleagues directly, moved nothing downstream. Misperceived norms motivate the interventions and our baseline data document them clearly, but our results do not establish that correcting them is what produced the household response.

The increase in spouses' labor supply also illustrates that employer-sponsored train-

ing can generate positive externalities extending beyond the firm. These findings have implications beyond Japan, as workplace norm barriers to paternal involvement persist across diverse national contexts, from traditional societies to progressive countries like Sweden where organizational norms still constrain policy effectiveness (Haas and Hwang, 2019).

The information intervention reduced misperceptions about colleague attitudes, raising participants' second-order beliefs about coworkers' and managers' support for paternity leave. Correcting these misperceptions did not, however, generate effects on downstream outcomes including paternity leave intentions and information-seeking behavior. Awareness of others' supportive attitudes is thus not sufficient on its own to change behavior; more active or intensive forms of engagement appear to be needed to translate updated beliefs into action.

The two interventions trace a staircase: the information treatment moved second-order beliefs and reached nothing beyond them; the training robustly strengthened participants' own belief that leave benefits the workplace and additionally shifted flexible household time, but stopped short of the intention to take leave, itself one of our pre-specified outcome families. Lighter interventions reach only the lower, less costly margins.

Self-reported outcomes raise the possibility of socially desirable reporting, which Section 6.1 examines in detail. The pattern of responses is not one of uniform inflation. Effects concentrate in the margins the training targeted, while weekday time use and the non-target subgroups are unchanged. The same holds for outcomes an image-conscious respondent might equally have inflated, such as perceived psychological safety and willingness to recommend the firm. The reported household time budget is internally consistent, with spouses' additional paid hours accommodated by the housework hours they give up, a coherence that inflating any single answer would not produce. We also conducted the surveys under strict anonymity and verified, using the Social Desirability Scale (Crowne and Marlowe, 1960), that treatment effects are not larger among respondents more prone to favorable self-presentation.

We contribute to a large literature on workplace interventions designed to promote gender equality in labor markets. Much of this work focuses on empowering women, for example through leadership training, mentoring, or programs that aim to increase willingness to compete (e.g., Kelly et al. (2014); Moen et al. (2016); Bandiera et al. (2020); Boneva et al. (2024); Macchiavello et al. (2024)). Other studies seek to reduce bias in evaluation and promotion by targeting supervisors or organizational procedures, such as structured performance reviews or blind screening (e.g., Goldin and Rouse (2000); Bohnet et al. (2016)). These approaches have generated significant progress, but they leave largely unaddressed the unequal distribution of household responsibilities, which remains a central constraint on women's career advancement. Our study takes a different approach: rather than helping women accommodate a male-dominated workplace or

making evaluation procedures fairer, we target men directly. By using a workplace-based program to increase paternal involvement in childcare, we address a structural source of gender inequality that has received comparatively little experimental attention, namely the norms that discourage men from sharing domestic labor.

Our paper is also related to the literature on pluralistic ignorance and belief correction. Bursztyn et al. (2020) show that individuals systematically underestimate support for progressive norms and that revealing actual peer beliefs can shift intentions. Our study applies a similar idea in the field by using survey data from respondents' own organization. Compared to Bertrand et al. (2024), who draw on public opinion data from the general population, our intervention uses information drawn from participants' own colleagues and supervisors, a distinction that matters because individuals may reasonably discount broad societal views when making workplace decisions, whereas information from colleagues is more difficult to ignore. Even with information drawn this close to the decision, we find no effects beyond beliefs. The contrast with our training arm suggests that information alone may not move behavior without accompanying elements such as persuasive content, credible messengers, or visible employer sponsorship.

Beyond its contribution to the labor economics literature, this study speaks to a broader question in organizational economics: whether organizations can deliberately shift workplace norms through targeted interventions. Prior work shows that corporate culture and workplace norms can shape employee behavior and firm outcomes (Guiso et al., 2015), but evidence on how such norms can be changed through low-cost, scalable programs remains limited. In the closest large-scale experiment, Chang et al. (2019) find that an online diversity training moved attitudes among the employees least supportive to begin with but changed behavior only among those already supportive, leaving it least effective where policymakers most wanted it to work. Our training reaches one step further, moving participants' own attitudes and their household behavior within a relatively short horizon. At the same time, our results delineate the limits of such interventions: organizations can strengthen employees' belief in the workplace benefits of paternity leave and move flexible household time, but we find weaker evidence that they change perceived workplace norms, and none that they move the costlier, career-relevant intention to take formal leave, which appears to require stronger organizational levers such as staffing support or managerial accountability.

The remainder of the paper is organized as follows. Section 2 describes the institutional background, focusing on Japan's parental leave system and the participating organizations. Section 3 outlines the research design, detailing the two interventions and randomization procedures. Section 4 presents descriptive statistics and examines baseline balance across treatment groups. Section 5 reports the main results on attitudes, behaviors, and spousal labor outcomes, and Section 6 provides robustness checks and additional analyses. Section 7 concludes with policy implications and directions for future

research.

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, placing it near the bottom of OECD rankings. The imbalances are even more striking within the household: women spend roughly seven hours per day on unpaid care work, more than four times the amount done by men. Yet because total hours of paid and unpaid work are similar across genders, the core issue is not overall work effort but the rigid allocation of tasks within households. These patterns limit women's career advancement and constrain men's involvement at home, shaping the institutional environment relevant to our study (OECD, 2023).

Japan has adopted one of the world's most generous paternity leave schemes to promote fathers' involvement in childcare. Eligible fathers can take up to twelve months of leave with substantial income replacement, yet uptake remains limited. As of our study period, roughly 30 percent of eligible fathers used paternity leave (Ministry of Health, Labour and Welfare, 2024), still far below participation rates in countries with dedicated father quotas such as Norway, Sweden, or Canada (UNICEF, 2022).[6] This sustained gap between statutory generosity and actual behavior indicates that policy reforms alone have not been sufficient to shift paternal involvement.

A growing body of evidence points to the role of persistent social norms and workplace norms in shaping these outcomes. Survey data show that traditional gender attitudes remain widespread; more than one quarter of respondents agree that maternal employment harms children (Haerpfer et al., 2020). These views manifest directly in organizational settings: among fathers who forego leave, a substantial share report an unsupportive workplace atmosphere as a key deterrent (Mitsubishi UFJ Research and Consulting, 2018). These patterns reveal a structural implementation gap in which formal entitlements exist but prevailing norms and expectations limit their use. This points to the role of social and workplace norms, rather than financial incentives alone, as the binding constraint on paternal involvement. Addressing this gap requires interventions that operate at the workplace level rather than relying solely on statutory policy.

Japan's childcare system is dominated by publicly subsidized centers, serving the vast majority of children. Nearly all children aged 3 to 5 attend licensed facilities, which have been tuition-free since 2019, and care for infants and toddlers is heavily subsidized

[6] Japan'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.

for households with children under age 3.[7] Public funding covers most operating costs, keeping out-of-pocket expenses low for families. Another notable feature is that childcare services remain concentrated in standard weekday hours. Only 6.5 percent of centers offer weekend care, and even where extended-hour services are available, 83.6 percent provide less than two additional hours per day. Because childcare services operate almost entirely within standard weekday hours, most parental care on weekdays is compressed into the short window between pickup and bedtime, and fathers' involvement concentrates around drop-off and pickup (National Council of Nursery Schools, 2022).

Alternative childcare arrangements are extremely limited in Japan. National statistics indicate that only 0.4 percent of two-year-old children receive care from babysitters (Ministry of Health, Labour and Welfare, 2012). The use of paid home-support services is similarly rare, with only 1.8 percent of households reporting current use (Ministry of Economy, Trade and Industry, 2024). These figures show that supplemental childcare options outside formal centers are scarce, leaving most families without flexible coverage. Weekday routines are therefore rigid, and the weekend is the main margin on which parents can reallocate childcare time.

Social norms in Japan place unusually demanding expectations on mothers, particularly in the area of household production. A well-known illustration is the *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). Time-use statistics are consistent with this pattern. 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). Given these high standards, simply lowering expectations for daily cooking and cleaning can meaningfully reduce mothers' time burdens. These norms suggest that fathers' taking on even a modest share of domestic tasks could substantially free up women's time, an outcome consistent with the housework reductions we observe among spouses of treated fathers.

3 Research Design

3.1 Overview

This section describes the research design, which consists of two randomized interventions combined with a three-wave survey conducted across four Japanese organizations. Figure 1 summarizes the overall timeline.

[7] See OECD (2024). Although Japan previously faced substantial childcare waiting lists (Yamaguchi et al., 2018b,a), capacity expansion and declining birth rates have eased shortages.

We collaborated with a nonprofit organization[8] to recruit four employers (two private firms and two municipal governments) that voluntarily applied to join a program designed to promote paternal involvement in childcare and housework, including paternity leave uptake. Within each organization, employees self-selected into the program by registering through an online form, and registered employees subsequently received a link to the baseline survey. In total, 1,962 employees signed up across the four organizations. The four organizations comprise 84 branch offices, which serve as the units of randomization for the training; the sign-ups came from 80 of these offices, while the remaining four produced no registrations and therefore drop out of the analysis at this stage.

We implemented two distinct interventions targeting different mechanisms through which workplace norms may constrain paternal involvement. The first is a two-hour WLB training program for male employees and their managers, designed to shift attitudes toward fatherhood and provide organizational endorsement for paternity leave. The second is an information intervention that corrected employees' misperceptions about their colleagues' and supervisors' actual views on paternity leave, targeting pluralistic ignorance as a barrier to leave-taking.

The experiment comprised a baseline survey and two follow-up surveys, with the training delivered between survey waves. As shown in Figure 1, randomly assigned treatment offices received the training between the baseline and the first follow-up surveys, while control offices received it between the two follow-up surveys. The information intervention was embedded in the first follow-up survey and administered only to non-managers. All surveys and interventions took place between August 2023 and March 2024. This staggered design allows us to estimate the short-run effects of the training program using first follow-up responses from control offices as the counterfactual, and to assess whether correcting norm misperceptions alone, absent the training's attitudinal and organizational content, is sufficient to change behavior. We adopted this waitlist (crossover) design because the participating organizations wanted all employees to receive the training; delivering it to control offices only later was a condition for conducting the study and ensured equitable access. A tradeoff is that, because every office is eventually treated, we cannot identify long-term effects by comparing trained employees with employees who never receive the training.

Our main analytic sample consists of 1,225 male employees who completed both the baseline and first follow-up surveys and provided essential demographic and treatment information.

The trial was registered in the AEA RCT Registry (AEARCTR-0012315), which pre-specified five families of outcomes: willingness to take paternity leave; time spent

[8]The organization is Evidence Studio (ES), 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.

on childcare, housework, and market work; 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 families. Section 5 presents them 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.

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. The four participating organizations represent typical employers from both the public and private sectors. The two municipal government offices 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.

All four are broadly comparable to the national average and do not represent extreme outliers in terms of size, industry, or composition. 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 rates differ substantially across organizations: among the private firms, one did not disclose this figure, while the other reported a rate of about 20 percent, well below the average of 52.2 percent among publicly listed firms. Participants came from diverse occupational backgrounds across a wide range of functions at both headquarters and regional branches.

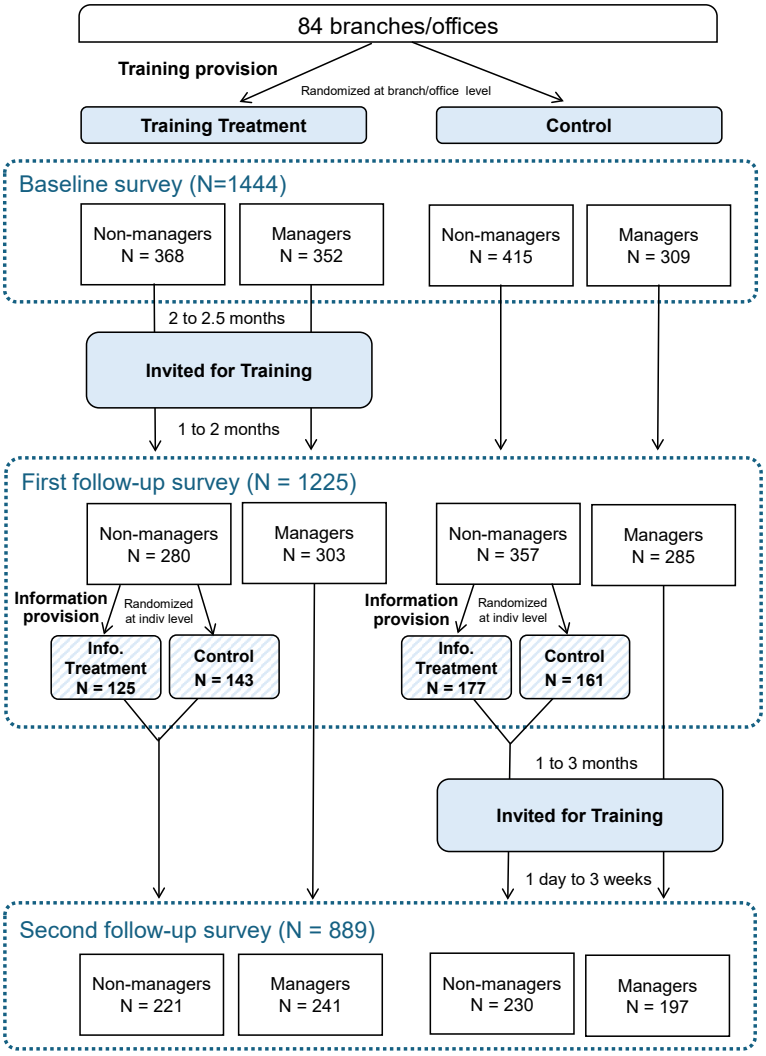
3.3 Interventions

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 employers. Separate sessions were provided for managers and non-managers and offered simultaneously, so that the program could foster workplace norms conducive to paternity leave uptake across hierarchical levels.[9] The sessions were conducted online, and the list of participants was not visible to attendees, so employees could not observe which of their colleagues took part.

