Managing Culture

Silvia Fernández Castro

Dissertation

Munich 2025

Ludwig-Maximilians-Universität

Managing Culture

Inaugural-Dissertation
zur Erlangung des Grades

Doctor oeconomiae publicae
(Dr. oec. publ.)

an der
Ludwig–Maximilians–Universität München

2025

vorgelegt von
Silvia Fernández Castro

Referent: Prof. Dr. Florian Englmaier
Korreferentin: Prof. Dr. Maria Guadalupe
Promotionsabschlussberatung: 16. Juli 2025

Datum der mündlichen Prüfung: 01. Juli 2025

Namen der Berichterstatter:

Prof. Dr. Florian Englmaier

Prof. Dr. Maria Guadalupe

Prof. Dr. Kristina Czura

Acknowledgements

This dissertation would not have been possible without the support of many people. It has been a journey of immense learning, and I am deeply grateful to those who have been there over the years.

First, I thank my first supervisor, Florian Englmaier, for his guidance in helping me navigate challenges and always pointing me in the right direction. I am also grateful to my second doctoral advisor, Maria Guadalupe, who has generously taken me under her wing despite being at a different school. I owe special gratitude to Kristina Czura, my third supervisor, who made me a researcher even before I began my doctoral studies.

Many others have guided me along the way. In particular, I appreciate Anik Ashraf, Karen Bernhardt-Walther, Vojtech Bartos, Guido Friebel, Stephen Leider, Mathias Lang, Rembrand M. Koning, Simeon Schudy, and Joachim Winter for their mentorship.

I also deeply appreciate my co-authors—Clarissa, Hoa, and Maren—who made it fun, and my peers—Carolin, Dominik, Friederike, Jae, Lena, Nathalie, Peter, and Svenja—who have provided support.

This research would not have been possible without Manuela Beckstein and Manuela Firmin, whose administrative support was essential to the success of the field experiments.

I am especially grateful to my field partners—Emily, Nelson, Robert, Solomy, Stefanie, and Martina—for hosting the projects and assisting me in running them.

Finally, I thank my parents, sister, family, and friends, who, despite the distance, have always supported me.

To all who contributed in ways big and small—thank you.

Contents

Preface

Humans are a uniquely cooperative species. We collaborate in large groups, often with people we do not know, to achieve goals that no individual could accomplish alone. Cooperation has shaped human history, from early societies working together to hunt and build settlements to modern economies where millions of people contribute to complex projects like manufacturing airplanes or launching spacecraft. The ability to work together is one of our greatest strengths—it drives innovation, economic growth, and social progress.

But cooperation does not happen automatically. It is guided by culture—the shared norms, values, and informal rules that shape how we interact. Culture helps people trust each other, work towards common goals, and solve problems without needing constant negotiation. It exists in all parts of life, from the way teams collaborate within a firm to the broader societal expectations that govern behavior in different countries.

Understanding how culture shapes cooperation can help explain why some organizations perform better than others, why certain policies succeed or fail, and how trust and shared values influence economic decisions. This is the focus of this dissertation.

At its core, this dissertation is about people—how they help each other, build trust, and overcome barriers to working together. Through three chapters, it empirically examines how culture and cooperation interact across different organizations and continents. It explores how workplace culture influences peer collaboration, how managers shape team culture, and how cultural norms affect health-related decisions. The findings highlight that culture is not fixed—it can be shaped and strengthened to create better environments for people to work, learn, and live.

The first chapter studies how organizations can incentivize a culture of employee cooperation. Using a randomized controlled trial in a large bank, it evaluates the impact of a public recognition award for peer support on knowledge-sharing. The study empiri-

cally examines the interpersonal dynamics of tacit knowledge exchange, showing how social image—shaped by peers—and management-related reputation serve as strong incentives for cooperative behavior.

The second chapter examines how leadership behaviors shape team culture. Using a randomized controlled trial in a large global pharmaceutical company, it evaluates how small managerial shifts affect employees' perception of their team as a safe space to speak up openly and experiment with new ways of getting things done. The study shows that leadership behaviors are not fixed but can be shaped to foster a more open and supportive workplace culture.

The third chapter examines how culture and social norms shape women's health decisions. Using a field experiment in a factory in Bangladesh, we study how stigma and shame surrounding menstruation act as barriers to adopting menstrual health technologies, even when they are freely available in the workplace. The study shows that group discussions fostering open communication and normalizing menstruation encourage women to adopt these technologies. The findings prove that deeply embedded restrictive social norms influence personal health choices and demonstrate that shifting cultural perceptions can drive meaningful behavioral change.

Together, these chapters explore how culture shapes the way humans interact and assist one another across different contexts, from the workplace to personal health decisions. They highlight the mechanisms that enable individuals to support one another, as well as the barriers—whether rooted in social stigma, organizational norms, or incentive misalignment—that can hinder collaboration. By identifying ways to develop a supportive and cooperative culture, this research offers insights for organizations, policymakers, and practitioners seeking to create environments where people can work, innovate, and make decisions that improve both individual and collective well-being.

Understanding how we work together is not just an academic pursuit—it is essential for addressing challenges in organizations and society. As workplaces evolve and global interdependence grows, the ability to build cultures that encourage trust, communication, and collective action becomes increasingly important. This dissertation contributes to that effort by shedding light on the social and cultural dynamics that drive cooperation and by providing evidence-based strategies to foster it in practice.

1

Cooperation in the Workplace

Experimental Evidence from Knowledge Workers

1.1 Introduction

Organizations depend on employee collaboration. Much of the expertise within firms is tacit —difficult to codify and often acquired only through experience. This tacit knowledge is transferred through peer-to-peer exchange (Battiston et al., 2021). When experienced employees share job-specific knowledge with their peers, they reduce inefficiencies, enhance productivity, and accelerate learning (Hamilton et al., 2003; Sandvik et al., 2020; Battiston et al., 2021).

Yet, fostering cooperative behaviors is challenging. In many organizations, meaningful support to peers is difficult for management to observe, measure, or integrate into standard contracts. Consequently, firms incentivize easily quantifiable individual pro-

*This chapter is joint work with Hoa Ho (LMU) and Maren Mickeler (ESSEC). It was pre-registered at the AEA RCT Registry with num. AEARCTR-0012191.

ductivity, leading employees to focus their efforts on these tasks (Holmstrom & Milgrom, 1991; Auriol et al., 2002), even when less quantifiable actions are valuable to the organization. This misalignment creates a tension between individual and organizational goals. Addressing this tension would require firms to measure and incentivize both individual performance and cooperative behaviors.

In this paper, we evaluate a human resources intervention designed to incentivize peer collaboration in a large banking corporation. First, we developed a novel tool to assess the *cooperativeness* of bankers within the organization. Subsequently, in partnership with the bank's top management, we introduced an incentive for cooperation. In particular, top supportive employees received an award acknowledging their *outstanding support to peers*. The award was identified as a suitable incentive, as certificates were considered a valued signal of cooperation—an important skill in the organization and industry. Moreover, monetary bonuses were dismissed due to their non-negligible cost and uncertainty about unintended consequences.[1]

We evaluate the intervention through a pre-registered cluster randomized controlled trial (RCT) conducted across 78 bank branches. To measure employees' *cooperativeness*, we relied on peer voting: each banker identified colleagues in their network from whom they frequently received help and selected the single most supportive person. To mitigate strategic voting, supervisors also assessed their subordinates' cooperation, contributing one-third of the final score. Branches were then randomly assigned to either the treatment or control group. In treated branches, employees ranked in the top 20% of cooperativeness received public recognition during a branch-wide meeting. Control branches continued business as usual without the recognition program. Our RCT design overcomes concerns about omitted variable and selection biases by randomly awarding the most supportive employees in half of the branches, allowing us to isolate the causal impact of the recognition incentive.

The organization involved in this study employs 3,000 staff across various sections of commercial and corporate banking. Our study involved their credit division, which comprises over 1,000 bankers engaged in knowledge-intensive, non-routine lending tasks. Their roles encompass both independent and collaborative activities, with employees frequently needing help from their peers for analyzing complex credit proposals, recovering disbursements from non-compliant borrowers, formulating loan cases

[1]Economic research has shown that monetary incentives can have adverse effects, particularly when used to promote altruistic or prosocial tasks (Alfitian et al., 2023; Ashraf et al., 2020a; Wagner et al., 2020).

for committee adjudication, and appraising large projects. Compensation includes a fixed base salary and a sizable variable incentive tied to individual performance.

Our study evaluates the impact of the intervention on workplace cooperation. Measuring daily cooperative behaviors, however, presents a challenge for researchers, as these interactions occur in person at the branches and vary depending on the task. To address this, in collaboration with the firm's leadership, we identified two systematically measurable and organizationally valuable cooperation metrics: (1) *expertise contributions*—employees' decisions to provide expert feedback in two structured calls for input, and (2) *peer mentoring*—the decision to volunteer for a junior mentoring program. These metrics were collected at two and six months post-intervention. *Expertise contributions* were measured twice, while *peer mentoring* was measured once, both concurrently with a wave of workplace surveys. After completing the survey, employees received an on-screen prompt inviting them to volunteer for one of these cooperative tasks. They saw a brief task description and could choose to participate, with their decision recorded in the system. For one task, we also tracked the time spent and the content of the contributions. Thus, we are able to assess both the decision to help and the quality the help provided. These metrics capture a broad willingness to contribute to the organization's collective good. To complement these measures, we will analyze workplace surveys on perceived help and support within the organization.