[9]Those unable to attend in real time were given access to a recorded version for ten days, and participating organizations encouraged registrants to complete the program by the deadline; watching the recording within this window counts as participation.

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.

Both sessions followed a common curriculum delivered by male instructors who were successful business professionals and actively involved fathers. The curriculum introduced paternity leave policies in Japan as part of a broader discussion on fathers' involvement in childcare and emphasized the benefits of greater paternal engagement for families, fathers themselves, and workplaces. For families, it emphasized that childcare is difficult for one parent to shoulder alone and noted that fathers often overestimate their contributions to housework and childcare relative to their partners' assessments. For fathers, the curriculum emphasized that taking leave need not harm career prospects and that active involvement in childcare can enhance work-life satisfaction. For workplaces, it emphasized that paternal involvement can strengthen teamwork and management skills by encouraging better communication and promoting shared task coverage as employees coordinate childcare responsibilities within teams. By correcting fathers' overestimation of their own domestic contributions and demonstrating the workplace benefits of shared caregiving, the curriculum aimed to change both household-level perceptions and workplace-level attitudes. The manager session additionally covered strategies for supporting employees with childcare responsibilities, including communication practices and evaluation systems that account for family commitments; both sessions concluded with a group discussion.

We interpret these curriculum elements as operating through three possible channels: the reduction of perceived career costs of leave-taking, the role model effect of male instructors, and the coordination device provided by employer endorsement. The first two channels worked through individual beliefs and attitudes. The curriculum directly addressed the perceived trade-off between career commitment and childcare involvement: by demonstrating that paternity leave can enhance rather than harm workplace performance through improved team coordination, skill development, and time management, the training aimed to reduce fathers' concerns that taking leave signals weak commitment. The male instructors, themselves active fathers with successful careers, provided visible role models; experimental evidence suggests that observing credible examples of others who have combined demanding work with active caregiving can shift beliefs about what is feasible and socially acceptable (e.g., Beaman et al., 2012). The third channel also operated through participants' exposure to the training. Employer sponsorship and managerial participation served as a coordination device: as in models of implicit tournaments in the workplace (Johnsen et al., 2024), men may avoid taking leave when they fear falling behind peers who remain at work. Attending conveyed the employer's position directly. Participants sat through a session their employer had funded and scheduled during paid working hours, and learned that a parallel session was being delivered to their managers, so attendance was itself evidence that the organization treated paternal involvement as legitimate rather than as a private accommodation to be arranged quietly. As noted above, the online format did not reveal participant lists, so this channel does not rely on attendees observing which coworkers took part; the intervention created no

observable participation signal among peers, and any coordination effect must instead operate through the content and organizational framing that participants experienced. What the training plausibly shifted is each participant's own assessment of how his employer would regard leave-taking, which is the belief that governs the perceived career risk of acting first.

Randomization of Training Sessions Across the four organizations, there were 84 offices, corresponding to branches in the private firms and municipalities in the local governments. We randomly assigned these offices into either a training treatment group (early treated) or a control group (later treated). 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 for two reasons: the training was delivered to groups of employees, and colleagues within an office interact frequently, so office-level assignment keeps any within-office spillovers inside treated units. Spillovers across offices are also possible if employees in different offices communicate, but the structure of the sample limits the scope for such contact: in the municipal stratum, offices are separate employers in different municipalities, and in the private firms, branches are geographically dispersed nationwide. To the extent residual contact occurs, it would attenuate the contrast between treatment and control offices and render our estimates conservative.

Employees from 80 of these offices signed up for the training program. Sign-up rates at the office level were not significantly correlated with treatment assignment (Appendix Table B.2), and the main findings are robust to restricting the sample to the organization with the highest sign-up rate (Appendix C).[10]

3.3.2 Information Provision

Intervention Content We designed the information experiment to shift the perceptions of male subordinates who hesitated to take paternity leave because of concerns about how they were viewed in the workplace. Our baseline survey provides evidence of pluralistic ignorance: although supervisors and coworkers generally expressed favorable views toward paternity leave, subordinates tended to underestimate this support. Building on previous studies documenting the role of information provision in changing perceived social norms (Bursztyn et al., 2020; Bertrand et al., 2024), we provided participants with

[10]HR divisions circulated the invitation to all employees; the invitation letter is reproduced in Appendix A.1. In total, 1,962 employees across 80 offices signed up, representing 12.4 percent of all employees in the four organizations. The invitation was sent to both male and female employees; the sign-up rate includes female participants, although the registration form did not collect gender information.

accurate information about their colleagues' and supervisors' actual (first-order) attitudes toward paternity leave, thereby correcting participants' second-order beliefs. By correcting the overestimation of negative second-order beliefs, sometimes referred to as a "wedge," we aimed to reduce the perceived social costs of taking paternity leave and to encourage greater uptake.

In the baseline survey, respondents and their managers reported on a six-point scale the extent to which they personally agreed with the following statement, drawn from standardized questions used in public opinion surveys including the Cabinet Office Opinion Poll:

> "It is good for men to take leave for childcare or caregiving."

We also elicited second-order beliefs by asking participants how much they thought their coworkers and managers would agree with the same statement. Because we posed the same question to both managers and non-managers, we were able to calculate the extent to which employees over- or underestimated their colleagues' and supervisors' actual views. In the first follow-up survey, treated non-managers were shown two charts for each target group (managers and colleagues): one displaying the distribution of non-managers' second-order beliefs about that group, and one displaying the group's actual survey responses, making the discrepancy immediately visible. Appendix A.3 reproduces the screens presented to the treatment group.

A key feature of our design was that the information was drawn directly from participants' own colleagues and managers within the same organization, rather than from society at large. We explicitly emphasized this in the treatment materials, presenting the charts as reflecting the actual views of participants' own coworkers and supervisors. This focus was important because the views of one's own colleagues and supervisors are more directly relevant to the decision to take leave than aggregate societal norms.

Randomization The intervention was implemented only among non-managers. We used stratified randomization at the organization level and randomly assigned half of the non-managers within each stratum to the treatment group and half to the control group. We randomized this intervention at the individual level, rather than the office level, because it was delivered privately within the online survey and involved no group component, so individual assignment raises no comparable spillover concern. Treated respondents could in principle mention the information to colleagues, but the belief outcomes were measured within the same survey session, before any such contact could occur. The one outcome measured later, information-seeking in the second follow-up survey, could be affected only through subsequent conversations, which would, if anything, attenuate the estimated effect.

3.4 Survey Waves and Sample

To minimize experimenter demand effects, we ensured strict anonymity and separated responsibilities across three organizations: the training was delivered by an external nonprofit, data collection was coordinated by the managing NPO, and all analyses were conducted independently by researchers.[11]

We allowed sufficient time between the surveys and interventions to capture potential behavioral and workplace changes. Approximately two months after the baseline survey, the training was delivered to the treatment group. The first follow-up survey was fielded one to two months later. The control group received the training about one to three months after that, followed by the second follow-up survey within three weeks.

The baseline and first follow-up surveys constitute our main analysis sample. Both surveys collected information on workplace environments (for example, psychological safety) and behavioral outcomes (for example, household time allocation), as well as measures of personal norms and perceptions of colleagues' and managers' norms. The baseline survey also gathered demographic characteristics, while the second follow-up survey provided supplementary data, including the social desirability scale. We assessed whether results might reflect a general tendency toward socially desirable responding by administering a Japanese adaptation of the Social Desirability Scale (SDS) (Kitamura and Suzuki, 1986) in the second follow-up survey; details are provided in Appendix Table C.1, and results are reported in Section 6.1.

Survey links were distributed to all employees who had signed up for the training program (N = 1,962). Of the 1,771 employees who answered the baseline survey, 1,444 were men, and 1,225 of them, or 84.8 percent, also answered the first follow-up: attrition between the two waves that define our main sample is thus 15.2 percent. Retention to the second follow-up is lower, with 889 men, or 61.6 percent of male baseline respondents, answering it; that wave, however, supplies only the social desirability scale of Section 6.1 and one outcome of the information intervention. Measured against all sign-ups rather than against baseline respondents, response rates were 90.3 percent at baseline, 75.6 percent at the first follow-up (N = 1,483), and 55.0 percent at the second (N = 1,080).[12] Appendix Table B.2 shows that office-level response rates were not significantly correlated with treatment assignment.

Our main analytic sample consists of male employees who completed both the baseline and first follow-up surveys and provided essential demographic and treatment information, including age, marital status, managerial status, parental status, and training assignment. This restriction yields 1,225 employees across 74 offices (six offices had no

[11] HR divisions had no access to survey data, researchers had no access to respondents' identities, and the managing NPO linked training records with survey responses while keeping all personally identifiable information confidential. This arrangement was explained at the beginning of each survey; Appendix A.2 presents the informed consent form.

[12] All respondents who completed a follow-up survey also completed the baseline survey.

male respondents in the baseline survey).

3.5 Econometric Specification

We estimate the causal effects of the training and information interventions on participants' attitudes and behaviors. We adopt different estimands for each intervention, reflecting their distinct implementation structures. For both interventions, we employ an analysis of covariance (ANCOVA) specification, in which control variables include baseline outcomes to enhance statistical precision and comparability across estimations (McKenzie, 2012).

Training Intervention

For the training intervention, we take the local average treatment effect (LATE) as our primary estimand. Our interest lies in the efficacy of the training program itself rather than the effect of being invited to participate. In scaling such interventions, participation is a key determinant of overall impact, but firms can influence participation through incentives or mandates. The LATE therefore provides a more policy-relevant measure of the program's effectiveness. In the appendix, we provide estimates for intention-to-treat (ITT) for reference.

We estimate the following model using two-stage least squares (2SLS), instrumenting actual participation with training assignment at the office level. The first stage is:

$$TrainingParticipation_{ij} = \pi\, Training_j + \gamma' X_{ij} + u_{ij}, \tag{1}$$

where $TrainingParticipation_{ij}$ is an indicator for whether individual i in office j participated in the training, and $Training_j$ equals one if the office was assigned to the early training group.

The second stage is:

$$Y_{ij} = \beta\, \widehat{TrainingParticipation}_{ij} + \theta' X_{ij} + \epsilon_{ij}, \tag{2}$$

where Y_{ij} denotes the outcome of interest measured in the first follow-up survey, and $\widehat{TrainingParticipation}_{ij}$ is the predicted value from the first stage.

The vector X_{ij} includes the baseline value of the outcome variable, randomization-strata fixed effects, and additional controls selected using the post-double-selection (PDS) LASSO procedure (Belloni et al., 2014). The candidate control set consists of age, age squared, managerial status, marital status, parental status, information treatment status, and all interactions among these variables.[13] Standard errors are clustered at the office

[13]We 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.

level. Participation is self-reported in the follow-up survey; the organizers did not maintain individual attendance or archive-viewing records permitting administrative verification. Because participation is instrumented by random training assignment, self-selection into participation does not bias our estimates; any misreporting of participation would rescale the LATE rather than reverse its sign, and we also report ITT estimates, which do not rely on the self-reported participation measure. Among men in offices assigned to the early training group and signed up for the training, 49 percent participated in the training before the first follow-up survey. No employee in the control group received the training before the first follow-up survey, so there is no cross-over contamination.

Interpreting β as the effect of a respondent's own participation additionally requires that assignment affect his outcomes only through that participation. Two features of the design support this. The sessions were delivered online and the attendee list was not visible, so employees could not learn which colleagues had taken part; and the invitation to the program was circulated to all employees in every organization, so exposure to the announcement itself does not distinguish early- from late-assigned offices. Managers in early-assigned offices did, however, attend their own sessions, and we cannot rule out that a trained manager influenced subordinates who did not attend. Any such exposure would be attributed to the respondent's own participation, so β is best read as the effect of the training as experienced in a treated office rather than as the effect of attending in isolation.