We find that the intervention significantly increases employees' willingness to help their peers. In treated branches, employees are 21% (p-value < 0.05) and 25% (p-value < 0.01) more likely to participate in *expertise contributions* at two and six months post-intervention, respectively, and 12% more likely to volunteer as *peer mentors* (p-value < 0.05). Peer ratings strongly predict cooperation—employees with higher peer ratings have higher values of expertise contribution, reinforcing the validity of our proxy measures for helping behavior. The intervention's effects do not differ between awarded and non-awarded employees, suggesting that the incentive promotes helping behaviors even among those who were not directly recognized but worked in a branch where such behaviors were acknowledged.

The incentive does not compromise contribution quality. For the *expertise contribution* task measured two months post-intervention, employees provided strategies for a repository supporting low performers. Evaluating both effort (proxied by time spent) and quality (assessed through treatment-blind expert evaluations), we find that while effort levels are similar across treated and control branches, the novelty and helpfulness of contributions are over 6% higher in treated branches. Lastly, self-reported

survey metrics indicate a significant improvement in perceived assistance in treated branches post-intervention. Employees in these branches report 0.17 standard deviations higher agreement (p-value < 0.05) with the statement that they received assistance from peers in the past month. Thus, the intervention increases cooperative behaviors without compromising the quality of help.

Having established that the intervention increased cooperation on average, we next examine the mechanisms behind this effect. The changes could be driven by shifts in individual motivation, horizontal relationships (among peers), or vertical perceptions (toward management). To explore this, we embedded a variation within the first *expertise contribution* measurement, conducted two months post-intervention. When employees were invited to share expertise, we experimentally manipulated the visibility of their participation along three dimensions. One group was informed that their participation would remain *private*. A second group was told that their involvement would be known to their *peers in the branch*. The third group was informed that *top management* would be aware of their decision to contribute. This variation was randomized at the individual level within both the treatment and control arms. This feature allowed us to assess how visibility influences helping behaviors independently of the incentive and how it interacts with the intervention.[2]

The intervention influences cooperation through two mechanisms: peer reputation and career concerns, while leaving intrinsic motivation unchanged. Under the private condition, 43% of employees contribute in the control group, compared to 44% in the treatment group, suggesting that the intervention does not affect intrinsic motivation (p-value $= 0.82$). However, under peer visibility, contributions in treated branches increase by 15% (p-value < 0.05). Employees may now feel a stronger obligation to help or become more aware that their contributions are noticed and valued by colleagues. Survey data further support a shift in social expectations: six months after the intervention, treated employees report a 0.33 standard deviation increase, relative to the control group, in how difficult they find it to decline requests for assistance from peers (p-value < 0.01). This effect is particularly pronounced among award recipients, who report a 0.6 standard deviation increase. Lastly, when contributions are visible to management, cooperation rises by 30% (p-value < 0.01). Plausibly, the award serves as a clear managerial signal that cooperation is valued. Employees may adjust their behavior to align with organizational expectations, either to enhance career prospects

[2]For details on ethical considerations and the handling of potential deception, see Subsection 1.2.5 on study design, which outlines measures ensuring fairness and transparency.

or because they recognize that helping others is rewarded.

Our article offers two main contributions. First, we provide a proof-of-concept that it *is* possible to incentivize peer support, the sharing of tacit knowledge, and cooperation without relying on team incentives. This is an important contribution, as team incentives may not be feasible in some organizations, or a firm may need to enhance cooperation not only within teams but also across teams. Second, we make a methodological contribution, introducing a simple, effective, and scalable scoring tool for organizations to measure cooperation and implement incentives accordingly. Lastly, this study empirically unpacks the complex interpersonal dynamics of tacit knowledge exchange, showing how both social image toward peers and management-related reputation serve as strong incentives for cooperative behavior.

Our findings have broad applicability. Work settings characterized by complex, interdependent tasks are prevalent not only in corporate environments but also in the public sector and academia (Autor et al., 2003). Our partner organization, a large financial sector firm, shares structural features with commercial banks worldwide, including standardized hiring processes, performance-based incentives, and industry-standard practices for managing productivity and collaboration. These shared characteristics make our results relevant to similar organizations in financial and other knowledge-intensive industries.

Replication is essential for the credibility of the sciences (Camerer et al., 2018) and particularly affects field experiments, where results are often context-dependent and tied to a group of researchers carefully designing and implementing the interventions. To address these concerns, we conducted a pre-registered replication study in a different bank to test the robustness of our findings. This institution, an international corporate bank with headquarters in the United States and 30 branches in Uganda, differed from our original partner in size, culture, and geographic scope. Using the same *Most Supportive Employee* award procedure and outcome metrics, we observed a 13% increase (p-value $= 0.08$) in willingness to help in treated branches, consistent with the main study. Baseline levels of helping behavior and treatment effects were strikingly similar, reinforcing confidence in the internal and external validity of our main results and suggesting the intervention's effectiveness across organizations. Due to the smaller size of the second collaborating partner (30 branches and a total of 200 bankers), we could only pre-register the replication of the main average effects of treatment versus control; heterogeneity analysis and mechanisms were not possible due to insufficient power. While these findings support the intervention's scalability,

further studies in diverse industries and contexts are necessary to fully generalize these results.

Our paper relates closely to two strands of the literature. First, we contribute to the literature on knowledge flows within organizations. Prior theoretical and empirical research demonstrates that efficient information exchange and communication between peers positively affects productivity (Sandvik et al., 2020; Battiston et al., 2021; Guillouet et al., 2024). The literature has highlighted the role of hierarchy as a barrier to employee voice and cooperation in organizations (Auriol et al., 2002; Castro et al., 2022). It has examined barriers from the demand side—why employees may hesitate to seek help—and found that concerns about reputational damage are particularly powerful, creating frictions that discourage help-seeking (Mickeler et al., 2023; Heursen et al., 2024). Our study complements this research by focusing on the supply side—the willingness of employees to *contribute* their knowledge to their *colleagues*.

Second, our work contributes to the literature on awards as influential social signals (Gallus & Frey, 2017; Gallus et al., 2023). Research has extensively discussed the power of different incentives to change behavior, and their unintended effects (Gneezy et al., 2011). We contribute not only by focusing on recognition as a particular type of incentive but also by identifying the specific channels through which it operate in the workplace. Additionally, we contribute to this literature by examining the effect of the incentives on cooperative behavior over time. Previous experimental research has studied cooperation and incentives in one-shot public goods games (Deversi et al., 2020), our study extends this analysis to a longitudinal workplace setting.

In Section 1.2, we detail the experimental design and empirical strategy. In Section 1.3, we present the results, and Section 1.4 discusses internal and external validity. Finally, Section 1.5 concludes.

1.2 Experimental Design

1.2.1 Setting and Sample

The field experiment was conducted with one of the largest commercial banks in East Africa. This firm has its headquarters in Kampala, Uganda. In 2023, the organization employed over 3,000 individuals across various sectors of commercial banking and served over 2.5 million customers, with assets of 6.3 trillion UGX (about 1.6 billion

USD). The bank's core business is divided across two major divisions: banking and credit. Our study involved their credit division, which comprises over 1,000 bankers engaged in knowledge-intensive, non-routine lending tasks.

The organization operates across the country through branches. In the credit section, each branch is structured into three hierarchical levels: branch management, supervisors, and bankers. The bank employs approximately 80 branch managers, 200 supervisors, and 1,000 bankers.

At the top of the hierarchy are the *branch managers*. They are responsible for the strategic direction, performance oversight, and reporting of the branch. Their role includes managing all levels of staff, providing leadership to supervisory personnel, ensuring operational efficiency. Their incentives are tied to the overall performance of the branch. At the next level are the *supervisors*. They are promoted from the ranks of bankers after at least six years of experience. Supervisors usually oversee a team of 5-6 bankers, and their core responsibilities are approving loan appraisals and coordinating their team's work. Their incentives depend on the performance of the bankers that they supervise. Supervisors rotate subordinates every 6-12 months.

At the lowest level of the hierarchy are the *bankers*, who are our study sample. They are the main field staff responsible for all client-facing work. Their tasks include appraising loan applications, presenting loans to the approval committee, monitoring the contractual use of funds, and ensuring repayment throughout the loan period. Approximately 60% of a banker's salary depends on variable performance and is tied to the number of loans disbursed in the current month within certain parameters of net performing ratio and volume of the portfolio. There are two main categories of bankers: *business bankers*, who handle large business loans and have a target of at least four disbursements per month, and *personal bankers*, whose loan targets vary between 20 to 40 loans per month depending on the subcategory. As part of its employee incentive structure, the bank offers a small financial reward each month to the most productive employee in each branch.

Of the 80 branches that the bank has across the country, 78 participated in our study.[3] We include all bankers in these 78 branches in our sample. Branches have on average 13 bankers. As of November of 2023, the bankers have been on average for 5 years at the organization and at the particular branch for 2.4 years, 39% are female. Table 1.1 presents summary statistics of our sample in Column (1).

[3]Two branches were excluded: one is located within a university campus, presenting unique operational characteristics, and the other was established after the study commenced.

Table 1.1: Sample Descriptive Statistics and Balance

	(1)	(2)	(3)	(4)
		Means		Difference
Variable	Full sample	Award Branches	Control Branches	Treat - Control
Share of Women	0.39	0.38	0.41	-0.04
	(0.49)	(0.49)	(0.49)	(0.03)
Tenure in Branch	2.40	2.38	2.44	-0.16
	(2.46)	(2.32)	(2.60)	(0.16)
Tenure in Firm	5.14	5.02	5.33	-0.37
	(3.57)	(3.69)	(3.45)	(0.23)
Survey metrics at baseline (0-100)				
Engagement	75.37	76.49	74.06	2.56*
	(22.66)	(22.81)	(22.66)	(1.43)
Turnover Intentions	25.70	24.99	26.47	-1.55
	(28.01)	(28.95)	(27.23)	(1.77)
Trust in Leadership	68.01	68.90	66.88	2.58
	(25.49)	(26.01)	(25.00)	(1.61)
Feeling Recognized	68.01	68.83	66.93	2.33
	(23.80)	(23.38)	(24.29)	(1.51)
Receive Assistance from Peers	74.12	74.60	73.38	0.78
	(23.31)	(23.43)	(23.36)	(1.48)
Productivity Metrics				
Portfolio Case Load	286.95	283.50	291.36	-10.37
	(306.45)	(296.65)	(315.54)	(19.37)
Number of Disbursements	21.28	22.25	20.57	1.67*
	(14.25)	(14.85)	(13.66)	(0.90)
Non Performing Ratio	0.03	0.03	0.02	0.00
	(0.06)	(0.06)	(0.05)	(0.00)
Observations	1,033	482	551	1,033

Notes: Summary statistics at baseline for the full sample of employees that participated in the workplace climate survey. Columns (1)-(3) report the means and standard deviations (in parentheses). Column (4) reports the difference in means by regressing the variable on treatment assignment and computing the difference along with the associated standard error. The variable *Share of Women* is the ratio of female to male employees. The variables *Tenure in Branch* and *Tenure in Organization* are measured in years. The survey metrics correspond to responses to the baseline survey, ranging from 0 to 100, with 100 representing the highest level of agreement. The *Portfolio Case Load* represents the number of loans handled by bankers as of November 2023. The *Number of Disbursements* refers to the number of loans disbursed in November 2023, and the *Non-Performing Ratio* indicates the share of the portfolio that is in late repayment as of November 2023. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

1.2.2 Evaluation Design

For this study, we collaborated with the bank's personnel department to collect primary data. Before the study began, employees were informed about the research collaboration under the broad theme of *personnel and productivity*. The department routinely gathers employee feedback through surveys, conducted either internally or externally, depending on the topic and area. In this case, employees were informed that survey responses would be collected externally by researchers via Qualtrics and that bank management would not have access to individual-level responses; only aggregated results would be generally shared unless explicitly stated otherwise.

We conducted three survey waves, timed one before and two after the intervention. The Chief Human Resources Officer (CHRO) distributed all surveys via email to employees. Employees could complete them on work computers or personal phones. Each survey took an average of 10 minutes to complete, and participants received 10,000 UGX (approximately USD 2.50) in phone credit as compensation. Figure 1.1 summarizes the timeline and the data collected at each stage.

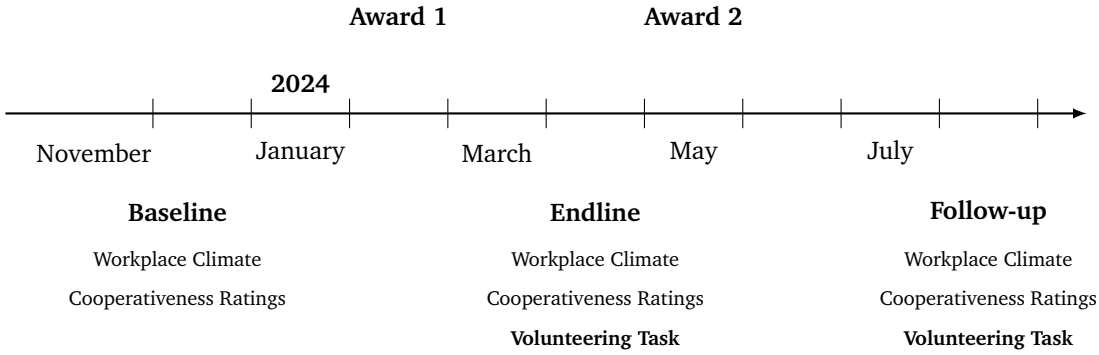


Figure 1.1: Timeline

The baseline survey followed a standard workplace climate design used in human resources, with the addition of the *peer cooperativeness ratings*. Following baseline data collection, we randomly assigned 38 bank branches to the treatment group and 38 to the control group.[4] Table 1.1 shows that we achieved balance in observables.

[4]Given the limited number of branches, we employed a matched-pair cluster randomization approach. Branches were paired based on key characteristics using an optimal greedy algorithm (King et al., 2007; Bruhn & McKenzie, 2009), which minimizes the Mahalanobis distance across pre-selected covariates. Matching relied on administrative and baseline survey data, including branch productivity, branch size, the share of business bankers, gender composition, and employees' perceptions of workplace recognition and peer support. Once pairs of similar branches were formed, one branch in each pair was randomly assigned to the treatment group and the other to the control group.

Column (2) shows average values for the treatment group, Column (3) for the control, and Column (4) tests the differences between both groups.

Between December 2023 and January 2024, branch management in the treatment group was briefed about the awards and given the names of the selected recipients to verify their appropriateness. No manager raised objections, and they were instructed to keep the names confidential until the award ceremony. The certificates were sent from headquarters to the treatment branches at the end of January, and employees received their awards in February 2024. This first round, which constitutes our main intervention, was presented as a surprise; employees were unaware beforehand that an award would be given. Therefore, the intervention was an unannounced, public recognition of employee cooperative behavior. Following the terminology of Bénabou & Tirole (2003, p. 504), unannounced public recognition corresponds to "discretionary" or "ex post" rewards, as opposed to "promised" or "ex ante" contingent incentives (e.g., innovation prizes).

In April 2024, the endline survey took place, followed by a second round of awards in May and a follow-up survey in August 2024. At the end of both the endline and follow-up surveys, employees were presented with *volunteering options* within the organization. These tasks, typically completed separately from personnel surveys, were timed concurrently with the workplace climate survey to allow matching responses to individual participants. Detailed descriptions of these tasks, which serve as primary outcome variables, along with other measures, are provided in Section 1.2.5. Each survey wave achieved a participation rate of 95%.

We pre-registered the first award as our main intervention and the volunteering task in the endline survey as our primary outcome. At the time of pre-registration, we were uncertain whether a second round of awards would be feasible. Additionally, while the first round was unexpected by employees, the second was not. As a result, we did not expect the second round to perfectly replicate the first but rather to be influenced by it. Nevertheless, examining the effects of a second award provides evidence of mid-term impacts and helps assess whether the surprise element drove any observed effects or if they persist over time.

1.2.3 Intervention: Most Supportive Employee Award

In this setting, cooperative behaviors are inherent to the nature of the job and are an integral part of daily work. Employees frequently seek help from peers for tasks such

as analyzing complex credit proposals, recovering disbursements from non-compliant borrowers, formulating loan cases for committee adjudication, or conducting in-person appraisals of large loans. The organization actively encourages collaboration as part of its workplace culture. However, despite this cultural emphasis on support, cooperative behaviors fall outside the firm's incentive scheme. Incentives are tied to individual performance through variable pay structures, and a *Most Productive Employee of the Month* monetary reward.[5] As a result, while cooperation is both common and valuable, it is neither formally tracked nor explicitly incentivized or recognized.

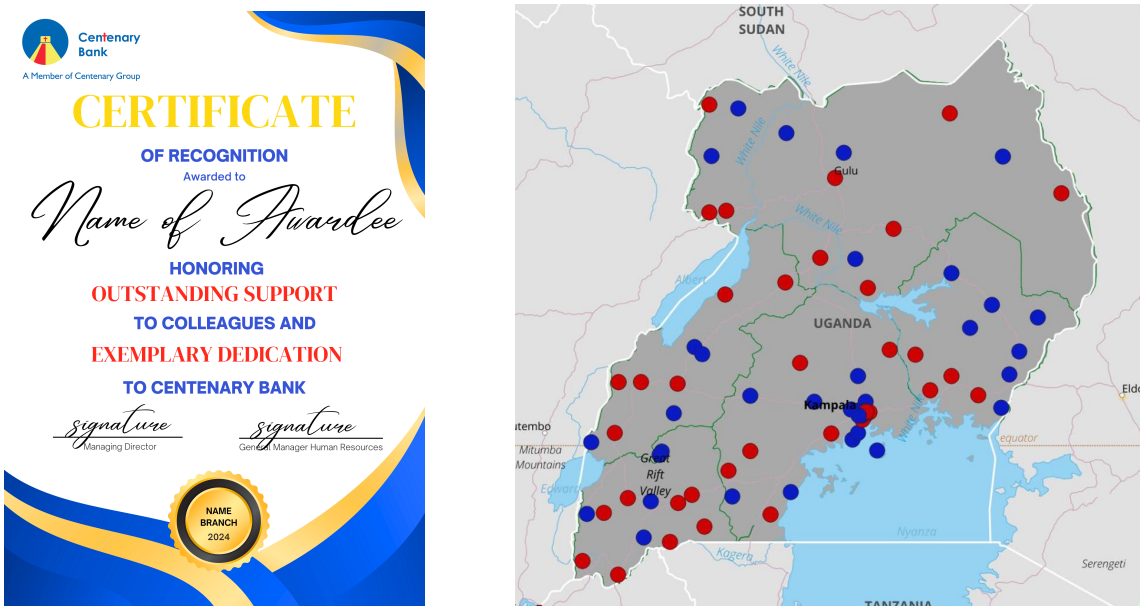
The intervention introduces a direct and formalized incentive for supportive behaviors. To implement this, we first measured employees' cooperativeness using both peer and supervisor ratings. For peer ratings, we used two survey questions inside the HR workplace climate survey. First, we recorded employees' support networks by asking: *From which credit officers do you receive frequent help?* Employees could select multiple colleagues from the full list of bankers in their respective branches. If they selected more than one peer, they were then asked a follow-up question to identify which one stood out as the most supportive. So we will know for each employee in the organization how many times a peer has marked him/her as supportive and as uniquely supportive. To prevent gamification of the metric, supervisors also rated each employee under their direct supervision. They were asked to rank on a scale from 0 (very rarely) to 100 (very frequently) how often their subordinates engaged in supportive activities towards their peers that went *above and beyond*.