Information Intervention

For the information intervention, implemented within the first follow-up survey among non-managerial employees, we estimate the average treatment effect (ATE). Because all targeted employees were exposed to the information during the survey, treatment compliance is effectively universal, so the ATE and ITT coincide.

We estimate:

$$Y_{ij} = \beta_1 \, Information_{ij} + \theta' X_{ij} + \epsilon_{ij}, \tag{3}$$

where $Information_{ij}$ equals one if respondent i in office j received the information treatment. The vector X_{ij} has the same structure as in equation 2, except that, because the information treatment was randomized at the individual level within each organization rather than across offices, it includes firm group fixed effects in place of the training's randomization-strata fixed effects. The candidate control set for PDS LASSO comprises age, age squared, managerial status, marital status, parental status, training invitation status, and all interactions among these variables. Standard errors are clustered at the office level. To assess potential complementarities between the two interventions, we also estimate a joint specification that augments equation 3 with a training participation indicator and its interaction with information treatment status.

Because the information intervention was implemented within the survey, we focus on its immediate effects on beliefs about men's paternity leave take-up and on intentions to take paternity leave. We also examine one behavioral outcome, whether participants subsequently researched paternity leave policies, measured in the second follow-up survey. By the time of the second follow-up survey, the control group had also completed the training, so this outcome captures the incremental effect of the information intervention over and above the training.

Multiple Hypothesis Testing

Each intervention is evaluated on several outcomes, which we group into the families described in Section 3, so conventional p-values overstate how often effects of the size we report should be expected to arise by chance. Alongside standard errors we therefore report Anderson (2008) sharpened two-stage q-values. A q-value is the false discovery rate analogue of a p-value: it is the smallest false discovery rate at which the hypothesis for that outcome would be rejected, so rejecting every hypothesis with a q-value at or below 0.10 implies an expected false discovery rate of 10 percent among those rejections. We compute q-values within outcome families rather than across all estimates in the paper, and report them in brackets in the tables and in parentheses in the text. A family consisting of a single outcome involves no multiplicity, and no q-value is reported for it.

4 Descriptive Statistics

4.1 Evidence for Pluralistic Ignorance

We begin by defining how we measure individual own beliefs and second-order beliefs about coworkers' and managers' attitudes regarding men's paternity leave. The concept of pluralistic ignorance refers to a situation in which individuals misjudge the prevailing social norms or attitudes of others, even when most people privately hold similar views. In our context, employees may personally support men's paternity leave but underestimate how supportive their coworkers or supervisors actually are. This misalignment between perceived and actual attitudes provides direct evidence of pluralistic ignorance in the workplace.

We construct two sets of measures from our baseline survey. The first captures respondents' own beliefs about paternity leave. Each respondent indicated, on a six-point scale, the extent to which they personally agree with the following two statements:

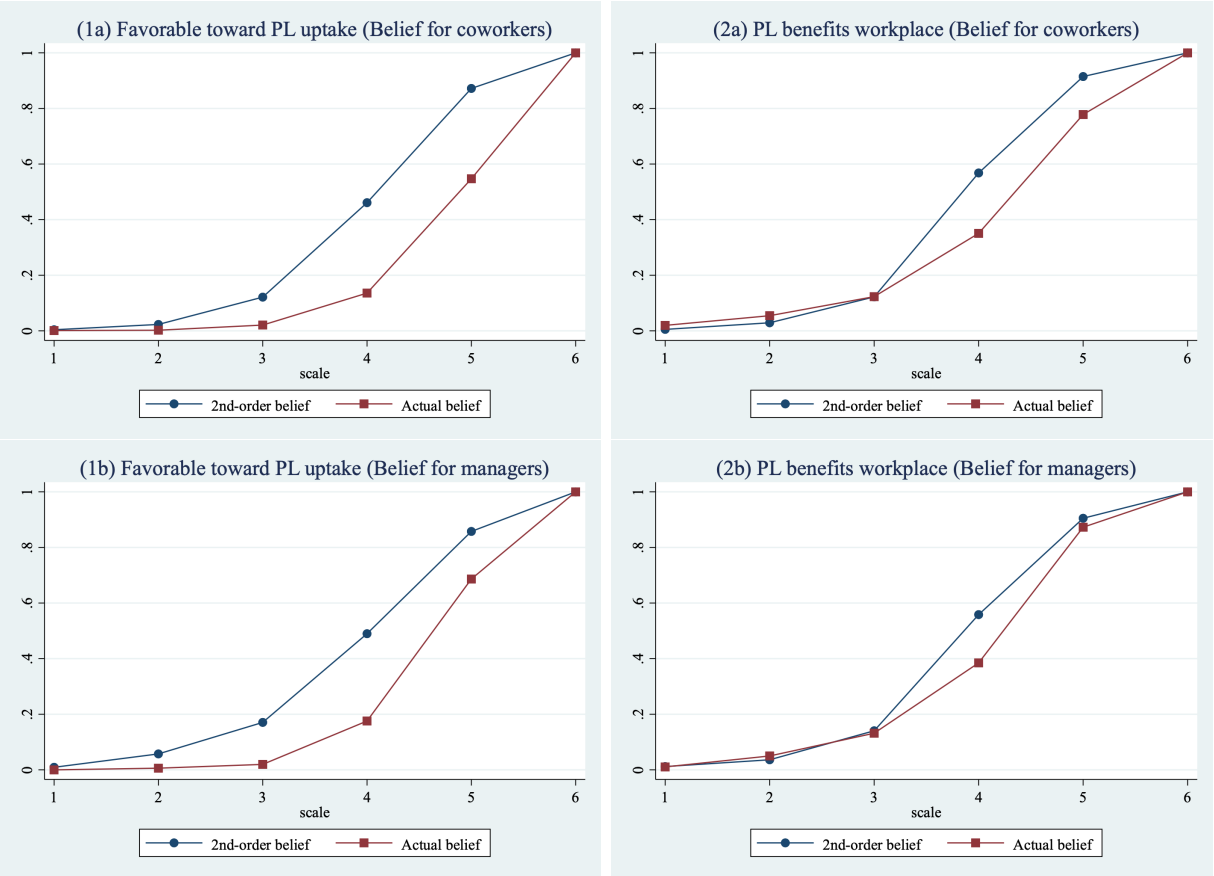
How much do you *personally* agree with the following statements?

(1) "*It is a good thing for men to take leave for childcare and caregiving.*" (hereafter "Favorable toward PL uptake")

(2) "*Increasing men's paternity leave uptake can improve company operations and promote a collaborative workplace culture.*" (hereafter "PL benefits the workplace")

Second, we use survey questions that ask each respondent how much they think their coworkers and their managers *would* agree with the same two statements, eliciting second-order beliefs about coworkers and managers separately.

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 presents the cumulative distributions of second-order beliefs about coworkers' (or managers') attitudes, alongside the cumulative distributions of the actual beliefs held by non-managers (or managers) regarding the same statements, labeled "Actual belief" in the figure. These beliefs are measured in the baseline survey. For the statement "Favorable toward PL uptake", the cumulative distribution of actual beliefs first-order stochastically dominates that of second-order beliefs, for both coworkers and managers, indicating that respondents are overly pessimistic about how supportive their colleagues are toward men's paternity leave. For "PL benefits the workplace" the same ranking holds over the upper half of the scale, where the great majority of the mass lies, although

the two distributions nearly coincide and cross among the few respondents in the lowest response categories.

The magnitude of the gap in beliefs is larger for the statement "Favorable toward PL uptake" than for "PL benefits the workplace." Thus, employees underestimate others' approval of paternity-leave uptake more strongly than they underestimate others' belief in its workplace benefits. The gaps in beliefs are broadly balanced between managers and coworkers, implying that misperceptions are not systematically biased toward one reference group.

Overall, the descriptive evidence clearly demonstrates pluralistic ignorance at baseline: employees personally support men's paternity leave but believe that their peers and supervisors are substantially less supportive than they actually are.

4.2 Balance Check

Before turning to the estimation results, we examine whether baseline characteristics are balanced between the treatment and control groups. Tables 1 and 2 present descriptive statistics for the training and information interventions, respectively. All variables are measured at baseline prior to the interventions. The tables report means for the control and treatment groups and the difference between the two (after controlling for the same strata fixed effects as in the main empirical specification), with p-values for the difference reported in parentheses.

For the training intervention (Table 1), randomization was conducted at the office level within organizational strata. The treatment and control groups are broadly similar in observable characteristics. Marital status, parental status, weekly working hours, and own and second-order beliefs are nearly identical across groups. Treated employees are slightly younger on average (44.2 vs. 45.5 years, $p < 0.10$). The share of managers is somewhat higher in the treatment group (0.520 vs. 0.444, $p < 0.05$), and weekday childcare hours are modestly longer among treated employees (1.54 vs. 1.31 hours, $p < 0.10$). These small differences are consistent with random variation and will be accounted for in the estimation stage by controlling for baseline covariates. We also note that, among non-managers, the training treatment group has a slightly lower probability of receiving the information treatment (0.462 vs. 0.520, $p<0.10$). We account for the potential impact of this imbalance in our estimation by including information treatment status in the candidate control set for PDS LASSO. Overall, the stratified randomization achieved balance in demographic and job-related characteristics.

For the information intervention (Table 2), randomization was conducted at the individual level within each organization. The treatment and control groups are again well balanced. Mean age, marital and parental status, and other baseline measures are very similar across groups. Two small differences emerge. The treatment group shows

Table 1: Descriptive statistics and balance test for training intervention

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

Notes. All variables are measured in the baseline survey. 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

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

Notes. All variables are measured in the baseline survey. 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.

a slightly higher baseline intention to take paternity leave (0.325 vs. 0.251, $p < 0.10$) and a lower probability of being invited to the training intervention (0.423 vs. 0.480, $p < 0.10$). Given the number of variables tested in the balance check, finding one or two statistically significant differences by chance alone is statistically expected and does not indicate a failure of randomization. These differences are further addressed by including baseline characteristics and training invitation status in the candidate control set for PDS LASSO, ensuring they do not confound the estimated treatment effects.

The two tables together indicate that both randomization procedures produced comparable treatment and control groups. Baseline attitudes and behavioral indicators are well aligned, suggesting that subsequent differences in outcomes reflect the causal effects of the interventions rather than pre-existing differences.

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.

5.1 Effects of Training Intervention

We first analyze changes in attitudes and perceptions of workplace norms, as belief shifts are a first step toward behavioral change. We then assess effects on time use and engagement with paternity leave policies, shedding light on how workplace training can promote greater paternal involvement at home. Throughout this section, the primary estimand is the Local Average Treatment Effect (LATE) of training participation, estimated by two-stage least squares (2SLS) using training assignment as an instrument for actual participation; Intention-to-Treat (ITT) estimates are reported in the appendix for reference. We report Anderson (2008) sharpened q-values alongside standard errors throughout; Section 3.5 describes their construction and interpretation.

5.1.1 Beliefs

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: LATE						
Training	0.076	0.447***	0.145	0.291**	-0.108	0.094
	(0.057)	(0.105)	(0.145)	(0.143)	(0.114)	(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
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 perception about workplace norms. The coefficients are estimated using training assignment as the instrumental variable, and they represent the Local Average Treatment Effect (LATE) of training participation. 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.

The first set of results concerns how the training intervention affected participants' own beliefs and their perceptions of workplace norms related to paternity leave. We focus on two key statements: (1) that taking paternity leave is a good thing for men ("Favorable toward PL uptake") and (2) that such uptake can improve company operations and promote a collaborative workplace culture ("PL benefits the workplace"; see Section 4.1 for variable definitions).

Table 3 Panel A reports the estimates for LATE of training participation on these

belief outcomes, using training assignment as an instrument for participation.[14] The corresponding ITT estimates are shown in Appendix Table B.3. For own beliefs, the training increased agreement with the first statement by 0.08 points (Column 1; $q = 0.101$). It increased agreement with the second statement, "PL benefits the workplace," by 0.45 points on the six-point scale (Column 2), an effect that remains significant after correction for multiple hypothesis testing ($q = 0.001$). These results suggest that the training was particularly effective in strengthening beliefs about the workplace advantages of men's paternity leave.

Columns (3)–(6) of Table 3 Panel A present the estimated effects of the training on respondents' perceptions of coworkers' and managers' attitudes. The training also raised second-order beliefs about coworkers' support for the workplace benefits of leave (Column 4; $q = 0.199$), while the effects on perceptions of managers are smaller and statistically insignificant ($q = 0.488$). If the apparent coworker–manager difference is real, one explanation is that supervisors' views are harder to infer: interactions with managers are typically less frequent and more formal than those with peers. The training's robust effect on beliefs is therefore concentrated on participants' own conviction that paternity leave benefits the workplace, rather than on their perceptions of prevailing workplace norms.

Panel B of Table 3 reports the results of Wilcoxon rank-sum tests comparing the belief distributions between the training treatment and control groups. Beliefs are measured in the first follow-up survey. Appendix Figure B.1 presents the corresponding cumulative distribution functions. Consistent with the regression-based results, the treatment and control distributions differ significantly in Columns (2) and (4), the two statements concerning the workplace benefits of leave.