The above information was then aggregated, and a score was calculated for each employee. Peer ratings were normalized by dividing the absolute number of times an employee was mentioned as helpful and the most supportive by the total number of colleagues in the branch minus one, as voting for oneself was not allowed. Supervisor ratings were normalized between 0 and 1 by dividing them by 100. A weighted score was then calculated, assigning 2/3 of the weight to peer ratings and 1/3 to supervisor ratings. This resulted in a final rating for each banker in the organization.

In the treatment branches, employees ranked in the top 20% in their branch, based on the aforementioned scoring criteria, were selected to receive the award.[6] The 20%

[5] This reward includes a small monetary bonus of approximately 60,000 UGX (after tax), equivalent to around USD 16.00 at the time of the study, along with public recognition through the display of the employee's name and photo in the branch.

[6] In branches with less than five employees, we selected the employee with the highest score to ensure all treated branches have at least one awarded employee. Only three branches have less than five bankers.



(a) (b)

Figure 1.2: (a) Template of Certificate. (b) Map of Intervention.

Notes: In Figure (b), the geographical location of treated branches is represented by a red circle, while control branches are marked by a blue circle.

threshold was determined in consultation with senior management to balance making the award scarce enough to be meaningful and attainable enough to motivate employees. Additionally, by fixing the proportion of winners at 20%, we ensured consistent intervention intensity across branches, regardless of their size.

Figure 1.2, part (a), shows a template of the certificate that the employees received. The certificates were hand-signed by the Chief Executive Officer of the bank and the Chief Human Resources Officer and framed. They explicitly mention *outstanding support* and *exemplary dedication* to peers and the firm. Part (b) of Figure 1.2 shows in red the geographic locations of the branches where the intervention took place. The awards were presented by branch managers during the February and May monthly meetings, which were attended by all branch employees. During the event, the managers emphasized the importance of cooperation and support and acknowledged how the award recipients had personally contributed to fostering a cooperative culture.

1.2.4 Conceptual Framework and Hypotheses

Our intervention—the introduction of a peer-nominated award—can influence cooperation through multiple channels. First, the certificate may directly affect behavior

due to its intrinsic value or by enhancing intrinsic factors such as employee morale. Second, the award process might shape reputations among peers and influence social expectations around helping behaviors. Moreover, since the awards are signed by top organizational management and presented by branch management, they serve as a visible signal of the importance the organization places on cooperation. This signaling effect may reinforce employees' perceptions of managerial priorities and workplace norms, further shaping their willingness to help.

In this section, we develop a simple conceptual framework that formalizes these mechanisms and derives testable hypotheses to examine how the award influences employees' decisions to help.

Intrinsic Predisposition and the Cost of Helping An employee's decision to help[7] depends on intrinsic factors such as altruism, internalized norms, or personal identity, as well as the cost of providing help (e.g., time or effort). Let α_i represent an individual's intrinsic predisposition to help, e.g., the utility an individual gets from providing help to others. Additionally, let c denote the cost of helping. In the absence of external incentives, the utility derived from helping or cooperating ($h=1$) can be expressed as:

$$u(h=1) = \alpha_i - c \tag{1.1}$$

The presence of the award could increase α_i by enhancing morale, reinforcing self-image, or increasing the perceived importance of helping.

Hypothesis 1 (Intrinsic Predisposition Mechanism): Employees are more likely to cooperate because the intervention increases their intrinsic predisposition to help.

To test this hypothesis, we need to compare helping decisions in treatment and control groups when the decision is observable only to the individual, in the absence of external incentives.

$$P(h=1|\text{Award, Anonymous}) > P(h=1|\text{No Award, Anonymous}) \tag{1.2}$$

would confirm *Hypothesis 1*, where $P(h=1)$ denotes the probability that an individual chooses to help under a given condition.

[7] We use the terms help, cooperation, and support to peers interchangeably. We refer to actions that assist others in improving their productivity, even though these actions do not directly contribute to the productivity of the individual providing the help.

Besides intrinsic motivations, in workplace settings, employees' utility from helping may also be influenced by two important *external factors*: peer reputation and career concerns.

Peer Reputation Individuals care about how they are perceived by others, which can affect cooperation. In workplace interactions, employees may feel social pressure to conform to cooperative norms or seek to signal positive traits, such as being helpful, to their peers. Let R_p represent the social or reputational benefits of being seen as cooperative and adhering to these norms. The utility function can then be expressed as:

$$u(h = 1, \text{Peer Visibility}) = \alpha_i + R_p - c \tag{1.3}$$

where $R_p > 0$ if peer visibility increases the value of helping due to social expectations or reputational benefits. Introducing a public award may influence R_p by amplifying peer reputation concerns, making social expectations more explicit, or increasing the desirability of being seen as helpful.

Hypothesis 2 (Peer Reputation Mechanism): Employees are more likely to cooperate because the intervention increases the reputational value of helping among peers.

If the award heightens peer reputation concerns, we expect the effect of peer visibility on helping behavior to be stronger in award branches than in non-award branches. We test this by comparing helping behavior under peer visibility:

$$P(h = 1|\text{Award, Peer Visibility}) > P(h = 1|\text{No Award, Peer Visibility}) \tag{1.4}$$

To isolate the peer reputation mechanism, we compare the effect of the award under peer visibility (Equation 1.4) to its effect under anonymity (Equation 1.2). If the award affects intrinsic motivation, the incremental effect of peer visibility will be the difference between the two.

Career Concerns Employees may also consider the career implications of their helping behavior. If management values cooperation, employees might help to align with

managerial expectations or to signal attributes that are beneficial for career advancement. Let R_m denote the career concerns effect, which captures the utility derived from being observed as helpful by senior management. The utility function becomes:

$$u(h = 1, \text{Mgmt Visibility}) = \alpha_i + R_m - c \tag{1.5}$$

where $R_m > 0$ if management visibility increases the career-related benefits of helping. Introducing a public award may heighten career concerns by signaling that management explicitly values and rewards cooperative behavior. Employees may thus view helping as a strategic action to improve their career prospects.

Hypothesis 3 (Career Concerns Mechanism): Employees are more likely to cooperate because the intervention increases the career value of helping.

If the award increases career-related incentives for helping, we expect the effect of management visibility on helping behavior to be stronger in award branches than in non-award branches. If this is the case, then:

$$P(h = 1|\text{Award, Mgmt Visibility}) > P(h = 1|\text{No Award, Mgmt Visibility}) \tag{1.6}$$

Similarly, to isolate the career concerns mechanism, we compare the effect of the award under management visibility (Equation 1.6) to its effect under anonymity (Equation 1.2). If the award affects intrinsic motivation, the incremental effect of management visibility will be the difference between the two.

If the decision to help is observed simultaneously by both management and peers, we will not be able to disentangle these two external mechanisms. To do so, we will have to compare helping choices in treatment and control branches where decisions are only visible to peers or only visible to managers.

1.2.5 Outcome Variables

Our primary conceptual outcome is collaboration. In particular, the supply side of help decisions. The ideal measure would be to observe real-time interactions where various opportunities to help arise and track employees' choices. However, this is challenging for researchers, as we cannot continuously observe all employees, and we need a

standardized measure that allows for individual comparisons. Therefore, in consultation with the firm's leadership, we identified three key measures of cooperation that are both valuable to the organization and systematically measurable: two knowledge-sharing options that we will call *expert contribution I* and *II* and one enrollment option to become a *junior mentor*. In order to be able to systematically measure the decisions to help, these metrics were timed to concur with the endline and follow-up surveys so that they could pop up at the end of the surveys and be recorded. These provide us with revealed measures of cooperative behavior. In addition, we will examine self-reported survey measures that directly capture employees' perceptions of help and cooperation within the organization.[8]

Revealed Choices: Knowledge Sharing and Mentoring

When employees complete the workplace climate surveys, an option appears on their screen informing them about a call for volunteers for a specific task. Employees read a brief description of the task and they can decide to sign up or decline. There is one call for volunteers after the endline survey and two at follow-up.

Expertise Contribution I. This is our primary pre-registered outcome, collected two months after the intervention. Employees are informed that the organization is compiling a repository of productivity-enhancing strategies to support bankers across the organization, particularly low performers. Human Resources will compile the contributions and share them by the end of 2024. Employees can briefly read that the task would entail providing successful strategies from professional experience, including general advice they find particularly useful, as well as specific guidance when struggling with borrower repayment. They are informed that this task will take approximately 10 minutes of their time.

Our primary outcome will be the *extensive margin*—whether employees select "yes" or "no" to contribute. If the respondent selects "yes" , then it opens a detailed description of specific current issues that underperformers face and the employee has an open box to introduce their approach to solving them. If an employee selects "no" the answer is recorded in our system and the form closes.

We have access to the individual contributions made and we will use them to examine the *intensive margin*, analyzing both the effort and quality of contributions. To assess

[8]We have also requested access to administrative records. Results from these records will be available 12 months post-intervention.

effort, we will examine the time spent on the contribution and its length (measured in number of characters). To assess quality, we rely on expert evaluations conducted blind to treatment. We engaged two independent evaluators with relevant expertise in banking: one is a senior manager from the bank's head office with over 15 years of experience, and the other is a banking consultant with knowledge of commercial banking in Uganda but external to the bank. These experts assessed each contribution across three dimensions: (1) *helpfulness*—how beneficial the advice would be for other officers in the bank, (2) *innovativeness*—whether the contribution presents unique or creative insights, and (3) *actionability*—whether the advice provides clear and implementable steps. Both evaluator scores will be averaged over each contribution to generate a final quality assessment that is the mean average of both evaluations.

To disentangle the mechanisms laid out in subsection 1.2.4, we introduced a small variation when eliciting this response. It will allows us to decompose the average effect and examine its underlying drivers.

> **Mechanisms:** We implemented individual-level randomization across treatment and control, varying one sentence in the survey that specified the visibility of the decision to help. Employees were informed about the visibility of their involvement based on the following conditions:
>
> *Anonymous*: Only the employee knows that they have contributed ("*Your contributions will be anonymous, and no names will be included in the repository*").
>
> *Peer Visibility*: Employees were informed that their names would be shared with colleagues in their branch ("*Your contributions will be anonymous, and no names will be included in the repository. However, we will circulate a list of contributor names to the bankers in the branch so that the efforts of those who shared can be recognized by their direct peers"*).
>
> *Management Visibility*: In this condition, following the same structure as above, employees were told that their names would be shared with senior management (*"Your contributions will be anonymous, and no names will be included in the repository. However, we will share a list of contributor names with Head Office –HR and Microfinance teams- so that the efforts of those who shared can be recognized by management"*).

From an ethical perspective, while we aimed to examine mechanisms, we also wanted to avoid creating unfair advantages within the organization as a result of the study. For example, if the visibility of contributions were to provide a career benefit due to top management's awareness, employees in the *Anonymous* and *Peer Visibility* conditions would be at a disadvantage, as they would have contributed, but their names would not be known to management. A related and conflicting goal is to avoid deceiving respondents by not following through on promised behaviors. To address these concerns, we follow the strategy of Bursztyn et al. (2023). We inform respondents in the *Peer Visibility* and *Management Visibility* conditions that the list of names will be shared only if we receive at least one contribution from each branch in the bank. In practice, since at least one branch is unlikely to participate in the survey,[9] we ensure that the likelihood of the list being shared is very low. Due to this design, our outcomes should be interpreted as a behavioral choice that is conditional on respondents believing it sufficiently probable that all branches will be included and that at least one response will be received per branch.[10]

Expertise Contribution II. This metric also captures employees' willingness to share their expertise but is measured at a later stage—during the follow-up survey—under a distinct setup and focus. Specifically, employees were invited to volunteer for a consultation with an external expert to provide feedback on challenges in *customer relationship management*. The consultant aimed to gather insights from frontline bankers and propose recommendations to top leadership for improving systems and addressing inefficiencies. As the bank has expanded significantly in recent years, some bankers now manage portfolios of more than 300 customers. To adapt to this growth, the organization is continuously refining its processes. This consultation task serves as a measure of willingness to cooperate by sharing knowledge, as it represents an additional responsibility beyond contractual obligations. Improved processes would not only benefit the organization but also directly support bankers, as their performance depends on the effective management of their portfolios. We observe only the *extrinsic margin decision*—whether employees select "yes" or "no" to sign up for the consultation; the

[9] One branch is being set up during the study period, and its systems are not yet fully operational, making it unlikely to participate. However, the survey email is sent across the entire credit division, so they still received the survey.

[10] It is possible that some respondents may believe that the list is unlikely to be shared. However, for this to bias our estimated treatment effects, it would require not only that this belief differs systematically across treatment conditions but also that respondents who hold this belief are more or less likely to contribute in the visibility arms compared to the non-visibility condition.

actual meeting with the consultant is outside the scope of our research. Bankers were informed that the expected time commitment was 60 minutes.

Volunteering as a Junior Mentor. When a new employee joins a branch, an experienced banker is assigned as a second point of contact alongside the supervisor. The bank aims to pair new staff with experienced bankers during their initial months to provide guidance and support with integration into the workplace and community. Uganda has more than 32 languages and a diverse culture. Bankers must quickly adapt not only to a new job but also to new customs and, in some cases, a different language. Having a peer mentor during these first months can help facilitate this transition. At the end of the follow-up survey, the bankers saw a call for volunteers for this role. While junior mentors do not need to be volunteers—management can directly assign them—having a roster of volunteers who specify their availability and preferences is beneficial for the organization. Again, we observe the decision on the *extrinsic margin* - whether the respondent decided to sign up for the next round of volunteers. Typical official pairings last three months. Whether voluntary or assigned by management, mentoring is not formally monitored and falls outside the contractual obligations of the banker. However, once assigned, it is expected that the mentor will fulfill this role. It falls within the tasks of the branch management to take care of the final matching of mentors to juniors.

These two measures of cooperative behavior at endline do not allow for an examination of mechanisms, as they inherently involve some degree of visibility—either to management or peers. This reflects the common dynamics in organizations, where volunteering tasks, whether promotable or not, are typically observed by at least some colleagues. However, the endline measure, *Expertise Contribution I*, provides a unique opportunity to disentangle mechanisms. Since this metric captures voluntary knowledge sharing and can be directly entered into the system under an anonymous option, it allows us to isolate the effects of the incentive independently of the visibility component, which might otherwise influence behavior, especially once we add the award for cooperation.

Workplace Personnel Survey: Additional Measurement

The measures described in the previous section capture employees' decisions to offer help, and while it is logical to expect that an increase in help given would result in more help received, they do not provide a direct measure of whether bankers actu-

ally *received* more assistance. To assess whether the intervention led to more received assistance, we leverage self-reported metrics from the workplace climate survey as suggestive evidence. This survey covered various HR-related topics such as leadership quality, role clarity, and organizational transparency, which, while relevant, were not the primary focus of this study. However, it included a key question on peer assistance: *"In the last month, other bankers in my branch assisted me with challenges, even if it was not part of their job description"*. This measure allows us to test whether the intervention increased reported help received from colleagues over time, as the question was included in baseline, endline, and follow-up surveys.

Additionally, the workplace climate survey conducted six months post-intervention included measures to assess the mechanisms discussed in Section 1.2.4. One question captures peer pressure to provide help, evaluating whether the award influenced perceived expectations to assist others. Respondents rated their agreement with: *"I find it difficult to decline a request for help even when I have other tasks that should take priority"*. Comparing responses between treatment and control groups will provide evidence on whether the introduction of the award increased perceived pressure to help.

To assess whether employees viewed the intervention favorably—an important consideration for organizations introducing new programs—we included a set of perception measures in treated branches. One such measure captures beliefs about career benefits: *"Receiving the award is beneficial for career advancement"*. Since this question was asked only in treatment branches, we cannot compare it to a control group, but it provides insight into whether employees perceive the intervention as linked to career progression. Additional perception questions evaluate the broader organizational implications of the award.

Robustness Checks: Spillovers

Since treated and control branches are located in close proximity in some regions, particularly in Kampala, and employees are occasionally transferred between branches, there is a possibility that those in control branches were exposed to information about the intervention through informal communication, inter-branch interactions, or staff transfers. To assess potential spillover effects, we included a question in the endline and follow-up survey asking employees whether they had observed initiatives in other branches that they would like to adopt in their own workplace. Specifically, respon-

dents were asked: *"Are there initiatives in other branches that you would like to adopt in your own branch? If so, which ones?"*.

We will analyze the open-ended responses individually, searching for references to the intervention, particularly mentions of *awards* or *recognition*. If employees in control branches report awareness of the award, we will conduct robustness checks by excluding pairs of branches where spillovers are detected to ensure that our estimates are not biased by cross-branch contamination.

1.3 Results

1.3.1 Empirical Specification

To test the null hypothesis that the intervention has no effect on cooperation, we estimate the average treatment effect using the following specification of an Ordinary Least Squares regression:

$$y_{ib} = \beta_0 + \beta_1 T_b + \varepsilon_{ib}, \tag{1.7}$$

where y_{ib} represents the cooperation decision of employee i in branch b, taking a value of 1 if the employee cooperates and 0 otherwise. The variable T_b is a dummy equal to 1 if branch b is in the treatment group and 0 otherwise. ε_{ib} is the error term. β_1 captures the causal effect of the intervention on cooperation, representing the difference between employees in treated and control branches. Standard errors are clustered at the branch level.

For the intensive margin, the regression above will compare only those who have decided to contribute, so the dependent variable y_{ib} will represent the quality of the contribution, conditional on having chosen to contribute. Since baseline values for most primary variables are not available, we rely on a simple treatment-versus-control comparison.

1.3.2 Treatment Effect on Cooperation

Sign-ups to Help Table 1.2 presents the treatment effects on the probability of signing up for cooperative tasks across different volunteer options. Column (1) reports

the pre-registered primary outcome: *Expertise Contribution I*, measured two months after the intervention. In the control group, 42.8% of employees contributed. This rate increased by 8.9 percentage points (a 20.7% increase) in treated branches, indicating that the intervention significantly enhances employees' willingness to engage in cooperative tasks (p-value < 0.05).