The stronger response for the "PL benefits the workplace" statement is consistent with the structure and content of the training. Instructors emphasized how paternity leave can enhance workplace efficiency by improving communication, promoting shared task coverage, and encouraging broader skill development within teams. The employer-sponsored nature of the program may also have signaled that the organization values male participation in childcare. As we show in Section 5.2, a light-touch information intervention that corrected second-order beliefs without these reinforcing elements updated perceptions but did not produce behavioral change, consistent with the interpretation that the reduction of perceived career costs, role model exposure, and the coordination device provided by institutional endorsement together contributed to the training's effectiveness, in a way that belief correction alone did not.

Heterogeneity analysis (Appendix Table B.5) shows that the estimated effects do not differ significantly by participant's baseline beliefs. We also examine heterogeneity by job position (manager versus non-manager) and family status (with or without children)

[14] Appendix Table B.4 reports the LATE estimates without PDS LASSO-selected controls.

in Appendix Table B.6, but find no statistically significant differences across these subgroups.

5.1.2 Household time allocation

The training shifted participants' beliefs about the workplace benefits of paternity leave, though effects on general approval of leave-taking were more modest. We next assess whether the training translated into behavioral changes, examining effects on participants' time allocation, as well as those of their spouses. Because the three channels identified in Section 3 were all active simultaneously, the behavioral changes documented below reflect their combined effect.

Table 4 reports the estimated LATE of training participation using training assignment as an instrument, and Appendix Table B.7 presents the corresponding ITT estimates.[15] Results are presented for the full analytic sample. To examine whether the effects vary by the presence and age of children, Table 5 reports LATE estimates separately for three groups: (1) households with at least one young child (aged five or below), (2) those with only older children (aged six or above), and (3) households without children. The corresponding ITT estimates are presented in Appendix Table B.9.

The training increased fathers' weekend childcare time, while leaving weekday childcare and working hours unchanged (Table 4, Panel A, Columns (1)–(3)). In the full sample, treated participants spent about one additional hour per day on weekend childcare, a 16 percent increase relative to the control mean ($q = 0.176$). The effect is concentrated, as the training's emphasis on caring for young children would predict, in households with at least one child aged five or below, where fathers increased weekend childcare by 2.3 hours per day (Table 5, Panel A, Column (3)), an effect that remains significant after the same correction ($q = 0.034$). Working hours in this group also declined, though not significantly. The absence of weekday effects is consistent with how little discretionary time the weekday leaves for these full-time employees. Control-group fathers work about 44 hours per week, and once commuting and the fixed evening routine of pickup, dinner, and bedtime are accounted for, little adjustable time remains on a weekday evening. The contrast with spouses is instructive: spouses in our sample average about 20 paid hours per week, so their weekday afternoons are available for childcare, whereas fathers' 44-hour weeks occupy precisely those hours. The constraint on weekday adjustment is thus work schedules rather than the availability of childcare tasks. The weekend is therefore effectively the only margin on which fathers can reallocate time toward childcare, which is where we observe the response. The concentration of formal childcare in standard weekday hours, described in Section 2, reinforces this pattern but is secondary to the binding time constraint on weekdays.

[15] Appendix Table B.8 reports the LATE 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)
Panel A: Own hours					
Training	-0.833	-0.113	0.943**	0.022	0.040
	(1.764)	(0.220)	(0.453)	(0.113)	(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
Panel B: Spouse hours					
Training	3.578**	-0.028	0.108	-0.495	-0.084
	(1.729)	(0.385)	(0.453)	(0.390)	(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 the effects of the training intervention on the behaviors of male respondents and their spouses. The coefficients are estimated using training assignment as the instrumental variable, and they represent the Local Average Treatment Effect (LATE) of training participation. 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, and single-outcome working-hour families are not subject to the correction. 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

	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)
Panel A: Own hours					
Training × Young children	-3.958	-0.721	2.286***	-0.290	-0.875
	(2.670)	(0.502)	(0.864)	(0.233)	(0.757)
Anderson q-value		[0.229]	[0.034]	[0.229]	[0.229]
Training × Only older children	-0.166	0.251	0.331	0.386**	0.442
	(2.705)	(0.237)	(0.559)	(0.155)	(0.289)
Training × No children	2.308			0.182	0.711**
	(7.735)			(0.272)	(0.296)
Observations	1015	859	859	1001	1001
No. clusters	74	72	72	74	74
Baseline Mean	45.591	1.455	6.728	0.880	2.041
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes
Panel B: Spouse hours					
Training × Young children	4.353*	0.310	-0.451	-1.334*	-2.061**
	(2.475)	(0.864)	(1.186)	(0.719)	(0.906)
Anderson q-value		[0.562]	[0.562]	[0.106]	[0.101]
Training × Only older children	1.789	0.204	0.440	-0.329	0.631*
	(2.068)	(0.478)	(0.762)	(0.464)	(0.338)
Training × No children	12.39			0.644	1.615
	(8.397)			(1.330)	(1.386)
Observations	994	846	846	980	980
No. clusters	74	72	72	74	74
Baseline Mean	21.757	5.101	7.915	4.090	4.018
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. "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). "Baseline Mean" is the mean of the dependent variable in the baseline survey for the estimation sample. 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.

Having established that the training affected participants' own time use, we now turn to the effects on their spouses. Table 4, Panel B, shows that spouses' labor supply responded to the intervention. Column (1) indicates a statistically significant increase in spouses' working hours, with the estimated LATE corresponding to an increase of 3.6 hours per week, or about 18 percent relative to the control mean. Appendix Table B.11 documents that this increase is driven by women who were already employed part-time: the share of spouses working more than 10 hours per week rises by 9 percentage points and the share working more than 20 hours by 15, while the shares working more than 30 hours and more than zero hours are both unchanged, indicating no detectable movement at the extensive margin of employment. Thus, the intervention appears to lengthen hours among existing part-time workers rather than induce new labor-force entry. This is also the margin on which a response within one to two months is institutionally plausible: part-time schedules in Japan are typically set through monthly shift adjustments, so hours can expand quickly once household time is freed, whereas labor-force entry or a move to full-time work involves longer lags.

The rise in spouses' paid work hours was mirrored by a reduction in their housework, though one visible mainly in the households where fathers' childcare also rose. In the full sample, the estimates for spouses' housework and childcare are small and not statistically significant (Table 4, Panel B, Columns (2)–(5)). Among households with young children, however, spouses reduced both weekday and weekend housework (Table 5, Panel B, Columns (4)–(5); sharpened $q = 0.106$ and 0.101, respectively). The combined weekly reduction of about 11 hours more than accommodates their additional 4.4 paid hours, consistent with a reallocation of time within the household that both supported greater maternal labor market participation and lowered the overall burden of unpaid work, a meaningful margin in a context where domestic standards are unusually demanding (Section 2).

The spousal outcomes are reported by the husband rather than measured from the wife directly. The reported additional paid hours are fully accommodated by the reported reductions in housework, so the implied reallocation is internally consistent as a household time budget.

The increase in fathers' childcare time did not come at the expense of mothers' involvement. As shown in Appendix Table B.12, the couple's total childcare time did not decline, and among households with young children the couple's total housework time fell substantially. This reduction in unpaid work is consistent with greater intra-household coordination following the training: among treated 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). Because these post-training reports are available only for treated participants, they are descriptive rather than causal, but they corroborate the

interpretation that the intervention prompted concrete short-term action extending to the household level.[16]

For households with only older children, the training had no detectable effects on working hours or childcare time, but it led to a statistically significant increase in men's housework time (Table 5, Panel A). Effects on their spouses were mixed, with a slight increase in weekend housework (Table 5, Panel B). Among participants without children, the training increased weekend housework time (Table 5, Panel A, Column (5)). These patterns across non-target subgroups are difficult to interpret and should be treated with caution, as both groups are smaller and the training content was less directly relevant to their household circumstances.

Because our behavior measures are self-reported, Section 6.1 examines whether misreporting influenced by the training could account for them, showing that the cheapest desirable response, stating an intention to take leave, did not move, that the effects are concentrated in the margins the training targeted rather than spread uniformly across outcomes, and that respondents more prone to socially desirable responding do not drive the results.

5.1.3 Information-seeking and intention to take paternity leave

Finally, we examine participants' engagement with paternity leave itself through two outcomes that differ in how much they commit the participant: whether he sought out information about leave policies (a low-stakes, preparatory step) and whether he intends to take leave (a discrete, career-relevant decision).

The training markedly increased information-seeking. The share of participants who reported researching paternity leave policies rose by about 16 percentage points (Table 6, Column (1)), and the effect is similar, about 17 points, among men aged 50 or younger (Column (2)), the sample used for the intention outcome below. The effect appears among both non-managerial employees, who are more likely to use such policies themselves, and managers, who may have sought the information to better support employees with childcare responsibilities (Appendix Table B.10).

The decision to take leave, however, did not respond. Table 6, Column (3), shows that the training did not raise the overall probability of intending to take leave; the estimated LATE is small, negative, and statistically insignificant ($q = 0.163$). The contrast is itself informative: the training prompted concrete information-seeking and shifted flexible household margins, yet left unchanged the lumpier, career-relevant intention to take formal leave.

Two factors help interpret this estimate. First, intending to take leave is a discrete

[16] An alternative explanation is that couples reduced total unpaid work through outsourcing (for example, purchasing prepared meals) or by lowering standards for domestic chores, rather than through reallocation. Our data do not allow us to distinguish these possibilities.

decision with potentially large career consequences, and is therefore less responsive to a short attitudinal intervention than the continuous time-use margins or the low-stakes step of gathering information. Second, the question is closely tied to expecting an additional child: many participants are in their forties and do not anticipate having more children, so few are on the margin where the decision is live, which also limits statistical power among the younger men for whom it is most relevant.

Although the binary measure did not move, the training did reshape how participants reasoned about future leave-taking. It reduced the share stating they would not take leave even if they had a (new) child, while increasing the share reporting that childrearing was complete or not anticipated, a pattern consistent with the program making the time demands of childcare more salient rather than reinforcing outright opposition to leave-taking. We report these category-level results in full in Appendix B.2.3.

Workplace conditions also moderate the response, and the difference is statistically significant ($p = 0.043$; Table 6, Column (4)): the point estimate is positive for respondents with more coworkers available to cover their tasks and negative and significant for those facing tighter staffing constraints. Feasibility thus governs whether supportive attitudes translate into the leave decision: shifting beliefs may be necessary but is not sufficient without organizational support such as adequate staffing, the lever we return to in the conclusion.

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	0.155***	0.173***	-0.052	
	(0.052)	(0.049)	(0.048)	
Anderson q-value	[0.001]		[0.163]	
Training × Above-median replaceability				0.079
				(0.069)
Training × 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. 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 Local Average Treatment Effect (LATE) of training participation, using training assignment (and, in Column (4), its interactions with the indicated variables) as instrumental 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 management and worker relationships

The results above indicate that the training intervention affected participants' attitudes and household behaviors. To assess whether it also influenced broader workplace dynamics, we examine outcomes related to management practices and interpersonal relationships.

Appendix Table B.15 reports results for measures of workplace management, including the number of coworkers available to take over the respondent's tasks, time spent in meetings, and days worked from home. Appendix Table B.16 presents results for relationship indicators such as perceived psychological safety, relationships among coworkers, and willingness to recommend the firm to a job seeker.[17] Across all measures, the estimated effects are small and statistically insignificant.

Overall, the training did not produce detectable changes in these workplace outcomes, even before correcting for multiple hypothesis testing. The intervention thus appears to have influenced the personal attitudes and household margins it targeted more readily than broader workplace management practices or interpersonal relationships, which may take longer to shift than our follow-up window allowed. We return to these null results in Section 6.1, where they serve as falsification outcomes against socially desirable responding.

5.2 Effects of Information Intervention

We now examine the effects of the information intervention, which aimed to correct participants' misperceptions about coworkers' and managers' attitudes toward men taking paternity leave. As described in Section 3.5, this intervention targeted non-managerial employees and was implemented during the first follow-up survey. Participants in the treatment group were shown aggregate data on their colleagues' and supervisors' actual responses to the statement "It is good for men to take leave for childcare or caregiving," allowing us to test whether correcting second-order beliefs could shift attitudes, intentions, or subsequent behavior. Throughout this section, the primary estimand is the Average Treatment Effect (ATE), estimated by OLS. Relative to the training, which shifted both beliefs and flexible household margins, this light-touch intervention serves as a benchmark for whether correcting perceived norms alone, absent the broader persuasive and organizational content of the training, is sufficient to move downstream outcomes. A key question is therefore whether targeted belief correction by itself generates behavioral change.