Columns (3) and (5) present results from cooperation measures collected six months after the intervention. The treatment effects remain consistent across these metrics. Sign-ups for the second expertise contribution are 24.8% higher (p-value < 0.01) in the treatment group compared to the control, while sign-ups for peer mentoring are 12.3% higher (p-value < 0.05). These findings suggest that the intervention induces a sustained increase in cooperation over time, corroborated by different measures across different periods.

Table 1.2: Effect of the Intervention on Volunteer Sign-Ups

	(1)	(2)	(3)	(4)	(5)	(6)
	Endline		*Follow-up*			
	Expertise Contribution I		*Expertise Contribution II*		*Volunteer Mentor*	
Award Branch	0.089**	0.096**	0.118***	0.118***	0.066**	0.066*
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Most Supportive R1		0.113**				
		(0.06)				
Award Branch × Most Supportive R1		-0.044				
		(0.07)				
Most Supportive R2				0.044		0.094
				(0.05)		(0.06)
Award Branch × Most Supportive R2				0.001		0.002
				(0.08)		(0.08)
Mean Dep. Var	.428		.475		.534	
Observations	908	908	908	908	908	908

Notes: The table presents linear probability model (OLS) estimates of the treatment effect on sign-ups for different cooperation tasks: willingness to provide expertise to the repository (Columns (1) and (2)), willingness to meet with a consultant to improve customer management processes (Columns (3) and (4)), and willingness to sign up as a junior mentor (Columns (5) and (6)). The dependent variable is binary, coded as 1 if the person chose to sign up and 0 otherwise. Columns (2), (4), and (6) include an interaction with being selected as the most supportive employee in the previous survey. If a person was selected and was in the award branch, they received the award. Standard errors are clustered at the branch level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Interestingly, the treatment effect is not driven exclusively by employees who received the award. Column (2) introduces an interaction term between treatment assignment and being selected as the "Most Supportive Employee". Employees identified as the

most supportive in the control group serve as a counterfactual for those in the treatment group who received the award. Across Columns (2), (4), and (6), the coefficient on treatment assignment remains large and significant, while the interaction term is consistently insignificant and close to zero. This indicates that receiving the award is not the driver of the treatment effect. Instead, the intervention also appears to motivate the broader employee base—those who participated in the voting process and observed the award—leading to increased willingness to help among non-recipients in the treatment group.

Additionally, Column (2) reveals that employees selected as the most supportive are more likely to sign up to contribute, with a large and significant coefficient (p-value $<$ 0.05). This suggests that the voting score effectively identifies employees predisposed to helping. In additional analyses, we further test the correlations between supervisor ratings and peer ratings with the likelihood of contributing in the control group (see Appendix Table A4).[11] We observe that peer ratings are a strong predictor of singing up to the cooperation task. Employees with higher peer ratings at baseline showed a significantly stronger likelihood of signing up to contribute (p-value < 0.01). Specifically, a 10% increase in the number of peers selecting an individual as frequently providing support raised that individual's likelihood of volunteering to share knowledge by 2%. In contrast, supervisor rankings showed no correlation with actual helping behavior in this task (p-value $= 0.71$).[12]

The results in Table 1.2 are based on the most restrictive sample: employees who participated in both the endline and follow-up surveys and were not transferred between treatment and control branches [13] and without using any controls. The effects and coefficients remain unchanged when including pair-matched fixed effects (Table A1), incorporating the full sample (Table A2), and under alternative specifications such as logit estimation (Table A3). Furthermore, the results are robust to replication in a different organization (see Section 1.4).

Effort and Quality of Contributions We examine whether the observed increase in contributions affects the effort and quality of the knowledge shared. A potential

[11] We check only for the control group as their behavior was not affected by the intervention.

[12] We observe similar trends in the subsequent survey wave. Supervisor ratings remain a poor predictor of signing up to volunteer as a mentor (p-value $= 0.62$). Peer ratings continue to perform better, with coefficients of similar magnitude to those at baseline, though they are not significant (p-value $= 0.15$)

[13] Employees are typically transferred after three years in a branch; approximately 6% of the sample was transferred during the study period.

concern is that while more employees opt to contribute, the additional contributions might be of lower quality if the new contributors possess less expertise. Table 1.3 presents the results. Columns (1) and (2) examine effort, measured by the length of contributions (in characters) and the time spent composing them (in seconds). The estimates indicate no significant difference in contribution length between treatment and control groups. However, treated contributors spent 25% more time composing their responses, corresponding to an increase of 145.7 seconds (2.4 minutes). This effect is not statistically significant (p-value = 0.13). Note that due to a coding error, contribution times were recorded only from the second day of data collection, leading to a reduced sample size in Column (2).

Table 1.3: Intervention Impact on Effort and Quality of Contributions

	(1)	(2)	(3)	(4)	(5)
	Effort		*Expert Evaluation of Quality*		
	Length	*Time*	*Helpful*	*Actionable*	*Innovative*
Award Branch	22.314	145.726	0.149*	0.100	0.087**
	(45.89)	(99.51)	(0.08)	(0.07)	(0.04)
Mean Dep. Var	413	586	2.62	2.31	1.61
Observations	407	277	407	407	407

Notes: The table presents OLS estimates of the effect of treatment assignment on different intrinsic margin measures. Columns (1) and (2) show the effect of treatment on response length (measured in characters) and response time (measured in seconds). Column (2) has a reduced number of observations as the timer only started counting on the second day of data collection. Columns (3)–(5) present results for quality measures, evaluated by expert evaluators blind to treatment. Quality scores range from 0 to 6, with 6 indicating the highest quality. All expert-evaluated scores are averaged across evaluators. Standard errors are clustered at the branch level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Columns (3)–(5) assess contribution quality based on expert evaluations blind to treatment. The quality measures range from 0 to 6, with higher scores indicating better quality. The coefficients across all quality dimensions—*helpfulness*, *actionability*, and *innovativeness*—are positive, though not always statistically significant. Notably, the intervention significantly increased the innovativeness of contributions (Column (5)), with treated contributors scoring 5.4% higher than their control counterparts.

Overall, these findings suggest that the increased participation in knowledge sharing does not come at the expense of quality. If anything, contributions in the treatment group tend to be more innovative, implying that expanding participation can enhance the variety and originality of shared knowledge without diluting its value. However, we do not have intrinsic margin measures at follow-up to examine whether these effects persist over time.

Self-Reported Help We now examine whether employees report receiving more help as a result of the intervention. Figure 1.3 presents the estimated impact on self-reported perceptions of receiving assistance from peers. Compared to the control group, employees in treated branches reported a 0.14 standard deviation statistically significant increase (p-value < 0.10) in agreement with having received help from colleagues in the previous month at endline (coefficient from Award Branch X Endline). This effect further moves to 0.17 standard deviations (p-value < 0.05) six months after the intervention (coefficient from Award Branch X Follow-up). These results prove that the intervention not only encouraged employees to offer assistance but also translated into an increase in the self-reported reception of peer support over time.

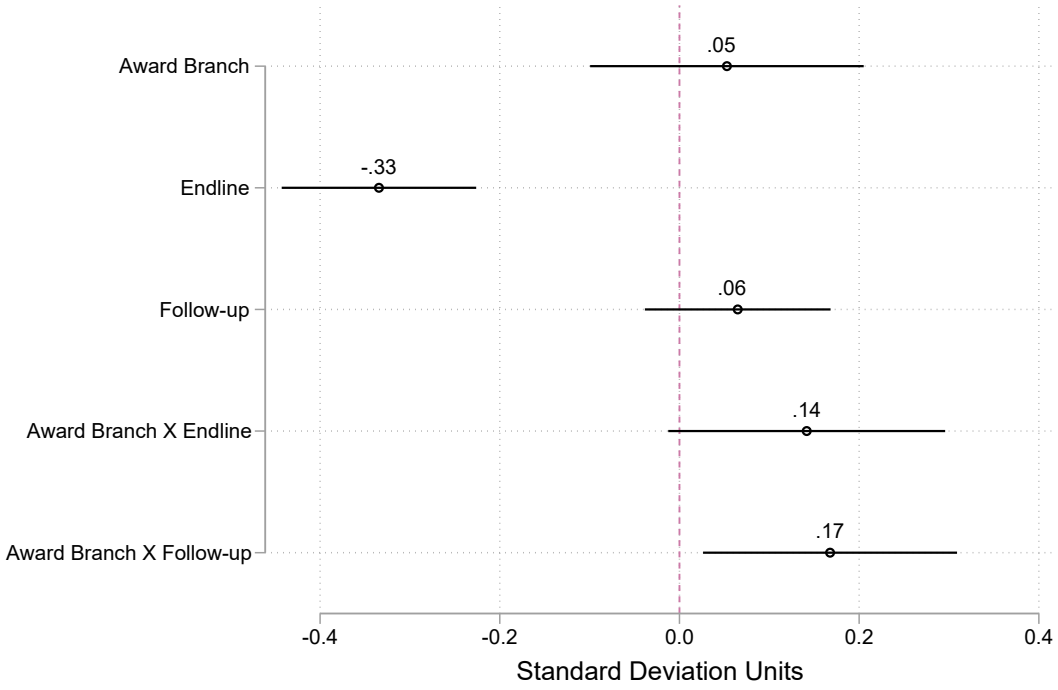


Figure 1.3: Impact of the Intervention on Perceived Frequency of Peer Help

Notes: This figure presents coefficient estimates from a linear regression analyzing the impact of being in a treated branch compared to a control branch on self-reported perceptions of receiving peer assistance over time. Employees rated their level of agreement on a scale from 0 to 100 with the statement: *"In the last month, other bankers in my branch assisted me with challenges, even if it was not part of their job description"*. The dependent variable has been standardized (demeaned and divided by its standard deviation), so coefficients are interpreted in standard deviation units. Confidence intervals are at the 95% level. Standard errors are clustered at the branch level.

These results complement the primary findings and provide suggestive evidence that the intervention is associated with an increase in reported peer assistance. While the self-reported nature of the measure introduces potential reporting biases, the consistency of the effects over time suggests a sustained perception of greater support

among employees in treated branches. Although there are several reasons why employees might indicate that they "provide more help" after the introduction of the award, self-reports of having *received* more help are likely subject to less bias. This is because individuals have fewer incentives to overstate the support they receive from colleagues, as doing so does not enhance their self-image. If anything, research suggests that individuals have reputational incentives to avoid being seen as someone who needs help (Heursen et al., 2024). This finding aligns with the observed increase in help offered and reinforces the idea that recognition-based incentives may shape workplace cooperation.

1.3.3 Mechanisms: What Drives Employees to Cooperate More?

Having established that the intervention increased cooperation on average, we now examine the underlying mechanisms driving this effect. Does the intervention alter individual predisposition to help, does it change the pressure or expectations to cooperate among peers, or does it signal from management that cooperation is valued, leading employees to behave strategically for career gains? Figure 1.4 disaggregates the decision to help in the *Expertise Contribution I* under different levels of visibility.

The first hypothesis is that the award could alter *individual predisposition* to help. To test this, we compare the percentage of employees who choose to help in the treatment branches to those in the control branches when their decision remains private. As shown in the leftmost bars of Figure 1.4, there is no statistically significant difference between the treatment and control groups under anonymity (p-value $= 0.82$). Around 43% of employees contribute in the control group, compared to 44% in the treatment group. Therefore, we reject the hypothesis that $P(h = 1|\text{Award, Anonymous}) > P(h = 1|\text{No Award, Anonymous})$ and conclude that the intervention does not affect intrinsic motivation or purely altruistic behavior. If an organization seeks to increase support for public goods without additional incentives, the award alone may be insufficient.

Second, we examine whether the intervention influences *peer reputation*. If this mechanism is at play, contributions under peer visibility should differ between treatment and control groups. As shown in the two bars in the center of Figure 1.4, contributions in treated branches increase by 15% (p-value < 0.05) under peer visibility, a statistically significant difference. This confirms our second hypothesis that peer reputation drives cooperation ($P(h = 1|\text{Award, Peer Visibility}) > P(h = 1|\text{No Award, Peer Visibility})$). While we cannot pinpoint which specific dimension of

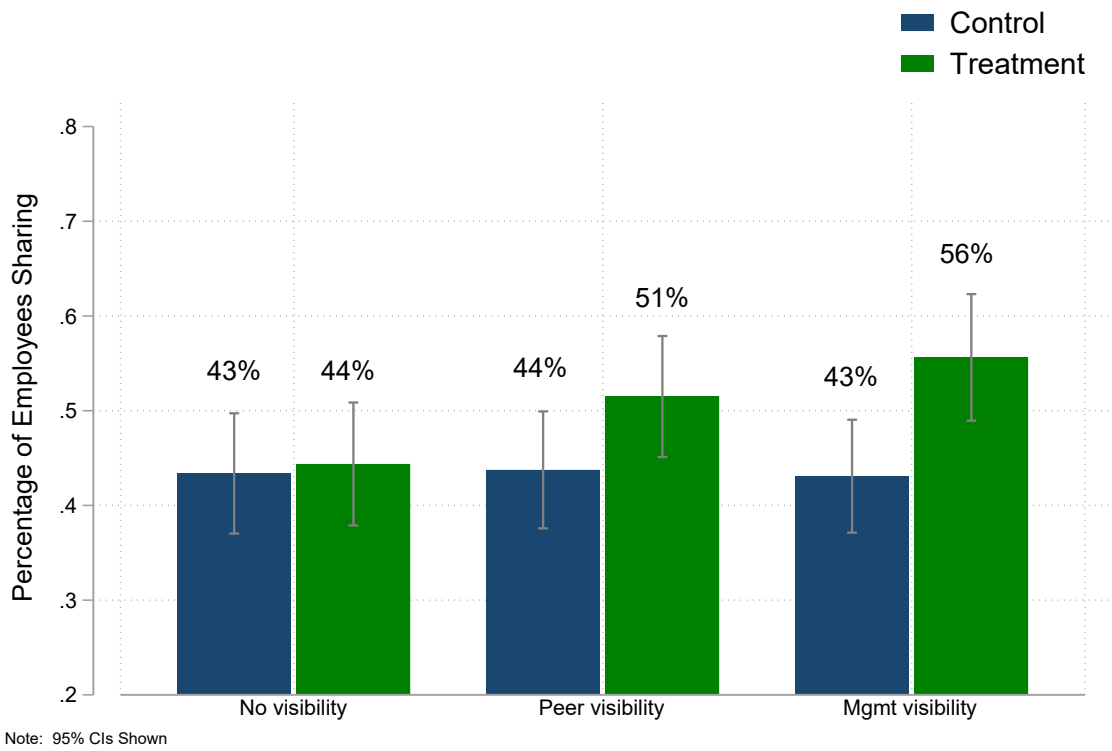


Figure 1.4: Sign-ups to Expertise Contribution I under Visibility Conditions

Notes: This figure presents the percentage of employees who signed up to contribute to the *Expertise Contribution I* task, disaggregated by visibility condition. The leftmost bars represent the condition where contributions were only known to the individual. The two bars in the middle correspond to conditions where contributions might be known to peers, while the two bars on the right correspond to conditions where contributions might be visible to management. The green bars indicate treatment branches where an award was introduced, while the control branches are shown in blue.

professional reputation is at play, this effect could reflect employees feeling more obligated to help or becoming more aware that their peers notice and value this behavior. Self-reported survey data further suggest that expectations around helping have shifted.

Figure 1.5 presents the comparison between treatment and control groups regarding the statement: *"I find it difficult to decline a request for help even when I have other tasks that should take priority"*. Six months after the intervention, employees in the treatment group reported a 0.33 standard deviation increase in difficulty declining help requests compared to the control group. This effect is even stronger for award recipients, who report a 0.64 standard deviation increase relative to their counterfactuals—employees in the control group identified as the most supportive.[14]

Overall, these results suggest that the intervention directly influences peer interactions

[14]The wider confidence intervals for this group reflect the smaller sample size, as it comprises only the top 20% of supportive employees.

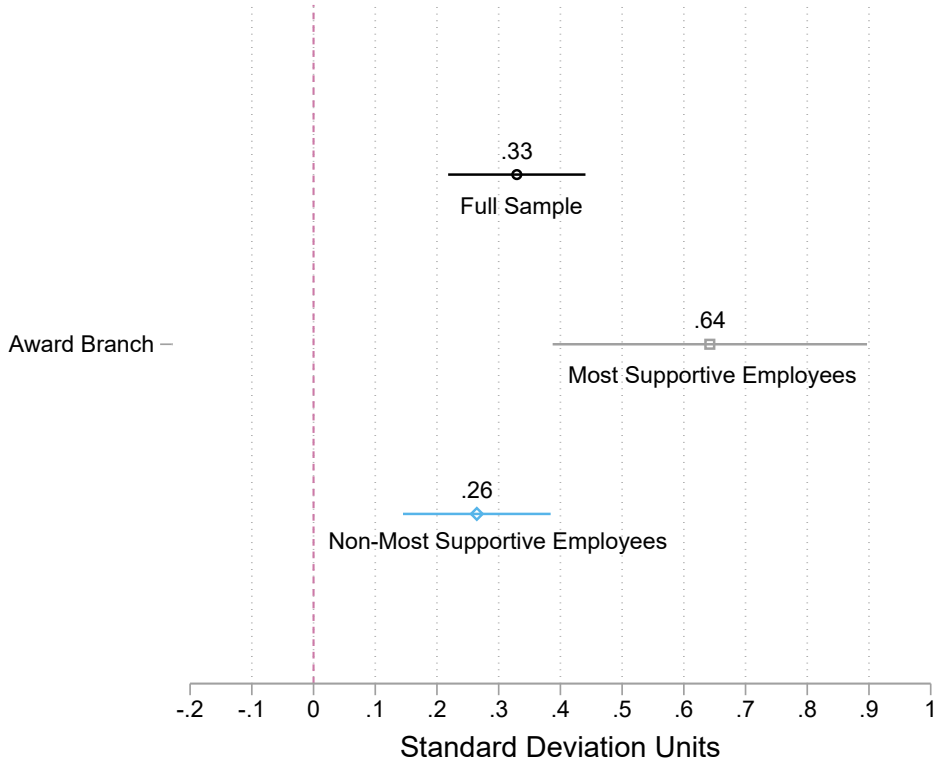


Figure 1.5: Impact of the Intervention on Perceived Difficulty of Declining Help

Notes: This figure presents coefficient estimates from linear regressions analyzing the impact of the intervention on employees' self-reported difficulty in declining help requests. The dependent variable is based on responses to the statement: *"I find it difficult to decline a request for help even when I have other tasks that should take priority"*. The analysis compares treated branches to control branches, with results reported for the full sample as well as separately for *Most Supportive Employees* (top 20% in supportiveness) and *Non-Most Supportive Employees* (remaining 80%). The dependent variable has been standardized (demeaned and divided by its standard deviation), so coefficients are interpreted in standard deviation units. Confidence intervals are at the 95% level, and standard errors are clustered at the branch level.

by shifting norms around cooperation. However, this is not the only way the intervention could affect cooperation through this channel. Anecdotal evidence from post-study qualitative interviews with awardees suggests that the award increases the attractiveness and awareness of being seen as a supportive colleague. Helping behaviors in professional settings often consist of minor yet meaningful gestures—answering a colleague's question, sharing advice, or offering quick support—that accumulate over time. Without formal recognition, these acts may lack visibility, leading employees to underestimate their significance or doubt whether peers notice and value them. The award may serve as a signal that such contributions are observed and appreciated, reinforcing the perception of who consistently engages in supportive behavior. For example, one awardee remarked:

"I was very surprised when I received the award. I've always helped others

> *because that's how I was raised, and I know I'm a supportive person. What surprised me most was learning that so many people in the branch noticed my efforts and saw me as someone who stands out in supporting others."*

Lastly, an alternative explanation is that overall cooperation did not increase, but rather that employees strategically participated in this particular task to signal their helpfulness and improve their chances of winning the award in the next round. Under this scenario, the observed increase in contributions would not reflect a genuine shift in cooperative behavior within the organization but rather a temporary response driven by award incentives. However, strategic voting is unlikely to account for this effect, as the call for contributors occurred after the voting for the next round of awards had already taken place, and the names of those who contributed were scheduled to be shared nine months later. While we cannot completely rule out this channel, we consider it the least plausible explanation.