Because the information was delivered and outcomes were measured within the same survey wave, the analysis captures short-run effects. Table 7 summarizes the results. Panel A reports the average treatment effect of the information treatment, and Panel B

[17] See Appendix A for variable descriptions.

shows the results of Wilcoxon rank-sum tests comparing the belief distributions between the treatment and control groups. Columns (3) and (5) show that the intervention increased second-order beliefs about coworkers and managers, by 0.111 and 0.147 points on the six-point agreement scale, respectively, relative to control-group means of 4.60 and 4.52 (sharpened $q \approx 0.13$ within the second-order belief family). The estimates move in the same direction for both coworkers and managers, and the rank-sum tests likewise indicate differences in the belief distributions between the treatment and control groups (see corresponding cumulative distributions in Appendix Figure B.4). These results indicate that the intervention reduced participants' underestimation of how positively others viewed men's paternity leave.

Despite correcting these misperceptions, the intervention produced only limited effects on downstream outcomes. The information provided referred exclusively to the first statement, "It is good for men to take leave for childcare or caregiving," and did not include any content related to the second statement, "Men's paternity leave uptake benefits the workplace." As shown in Column (1), participants' own first-order beliefs about the first statement did not significantly increase, although the point estimate is positive. The absence of a first-order belief effect is not itself unexpected for two reasons. First, the intervention was designed to correct second-order beliefs rather than to directly persuade participants of anything new. Second, participants' own first-order beliefs were already highly favorable at baseline, leaving limited room for further improvement. We also find little change in beliefs related to the second statement, as reported in Columns (2), (4), and (6).

The intervention also had no significant impact on participants' reported intention to take paternity leave in the future (Column (7)) among men aged 50 or younger.[18] This null is notable for interpreting the results: stating an intention to take leave is arguably the cheapest socially desirable response a participant could give, yet intentions do not move even immediately after respondents are shown supportive data on their colleagues. The point estimate is close to zero rather than merely imprecise (-0.007 against a control mean of 0.295), so the null reflects an absence of response rather than an inability to detect one. We also examine whether participants subsequently sought out information about paternity leave policies, measured in the second follow-up survey (Column (8)). By that point, all participants had received the training, so this outcome captures the incremental effect of the information intervention over and above the training. We find no significant effect, indicating that the information treatment did not add to the training's effect on information-seeking behavior. Appendix Table B.18 additionally explores the effects on other outcomes measured in the second follow-up survey including perceived

[18] Appendix Table B.17 reports results for the full set of response options to the original question, "Do you plan or hope to take paternity leave in the future?", and finds no significant effects on any response category.

psychological safety, relationships among coworkers, and willingness to recommend the firm to a job seeker.[19] Across all measures, the estimated effects are small and statistically insignificant.

We do not analyze household time allocation for the information intervention, for reasons of design and measurement. The information was delivered at the start of the first follow-up survey, and the time-use questions in that wave refer to typical behavior over a period preceding the survey, so no time had elapsed over which the treatment could have affected reported behavior. The second follow-up survey was designed as a short supplementary wave centered on the social desirability scale and post-training actions and did not re-collect the time-use battery.

Providing accurate information about colleagues' and managers' support for paternity leave moved participants' second-order beliefs (sharpened $q \approx 0.13$), and this belief correction did not propagate to any downstream outcome we measure, including paternity leave intentions and information-seeking behavior. In terms of the staircase of margins, the light-touch information intervention reaches only the lowest step, beliefs. Correcting misperceived norms is thus not sufficient on its own to generate behavioral change.

Finally, we assess whether the information intervention complemented the training intervention by including both treatment indicators and their interaction in a joint specification. Results reported in Appendix Table B.19 show that seven of the eight interaction terms are small and statistically insignificant, indicating that the two interventions largely operated independently. The exception is own belief that paternity leave benefits the workplace (Column (2)), where the interaction is negative and significant at the 10 percent level and offsets much of the training's own effect on that outcome, a pattern that would suggest the two interventions were partial substitutes on precisely the margin the training moved. This is one marginally significant coefficient among eight, with no same-signed interaction elsewhere.

6 Robustness Checks

6.1 Social Desirability Bias

Because our data are based on self-reported beliefs and behaviors, respondents may have tailored their answers to what they perceived as socially or organizationally desirable. Several features of the results are difficult to reconcile with a social-desirability account.

First, and most directly, the cheapest way for a participant to give the response the interventions appeared to encourage was simply to state an intention to take paternity leave. Yet stated leave intentions did not move under either intervention (Tables 6 and 7). The information treatment is especially telling: respondents did not raise their stated

[19] See 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)	(8)
Panel A: ATE								
Information	0.044	-0.007	0.111*	0.036	0.147**	0.099	-0.007	-0.026
	(0.046)	(0.077)	(0.058)	(0.065)	(0.073)	(0.061)	(0.031)	(0.035)
Anderson q-value	[1.000]	[1.000]	[0.126]	[0.171]	[0.126]	[0.126]		
Observations	589	588	579	579	587	587	425	582
Control Mean	5.339	4.678	4.599	4.602	4.517	4.498	0.295	0.220
No. clusters	65	65	65	65	65	65	57	65
PDS LASSO Control	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 second-order belief about either coworkers or managers' agreement to the indicated statement, respectively. In Column (7), we focus on the sample of respondents who are aged 50 or younger. In Column (8), 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.

intentions even immediately after being shown data documenting their colleagues' and supervisors' support for men's leave-taking. If the surveys merely elicited norm-congruent answers, making the desirable response this salient within the survey should have moved precisely this low-cost margin, but it did not.

Second, the effects we do find are differentiated rather than a uniform inflation of favorable responses. On beliefs, the training shifted participants' conviction that leave benefits the workplace but not their general approval of leave-taking. On behavior, the gains concentrate in households with young children and on weekend childcare, the margins the training content directly targeted, while weekday childcare and working hours did not move even in that group. Outside the target group the pattern is not a broad lift either: working hours and childcare are unaffected throughout, and the only movement is in housework, in scattered increases among men with only older children or no children that we read as difficult to interpret rather than as evidence of the program's reach (Table 5). A general tendency to over-report involvement would have raised these outcomes together; instead, the estimated effects track the specific content of the program, which undifferentiated social-desirability reporting would not predict. The same logic applies to outcomes the training did not target but that an image-conscious respondent would plausibly inflate after an employer-sponsored program: perceived psychological safety, coworker relationships, and willingness to recommend the firm are all unchanged (Appendix Table B.16), as are the workplace management measures in Appendix Table B.15. These serve as falsification outcomes for the social-desirability account.

Third, the survey procedures were designed to minimize incentives for image-conscious responding. The informed consent form (Appendix A.2) explained that the study was conducted independently by the University of Tokyo and the partner non-profit orga-

nization, that individual responses would remain fully confidential, and that the data would not be shared with supervisors or used in personnel evaluations. All answers were submitted directly to the research team, only anonymized and aggregated results were provided to the firms, and participants were encouraged to answer candidly.

Fourth, we directly measure each respondent's propensity for socially desirable responding. The second follow-up survey included a standard ten-item Social Desirability Scale (SDS; see Appendix Table C.1). We summed the number of socially desirable responses, defined a binary indicator, High SDS, equal to one if the score exceeds the sample median, and re-estimated the main specifications including the treatment indicator, the High SDS indicator, and their interaction. In this specification, the treatment coefficient represents the effect among low-SDS individuals and the interaction captures any differential response among high-SDS individuals. As reported in Appendix Tables C.2, C.3, and C.4, the interaction terms are mostly small and insignificant, and the estimated effects remain close to the main results, with somewhat larger standard errors. Respondents more prone to socially desirable responding thus do not drive our findings.

Two limitations remain. The SDS was collected in the second follow-up survey, which postdates the outcomes of the information intervention, so the SDS-based test applies cleanly only to the training results. More fundamentally, no design fully eliminates reporting bias that is correlated with treatment. The arguments above lead us to regard such bias as an unlikely explanation for our main results, but we cannot exclude it entirely.

6.2 Heterogeneity of Treatment Effects by Program Sign-up Rates

The overall sign-up rate was 12.4 percent across participating organizations, ranging from 4 to 44 percent. Selective sign-up does not threaten internal validity: because training was randomly assigned among employees who signed up, the estimates are unbiased for this population as long as randomization was properly implemented. What selection shapes is the interpretation of the estimand: our results identify the effect of the training among employees who chose to participate, not among all employees. The direction of this selection is not obvious a priori. Participants may already be more favorably disposed toward paternal involvement, leaving less room for the training to move their attitudes and behavior; alternatively, they may have greater latent demand for change that the training helps to realize, in which case effects would be larger than in the broader workforce. Selection could therefore either attenuate or amplify the estimates relative to a workforce enrolled by default, and we cannot sign its direction. We note, however, that the effect among voluntary participants is the policy-relevant parameter for the way such programs are ordinarily deployed, since employer-sponsored training is almost always offered on an opt-in basis. Our LATE captures the returns among employees induced to participate under the recruitment conditions studied here, which is what a

firm considering the same offer needs to know. It does not extend to the additional men reached by stronger promotion, incentives, or a mandate, who are precisely those who declined here.

Participation varied widely across organizations, largely reflecting how actively each human resources department promoted the program; participation was voluntary and carried no monetary incentives, which partner organizations considered inconsistent with their internal norms. This variation shows that participation is not a fixed feature of the program but a margin that firms can influence through encouragement or incentives, neither of which we sought to optimize. The sizable per-participant effects therefore provide a relevant input into an organization's decision about whether to invest in participation rather than merely offer the program. They do not, however, identify the effects among employees who would participate only under stronger promotion or incentives, so the population-level gains from such investments need not scale proportionally with participation.

To verify the integrity of randomization, we first confirm that office-level sign-up and survey response rates are not systematically associated with training assignment status (Appendix Table B.2). We then assess the robustness of our main results by replicating key analyses using data from the organization with the highest sign-up rate (44 percent), whose participants are the least selected in our sample. The behavioral estimates (Appendix Table C.5) and the information-treatment estimates (Appendix Table C.7) are similar in magnitude to the baseline estimates but, in a substantially smaller sample, are no longer statistically significant: the estimated effect on fathers' weekend childcare, for instance, is 1.17 hours here against 0.94 in the full sample, with a standard error nearly twice as large. The training's effects on beliefs (Appendix Table C.6) are larger than in the full sample and remain significant. Neither pattern suggests that our main findings are an artifact of variation in program participation rates across organizations; if anything, the estimates are no weaker in the organization where sign-up was broadest and selection into the program correspondingly milder.

7 Conclusion

This study provides experimental evidence that workplace interventions can promote paternal involvement in childcare even where formal family policies are already well developed. Using a randomized controlled trial across four Japanese organizations, we show that an employer-sponsored two-hour training program for employees and managers increased fathers' weekend childcare time, robustly so in households with young children, and led their spouses to work longer hours. In contrast, an information intervention that corrected misperceptions about colleagues' attitudes moved second-order beliefs only weakly and had no detectable effect on paternity leave intentions or subse-

quent information-seeking behavior. These results indicate that formal policies alone are insufficient to alter behavior and that workplace-level interventions targeting attitudes and norms can contribute to achieving greater gender equality in unpaid work.

The findings have several implications for the design of work-family policies. Interventions that target workplace norms can complement traditional policy tools such as statutory leave entitlements and financial incentives. Legal rights and monetary supports are necessary, but they are often insufficient when workplace norms discourage men from using them. Our results suggest that workplace-based training can help close this gap when it combines attitudinal content with visible employer support. Including managers in the training may further strengthen this signal, though we cannot separately identify their contribution from our data. These changes followed a single two-hour session delivered online during working hours, so the resources required per participant are modest relative to the statutory entitlements and childcare capacity that dominate current policy. Participation is itself a policy margin. Because participation in the training was modest and we did not try to maximize it, it remains a margin that organizations can influence through stronger promotion or incentives. The sizable per-participant effects establish the potential value of reaching voluntary participants, although the population-level gains from broader participation will depend on the effects among the additional employees induced to participate.

The two interventions trace a clear ordering of margins by adjustment cost. The information treatment moved second-order beliefs only weakly and reached nothing beyond them, while the training additionally shifted the flexible margins of household time but not the intention to take formal leave. Moving that costlier, career-relevant margin likely requires stronger organizational levers than attitudes alone, such as staffing arrangements that allow colleagues to cover for absent workers and managerial accountability for enabling leave.

The increase in spouses' working hours points to a substantive externality. The training increased spouses' working hours alongside fathers' weekend childcare time, contributing to higher overall labor supply and potentially to household income, though we do not identify fathers' childcare as the cause of their spouses' additional hours. Because these gains to spouses accrue outside the firm, the sponsoring employer does not capture them, creating a potential rationale for public support of employer-sponsored paternity leave training.

The information intervention is inexpensive and easy to deploy, yet it did not produce detectable effects on paternity leave intentions or information-seeking behavior in our setting. This suggests that correcting norm misperceptions alone may not be sufficient to change behavior, and that more active or intensive forms of engagement may be needed to translate updated beliefs into action.

This study has two main limitations. First, our follow-up period extends only to sev-

eral months after treatment, leaving open whether the behavioral changes we document persist over time, and whether fathers who increased their childcare involvement experience longer-term effects on their own careers or on their children's outcomes; because the crossover design eventually treats every office, we cannot make this long-run comparison between trained and never-trained employees. Second, while the four participating organizations are broadly comparable to the national average in Japan, whether similar interventions would be effective in other national contexts, where institutional arrangements and norms around fatherhood differ, remains an open question. Both limitations point to directions for future research.

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.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman (2020). Women's empowerment in action: Evidence from a randomized control trial in Africa. *American Economic Journal: Applied Economics 12*(1), 210–259.

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. Cortes, J. Pan, H. Shigeoka, and M. Yagasaki (2024). Understanding the barriers to paternity leave-taking: Evidence from Japan. Working paper.

Bohnet, I., A. van Geen, and M. H. Bazerman (2016). When performance trumps gender bias: Joint versus separate evaluation. *Management Science 62*(5), 1225–1234.

Boneva, T., T. Buser, A. Falk, and F. Kosse (2024, May). The origins of gender differences in competitiveness and earnings expectations: Causal evidence from a mentoring intervention. Discussion Paper 301, ECONtribute: Markets and Public Policy, University of Bonn and University of Cologne.

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.

Chang, E. H., K. L. Milkman, D. M. Gromet, R. W. Rebele, C. Massey, A. L. Duckworth, and A. M. Grant (2019). The mixed effects of online diversity training. *Proceedings of the National Academy of Sciences 116*(16), 7778–7783.

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.

Doepke, M. and F. Kindermann (2019). Bargaining over babies: Theory, evidence, and policy implications. *American Economic Review 109*(9), 3264–3306.

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.

Goldin, C. and C. Rouse (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review 90*(4), 715–741.

Guiso, L., P. Sapienza, and L. Zingales (2015). The value of corporate culture. *Journal of Financial Economics 117*(1), 60–76.

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.

Macchiavello, R., A. Menzel, A. Rabbani, and C. Woodruff (2024, October). Promoting women to managerial roles in the bangladeshi garment sector. Working Paper.

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 ikujikyuugyo 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.

Moen, P., E. L. Kelly, W. Fan, S.-R. Lee, D. Almeida, E. E. Kossek, and O. M. Buxton (2016). Does a flexibility/support organizational initiative improve high-tech employees' well-being? evidence from the work, family, and health network. *American Sociological Review 81*(5), 1–31.

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.

Sarkadi, A., R. Kristiansson, F. Oberklaid, and S. Bremberg (2008). Fathers' involvement and children's developmental outcomes: a systematic review of longitudinal studies. *Acta Paediatrica 97*(2), 153–158.

Tamm, M. (2019). Fathers' parental leave-taking, childcare involvement, and maternal labor market participation. *Labour Economics 59*, 184–197. Germany.

UNICEF (2022). Parental leave in oecd countries: Key findings. Technical report, UNICEF.

Yamaguchi, S., Y. Asai, and R. Kambayashi (2018a). Effects of subsidized childcare on mothers' labor supply under a rationing mechanism. *Labour Economics 55*, 1–17.

Yamaguchi, S., Y. Asai, and R. Kambayashi (2018b). How does early childcare enrollment affect children, parents, and their interactions? *Labour Economics 55*, 56–71.

Supplemental Appendix

A Details of Experiments and Survey Questions

A.1 Invitation to Program

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 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. It 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 has been processed into a form that does not identify you personally on the part of the researchers. Please be honest in your responses, as they will not be shared with everyone's 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.

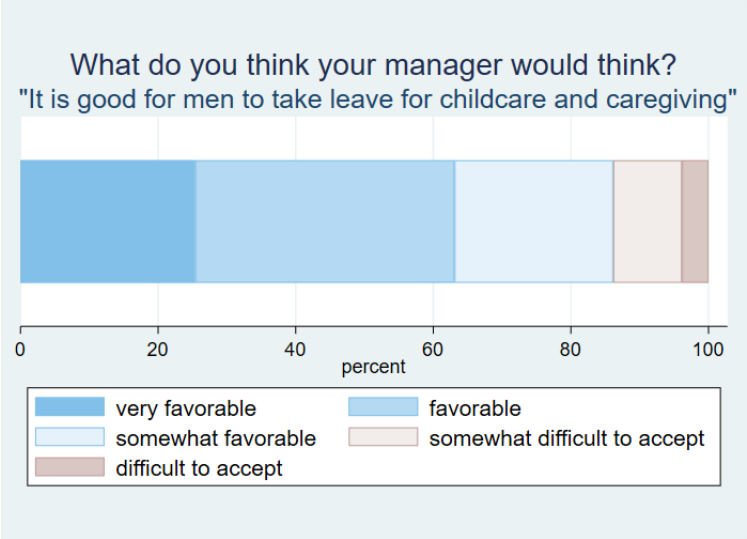
A.3 Information Provision (First Follow-up Survey)

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 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.



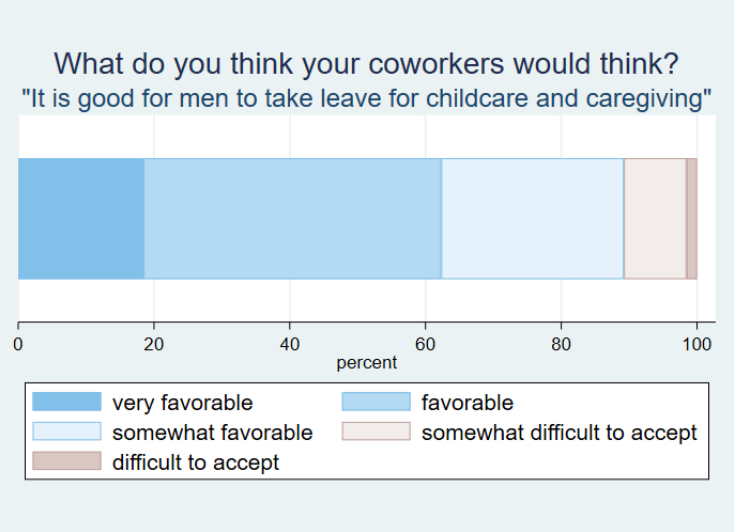
On the other hand, when responses from supervisors at [ORGANIZATION NAME] were analyzed, **94%** of them indicated that they "strongly agree" or "agree" with this view.



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 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.

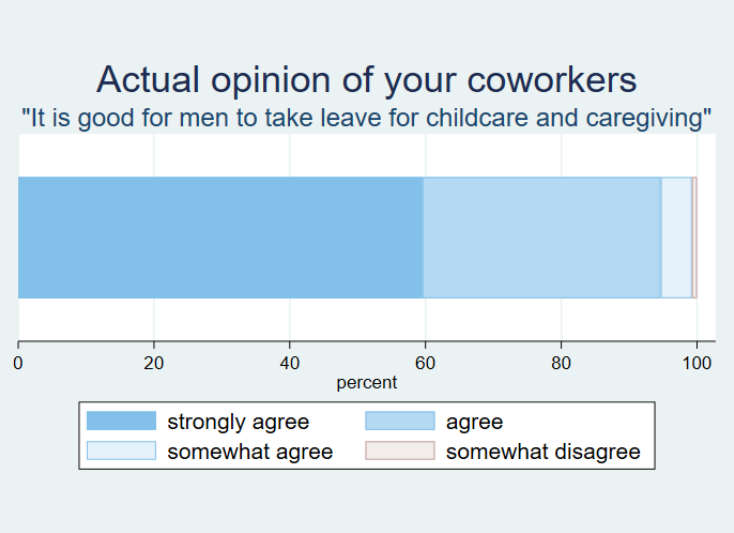


Image shown to treatment group (Page 3) and control group

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.4 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)	*It is a good thing for men to take leave for childcare and caregiving.*	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)	*Increasing take-ups of men's paternity leave can improve company operations and promote a collaborative workplace culture.*	1	2	3	4	5	6

Perception toward managers (colleagues)' belief- 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)	*Men take leave for childcare and caregiving.*	1	2	3	4	5	6

Perception toward managers (colleagues)' belief- 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)	*Increasing take-ups of men's paternity leave can improve company operations and promote a collaborative workplace culture.*	1	2	3	4	5	6

A.5 Questions about Hours

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.6 Questions about paternity leave

Intention to take paternity leave

Do you plan 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.7 Questions about Workplace Management 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 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 ratio 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 population 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

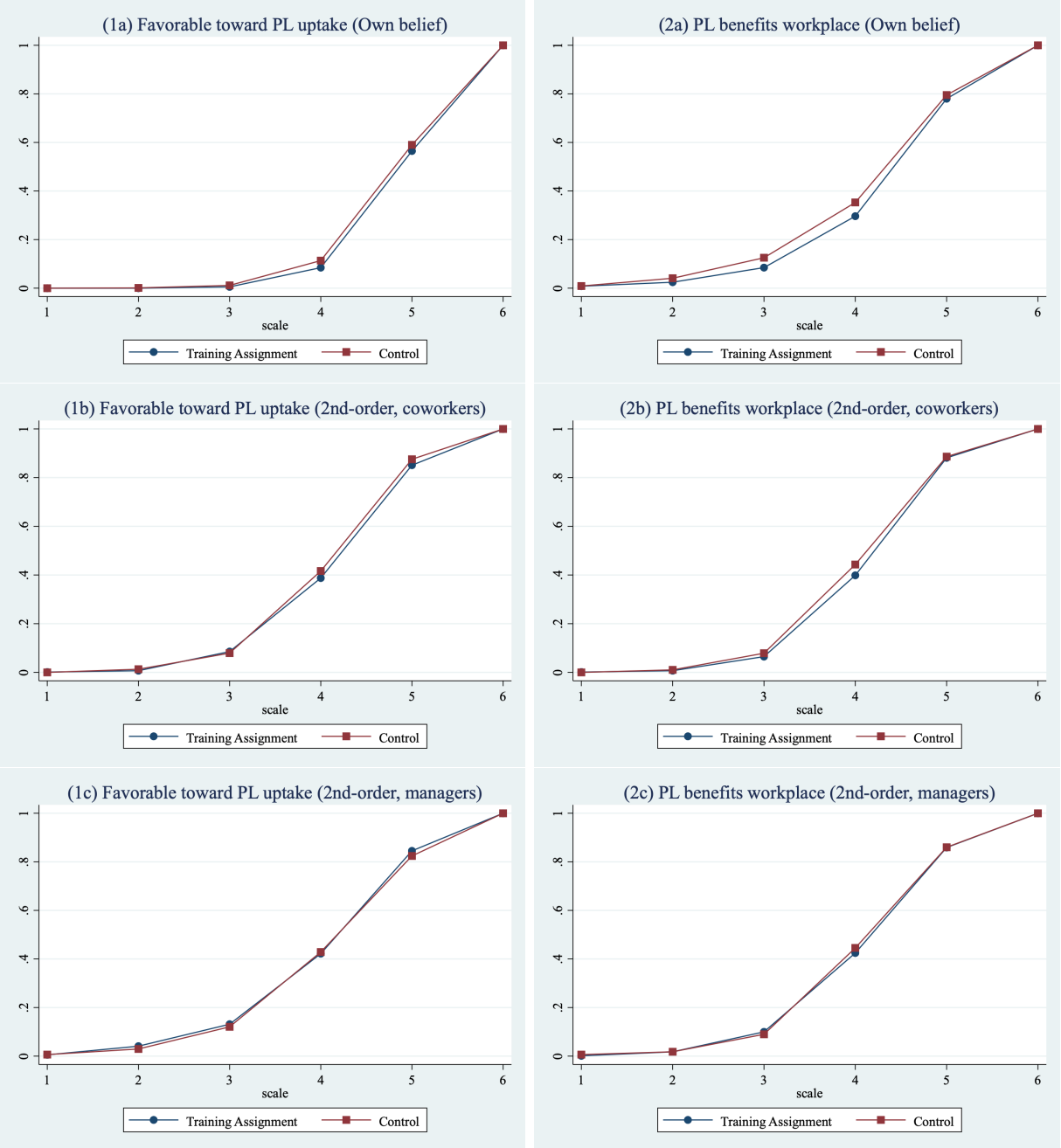
	Program Sign-up Rate		Survey Response Rate Baseline		Survey Response Rate 1st Follow-Up		Survey Response Rate 2nd Follow-up	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Training	0.028	0.021	0.051	0.050	-0.027	-0.027	0.038	0.038
	(0.076)	(0.066)	(0.037)	(0.037)	(0.050)	(0.051)	(0.061)	(0.062)
Observations	76	76	80	80	80	80	80	80
Strata FE	No	Yes	No	Yes	No	Yes	No	Yes
Mean Dep. Var	0.259	0.259	0.930	0.930	0.806	0.806	0.592	0.592

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 baseline and the first 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. 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 and Workplace Norms

Figure B.1: Cumulative Distribution: Training Treatment vs. Control in 1st Follow-up



Notes. Data on first- and second-order beliefs measured in 1st follow-up survey.

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

	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	0.031	0.182***	0.062	0.119*	-0.036	0.040
	(0.023)	(0.048)	(0.060)	(0.061)	(0.044)	(0.048)
Anderson q-value	[0.096]	[0.001]	[0.460]	[0.253]	[0.460]	[0.460]
Observations	1225	1223	1206	1208	1218	1218
Baseline Mean	5.196	4.553	4.512	4.330	4.443	4.323
No. clusters	74	74	74	74	74	74
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table reports the intent-to-treat (ITT) effects of the training intervention on own beliefs of the male respondents and their perception about workplace norms. The coefficients are obtained by regressing the outcome variables on an indicator for assignment of the office to the early training group. 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) 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	0.102*	0.446***	0.143	0.288**	-0.105	0.093
	(0.061)	(0.106)	(0.146)	(0.144)	(0.115)	(0.120)
Anderson q-value	[0.053]	[0.001]	[0.498]	[0.245]	[0.498]	[0.498]
Observations	1225	1223	1206	1208	1218	1218
Baseline Mean	5.196	4.553	4.512	4.330	4.443	4.323
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) 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 × Higher Baseline belief	0.031	0.317	0.228	0.262*	-0.306	0.220*
	(0.098)	(0.273)	(0.237)	(0.156)	(0.285)	(0.121)
Training × Lower Baseline belief	0.114*	0.489***	0.147	0.317*	-0.064	-0.025
	(0.067)	(0.113)	(0.144)	(0.168)	(0.110)	(0.192)
Observations	1225	1223	1206	1208	1218	1218
No. clusters	74	74	74	74	74	74
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Mean	5.196	4.553	4.512	4.330	4.443	4.323
P-val: High=Low	0.486	0.569	0.694	0.724	0.332	0.282

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 (below) the median among the baseline male sample. The specification follows that of Table 3, except that the training participation is interacted with these baseline belief indicators and training assignment interacted with the same indicator variables are used as instrumental variables.

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
Panel A: Manager / Non-manager						
Training × Manager	0.100	0.381***	0.125	0.170	-0.007	-0.049
	(0.066)	(0.112)	(0.158)	(0.155)	(0.120)	(0.116)
Training × Non-manager	0.054	0.555***	0.181	0.477**	-0.226	0.294
	(0.091)	(0.200)	(0.195)	(0.225)	(0.232)	(0.239)
Manager	-0.032	-0.005	-0.015	-0.001	-0.078	0.069
	(0.034)	(0.071)	(0.041)	(0.064)	(0.082)	(0.093)
Observations	1225	1223	1206	1208	1218	1218
No. clusters	74	74	74	74	74	74
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Mean	5.196	4.553	4.512	4.330	4.443	4.323
P-val: Manager=Non-manager	0.661	0.455	0.779	0.227	0.410	0.199
Panel B: Have a child / No child						
Training × Children	0.080	0.446***	0.168	0.268*	-0.029	0.043
	(0.061)	(0.104)	(0.149)	(0.150)	(0.093)	(0.112)
Training × No children	0.092	0.476	0.107	0.395*	-0.308	0.346
	(0.145)	(0.337)	(0.214)	(0.235)	(0.361)	(0.273)
Child	-0.043	0.022	-0.054	-0.004	-0.144	-0.038
	(0.058)	(0.122)	(0.077)	(0.100)	(0.109)	(0.093)
Observations	1225	1223	1206	1208	1218	1218
No. clusters	74	74	74	74	74	74
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Mean	5.196	4.553	4.512	4.330	4.443	4.323
P-val: Child=No Child	0.937	0.933	0.765	0.599	0.426	0.246

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 the training participation is interacted with these baseline characteristics and training assignment interacted with the same indicator variables are used as instrumental variables.

B.2.2 Behavioral Outcomes

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

	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)
Panel A: Own hours					
Training assignment	-0.325	-0.045	0.368*	0.004	0.049
	(0.712)	(0.089)	(0.190)	(0.046)	(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
Panel B: Spouse hours					
Training assignment	1.473**	-0.002	0.079	-0.213	-0.035
	(0.664)	(0.163)	(0.188)	(0.162)	(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 intent-to-treat (ITT) 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. Panel A shows the results for male respondents, and Panel B shows the results for their spouses. Standard errors in parentheses are clustered at 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.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)
Panel A: Own hours					
Training	0.013	-0.074	1.381**	-0.045	0.033
	(1.899)	(0.244)	(0.527)	(0.115)	(0.239)
Anderson q-value		[1.000]	[0.045]	[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
Panel B: Spouse hours					
Training	3.578**	0.326	0.647	-0.493	-0.082
	(1.746)	(0.449)	(0.580)	(0.396)	(0.350)
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 (ITT)

	(1) Weekly Working Hours	(2) Daily Childcare Hours (Weekday)	(3) Daily Childcare Hours (Weekend)	(4) Daily Housework Hours (Weekday)	(5) Daily Housework Hours (Weekend)
Panel A: Own hours					
Training × Young children	-1.810	-0.327	1.031**	-0.135	-0.399
	(1.234)	(0.218)	(0.403)	(0.0988)	(0.331)
Training × Only older children	-0.0724	0.101	0.133	0.155**	0.180
	(1.065)	(0.0974)	(0.234)	(0.0616)	(0.113)
Training × No children	0.821			0.0650	0.258**
	(2.664)			(0.108)	(0.118)
Observations	1015	859	859	1001	1001
No. clusters	74	72	72	74	74
Baseline Mean	45.591	1.455	6.728	0.880	2.041
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes
Panel B: Spouse hours					
Training × Young children	1.982*	0.140	-0.206	-0.605*	-0.936**
	(1.135)	(0.386)	(0.541)	(0.338)	(0.435)
Training × Only older children	0.714	0.0824	0.177	-0.136	0.259**
	(0.808)	(0.192)	(0.311)	(0.188)	(0.130)
Training × No children	4.271			0.227	0.568
	(2.843)			(0.440)	(0.372)
Observations	994	846	846	980	980
No. clusters	74	72	72	74	74
Baseline Mean	21.757	5.101	7.915	4.090	4.018
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. Standard errors in parentheses are clustered at 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.10: Impact of training on behaviors: HTE by manager/non-manager

	(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 × Manager	-0.699	0.070	0.749	0.170	0.158	0.153***
	(1.477)	(0.277)	(0.514)	(0.108)	(0.304)	(0.057)
Training × Non-manager	-1.349	-0.080	1.417	-0.195	0.001	0.170**
	(3.555)	(0.417)	(0.991)	(0.250)	(0.390)	(0.083)
Observations	1177	876	876	1161	1161	1211
No. clusters	74	72	72	74	74	74
Baseline Mean	45.140	1.443	6.638	0.912	2.049	0.180
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
P-values: Coefficients difference	0.865	0.766	0.574	0.200	0.748	0.849

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 the training participation is interacted with indicators for managers and non-managers, and training assignment interacted with the same indicator variables are used as instrumental variables. 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) >0	(2) >10	(3) >20	(4) >30	(5) >40	(6) >50	(7) >60
Panel A: Respondents							
Training	0.007	0.040	0.017	0.009	0.041	-0.034	-0.036
	(0.006)	(0.033)	(0.038)	(0.040)	(0.062)	(0.041)	(0.028)
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
Panel B: Spouse							
Training	0.016	0.091*	0.149**	0.031	-0.007	0.006	0.019*
	(0.037)	(0.048)	(0.060)	(0.047)	(0.032)	(0.015)	(0.010)
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

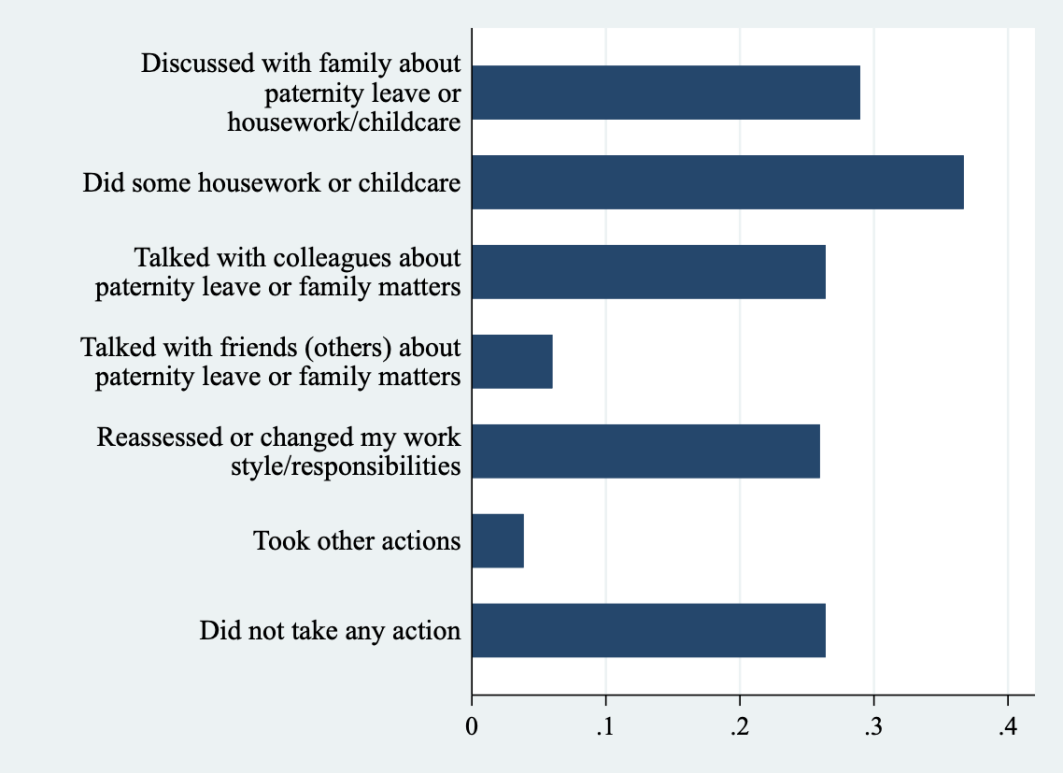
Notes. The dependent variables are indicator variables that take 1 if weekly working hours exceed the threshold indicated in each column and 0 otherwise. Panel A reports results for the respondents and Panel B for their spouses, corresponding to Panels A and B of Table 4. We control for the baseline value of the same indicator variable in all equations. The specification follows that of Table 4.

Table B.12: Behavioral changes: Total hours of couple

	(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
Panel A: LATE							
Training	2.868	-0.198	1.188	0.981	-0.401	-0.0784	-2.005
	(2.920)	(0.456)	(0.794)	(3.129)	(0.427)	(0.464)	(2.769)
Observations	994	845	845	845	979	979	979
Panel B: Heterogeneity by Child Presence and Age							
Training × Young children	1.738	-0.361	1.864	1.838	-1.696**	-3.274**	-14.83**
	(4.553)	(0.970)	(1.813)	(7.177)	(0.751)	(1.335)	(5.929)
Training × Only older children	2.222	0.133	0.720	2.052	-0.0377	1.008**	1.996
	(4.292)	(0.572)	(1.209)	(4.525)	(0.492)	(0.514)	(2.966)
Training × No children	13.49				0.825	2.350	9.041
	(13.73)				(1.243)	(1.494)	(8.602)
Observations	994	845	845	845	979	979	979
No. clusters	74	72	72	72	74	74	74
Baseline Mean	67.464	6.564	14.661	62.142	4.954	6.058	36.885
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 Local Average Treatment Effect of training participation, and Panel B reports heterogeneity by the presence and age of children, using the subgroup definitions of Table 5. The number of clusters, baseline 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 hope to take paternity leave in the future?"[20] 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 within 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.

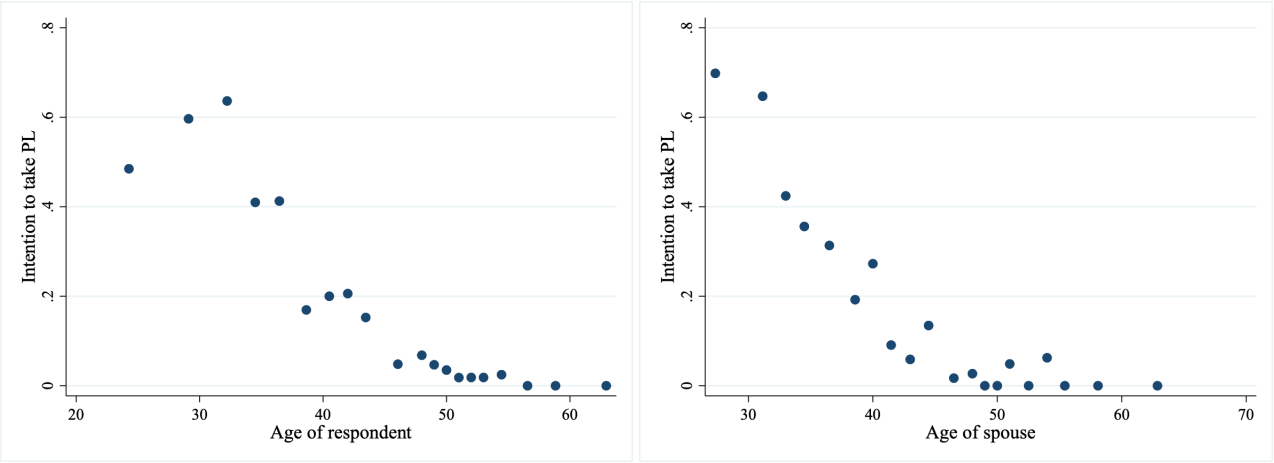
Table 6 reports the estimated LATE (see Table B.13 for ITT estimates). Column (3) shows no significant average effect of the training on reported intentions ($q = 0.163$). A likely explanation is sample composition: many participants are in their forties and may not expect to have additional children.

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 significant divergence ($p = 0.043$): the point estimate is positive among those with more potential substitutes and negative and significant among those with fewer. 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.

[20] Appendix 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 (ITT)

	Researched about PL		Intention to take PL	
	(1)	(2)	(3)	(4)
Training assignment	0.061***	0.075***	-0.023	
	(0.023)	(0.022)	(0.020)	
Training assignment × Above-median replaceability				0.042
				(0.033)
Training assignment × 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. The specification follows that of Table 6, except that here we estimate the intent-to-treat (ITT) effects of training assignment. The variable "training assignment" is defined as an indicator for being assigned to the early training group.

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 participation	-0.052	-0.068**	0.160**	-0.037
	(0.048)	(0.029)	(0.068)	(0.066)
Observations	775	775	775	775
No. clusters	69	69	69	69
Baseline Mean	0.272	0.059	0.459	0.209
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 hope to take paternity 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 drop the sample 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.

B.2.4 Workplace Outcomes

Table B.15: Workplace management

	(1) Num. workers for my task today	(2) Num. workers for my task in one month	(3) Meeting hours	(4) Days work from home
Training	-0.0418	-0.0227	-0.3642	-0.0684
	(0.1918)	(0.2164)	(1.2361)	(0.0438)
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 management 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. The coefficients are estimated using training assignment as the instrumental variable and represent the Local Average Treatment Effect (LATE) of training participation. 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.

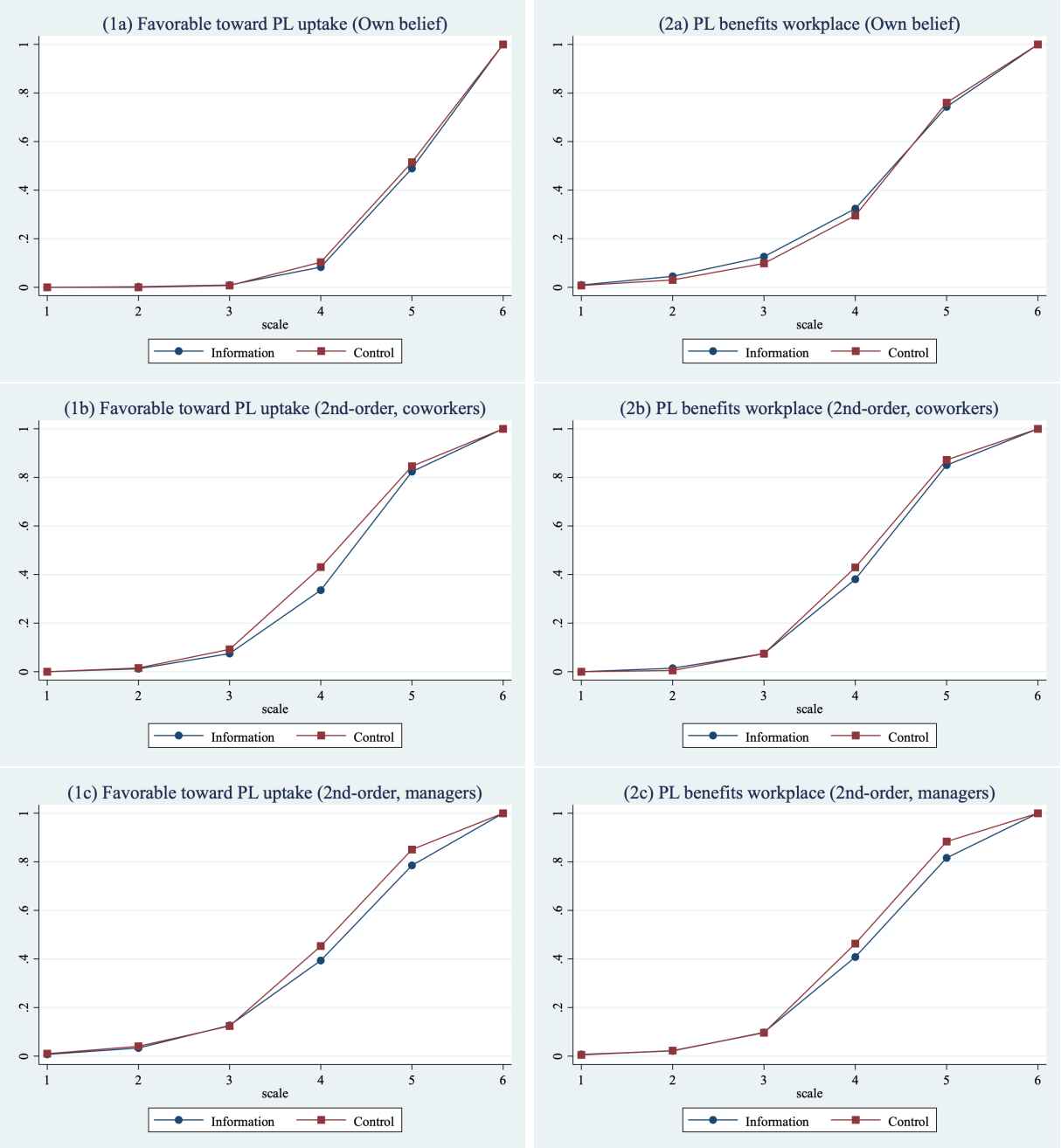
Table B.16: Workplace relationship

	(1) Manager knows family composition	(2) Psychological safety	(3) Coworkers relationships	(4) Recommend firm to jobseeker
Training	-0.006	0.005	0.070	0.012
	(0.037)	(0.139)	(0.161)	(0.062)
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 coworker relationships, measured on a scale. Column (4) is an indicator variable equal to 1 if the respondent would recommend the firm to a job seeker. The coefficients are estimated using training assignment as the instrumental variable and represent the Local Average Treatment Effect (LATE) of training participation. 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 1st follow-up survey.

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	0.004	0.036	-0.021
	(0.031)	(0.013)	(0.037)	(0.037)
Observations	425	425	425	425
No. clusters	57	57	57	57
Baseline Mean	0.376	0.071	0.287	0.266
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 hope to take paternity 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 drop the sample 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.

Table B.18: Impact of information treatment on outcomes measured in second follow-up survey

	(1)	(2)	(3)
	Psychological Safety	Work relationship	Recommend firm to jobseekers
Information	0.0777	0.148	-0.00735
	(0.0632)	(0.0985)	(0.0391)
Observations	410	410	405
Baseline Mean	4.000	3.835	0.511
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 workplace relationships, 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.19: Interaction of information and training on beliefs, intention, and information-seeking

	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)	(8)
Training assignment	-0.046	0.295***	0.102	0.118	-0.131*	0.116	0.038	0.014
	(0.047)	(0.090)	(0.089)	(0.108)	(0.075)	(0.104)	(0.036)	(0.032)
Information	0.002	0.108	0.159**	0.009	0.089	0.127	0.016	-0.060
	(0.072)	(0.086)	(0.068)	(0.087)	(0.091)	(0.088)	(0.042)	(0.038)
Training assignment × Information	0.094	-0.232*	-0.100	0.076	0.121	-0.051	-0.045	0.082
	(0.081)	(0.136)	(0.118)	(0.119)	(0.139)	(0.118)	(0.062)	(0.066)
Observations	589	588	579	579	587	587	425	582
Control Mean	5.329	4.523	4.516	4.539	4.542	4.439	0.295	0.213
No. clusters	65	65	65	65	65	65	57	65
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table examines complementarity of training and information interventions. The specification is the same as Table 7 except that training intervention 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, ensuring it captures the inclination to present oneself favorably rather than 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	0.104	0.386**	0.065	0.247	-0.190	0.032
	(0.071)	(0.166)	(0.131)	(0.157)	(0.162)	(0.130)
Training × High SDS	-0.094	-0.166	-0.044	-0.012	-0.202	0.006
	(0.159)	(0.285)	(0.269)	(0.237)	(0.258)	(0.181)
High SDS	0.051	0.162	0.005	0.060	0.131	0.091
	(0.057)	(0.102)	(0.100)	(0.099)	(0.082)	(0.072)
Observations	882	880	872	872	880	879
No. clusters	69	69	69	69	69	69
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Mean	5.223	4.560	4.517	4.337	4.444	4.342

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 interacting training participation with an indicator variable for the High SDS indicator and using training assignment with the same indicator variable as an additional instrumental variable.

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) 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	0.247***	0.074	0.117	0.105	0.190**	0.246***
	(0.068)	(0.146)	(0.093)	(0.091)	(0.094)	(0.079)
Information × High SDS	-0.338***	-0.058	-0.027	-0.094	0.080	-0.090
	(0.087)	(0.185)	(0.127)	(0.095)	(0.154)	(0.145)
High SDS	0.224***	0.096	0.013	0.122	0.030	0.208**
	(0.075)	(0.118)	(0.091)	(0.095)	(0.085)	(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.

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

Panel A: Own 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)	(6) Researched about PL
Training	0.152	-0.343	0.978	-0.0969	-0.283	0.123
	(2.612)	(0.246)	(0.675)	(0.162)	(0.345)	(0.0817)
Training × High SDS	-0.709	0.419	-0.492	0.178	0.667	0.0361
	(3.626)	(0.459)	(1.122)	(0.275)	(0.714)	(0.125)
High SDS	0.401	0.150	0.152	-0.0259	-0.149	0.00186
	(1.631)	(0.187)	(0.478)	(0.107)	(0.236)	(0.0315)
Observations	859	639	639	852	852	874
No. clusters	69	66	66	69	69	69
Baseline Mean	44.640	1.492	6.701	0.901	2.047	0.192
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Spouse's 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)
Training	3.278	0.393	-0.228	-0.558	-0.0186
	(2.369)	(0.603)	(0.888)	(0.630)	(0.452)
Training × High SDS	-0.542	0.128	0.920	0.255	0.190
	(3.644)	(1.051)	(1.561)	(0.880)	(0.817)
High SDS	0.794	-0.116	-0.422	-0.426	0.0931
	(1.835)	(0.408)	(0.714)	(0.336)	(0.296)
Observations	724	616	616	718	718
No. clusters	68	65	65	67	67
Baseline Mean	21.468	5.071	7.912	4.056	3.987
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 interacting training participation with an indicator variable for the High SDS indicator and using training assignment with the same indicator variable as an additional instrumental variable. 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 firm with highest program sign-up rate

Table C.5: Impact of Training on Behaviors: Sub-sample of firm with the highest program sign-up rate

Panel A: Own 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)	(6) Researched about PL
Training	1.096	0.242	1.167	-0.312	-0.317	0.086
	(3.873)	(0.405)	(0.797)	(0.241)	(0.360)	(0.091)
Observations	580	406	406	574	574	599
No. clusters	34	33	33	34	34	34
Baseline Mean	45.195	1.261	5.851	0.817	1.900	0.097
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Spouse's 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)
Training	5.543	0.303	-0.028	-0.991	0.191
	(3.801)	(0.685)	(0.899)	(0.704)	(0.637)
Observations	463	389	389	459	459
No. clusters	34	33	33	34	34
Baseline Mean	21.583	4.394	7.080	4.068	3.976
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes

Notes. In this table, we use the sample of a firm that achieved the highest program sign-up rate (44%). 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.6: Impact of Training on Beliefs: Sub-sample of firm 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	0.224***	0.621***	0.390*	0.614***	-0.259	0.057
	(0.077)	(0.159)	(0.230)	(0.172)	(0.271)	(0.189)
Observations	606	604	596	595	600	601
No. clusters	34	34	34	34	34	34
PDS LASSO Control	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Mean	5.031	4.454	4.337	4.225	4.115	4.108

Notes. In this table, we use the sample of a firm that achieved the highest program sign-up rate (44%). The rest of the specification is the same as that of Table 3.

Table C.7: Impact of information on beliefs: Sub-sample of firm 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	0.017	0.111	0.008	0.132	0.123*
	(0.069)	(0.085)	(0.086)	(0.078)	(0.125)	(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 a firm that achieved the highest program sign-up rate (44%). The rest of the specification is the same as that of Table 7.