The third and final hypothesis is that the intervention affects behavior by influencing *career concerns*. By introducing the award, management sends a clear and direct signal that helping others is valued. This may motivate employees to align their behavior with organizational expectations, either to enhance their career prospects or simply because they have become more aware that such behavior is rewarded. When comparing the two rightmost barss in Figure 1.4, we observe a 30% increase in cooperation under management visibility (p-value < 0.01). This finding supports the hypothesis that career concerns play a role in the observed behavioral change and therefore confirms the third hypothesis that $P(h = 1|\text{Award, Mgmt Visibility}) > P(h = 1|\text{No Award, Mgmt Visibility})$. Figure 1.6 further examines whether employees perceive career-related benefits from receiving the certificate. Two and six months after the intervention, employees in treated branches were asked about their perceptions of the award. Figure 1.6 presents results at two months (Appendix Figure A1 shows similar results at six months). On a scale from 0 to 100, where 100 represents the highest level of agreement, the average agreement level was 82 for the statement that winning the award has positive career implications. If employees believe that a behavior is valued by management, it is plausible that they will seek to comply with it.

1.3.4 Employee Perceptions of the Intervention

To assess the award's value from the organization's perspective, it is worthwhile to consider employees' direct opinions about it. Understanding how the award is per-

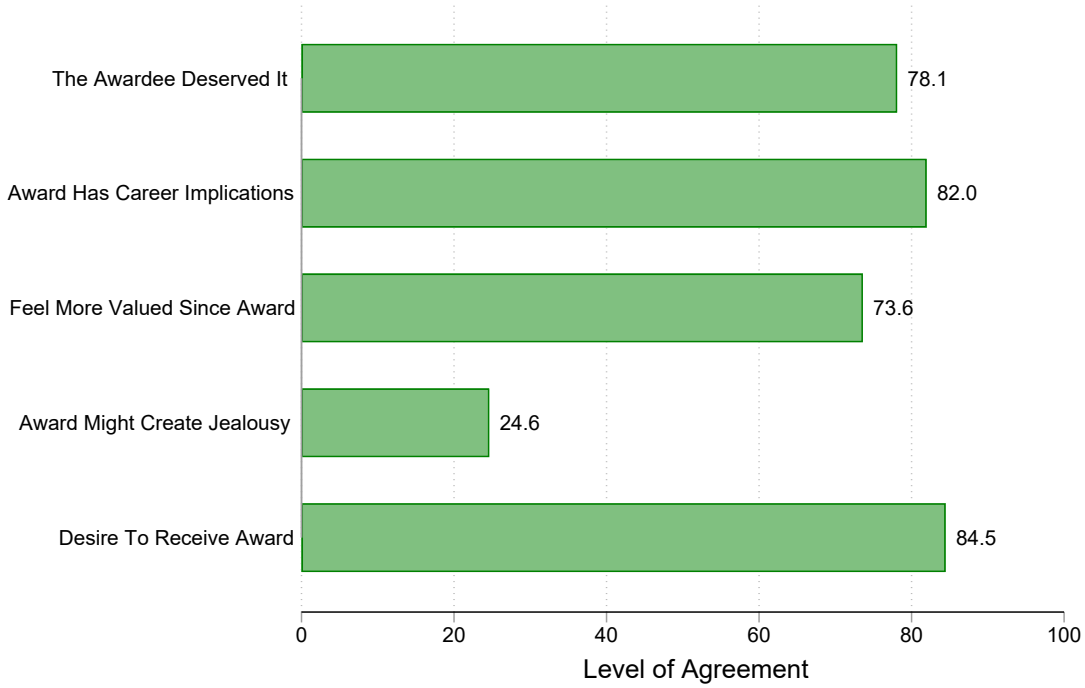


Figure 1.6: Perceptions of the Intervention

Notes: This figure presents employees' perceptions of the award six months after the intervention. Employees rated their agreement with various statements on a scale from 0 to 100, where 100 represents the highest level of agreement.

ceived can provide insight into whether it serves as a meaningful incentive, fosters positive workplace dynamics, or creates unintended negative consequences, such as resentment or competition.

To evaluate this, Figure 1.6 presents the level of agreement with several statements about the award, where employees rated their agreement on a scale from 0 (completely disagree) to 100 (completely agree). Employees in treatment branches overwhelmingly report positive perceptions of the award. Specifically, employees express an average agreement level of 84.5 with the statement that they would like to receive the award at some point, suggesting that the incentive is seen as highly desirable. Additionally, employees report an average agreement level of 73.6 with the statement that the award makes them feel more valued in their role, reinforcing the idea that public recognition enhances employees' sense of appreciation and organizational belonging. The award is also perceived as having professional value: employees report an agreement level of 82 when asked whether the award has career implications. This suggests that employees believe the recognition carries weight beyond symbolic acknowledgment and may influence professional development or advancement oppor-

tunities. Furthermore, employees reported an average agreement level of 78.1 with the statement that the awardee truly deserved the recognition, suggesting that the selection process was perceived as fair and legitimate.

A potential concern with workplace awards is that they might generate competition, resentment, or jealousy among employees. However, the average level of agreement with the statement that the award creates jealousy within the branch is only 24.6 out of 100. This suggests the award does not lead to negative interpersonal dynamics or workplace conflict. Instead, the responses imply that employees see the award as a constructive and meritocratic mechanism rather than a divisive or exclusionary one.

Taken together, these findings suggest that the award is broadly well-received and is viewed as both a desirable and meaningful form of recognition. Employees believe that it increases their sense of being valued at work, has professional relevance, and is awarded fairly, while concerns about jealousy or resentment appear to be minor.

1.4 External Validity and Replication

To assess the robustness of our findings, it is important to consider potential limitations arising from both our study's internal and external validity.

From an internal validity perspective, a concern when designing the study was potential spillovers between branches. Given the proximity of some treated and control branches, particularly in Kampala, and occasional staff transfers, employees in control branches may have been indirectly exposed to the intervention through informal communication or inter-branch interactions.

To assess potential spillover effects, we included a question in the endline and follow-up surveys asking employees whether they had observed initiatives in other branches that they would like to adopt in their own workplace. Analyzing over 200 open-ended responses from control branches, we found no direct references to the intervention. This is likely due to the broader organizational context in which the study took place. At the time of the intervention, the company was undergoing significant growth and cultural transformation, with multiple concurrent initiatives aimed at reducing hierarchical distance and improving communication. Consequently, the award was perceived as one of many human resource initiatives rather than a distinct program, making it less likely to be singled out and mentioned to other branches.

The most frequently cited initiatives employees wished to adopt were unrelated to

the intervention. Employees often referenced activities available in a previous branch but no longer present after transfer, such as team-building activities and management strategies like group recoveries. Given that the intervention lasted from February to August, the extent of inter-branch interactions and spillovers appears insufficient to compromise our results. Overall, these findings suggest that the stable unit treatment value assumption was preserved, and spillover effects are unlikely to threaten internal validity.

From the point of view of external validity, an active debate has arisen around the credibility and replicability of empirical findings in various scientific fields (Camerer et al., 2018; Baker, 2016), including psychology, management, and the biological sciences. Concerns that numerous published results may constitute false positives have led some scholars to describe the situation as a "credibility crisis" for modern science. Replication allows researchers to help in mitigating these issues. When multiple studies consistently corroborate an original finding—even under different contexts or with new samples—the claim that the result holds in a broader set of conditions is substantially strengthened.

In organizational economics and management, field experiments are typically conducted in partnership with specific organizations (Levitt & List, 2009; Chatterji et al., 2016). As a result, questions often arise about the degree to which their findings can be generalized—or "externally validated"—to other contexts. Even well-powered field experiments with robust designs can face challenges when tested elsewhere, as organizational culture, size, geographic location, and workforce characteristics could all shape outcomes differently. Replication, therefore, is essential not only for academic rigor but also to help decision-makers in firms determine whether a given research result is replicable in their specific setting.

In light of the concerns about generalizability, we conducted a replication study[15] in a second financial institution to determine whether our main results would hold under different conditions. We deliberately chose an organization in the same industry—a bank employing knowledge workers—but with distinct characteristics from our original partner. Specifically, this second entity is an international corporate group headquartered in the United States rather than a local bank, operating at roughly one-tenth the size (30 branches) of the initial institution. Although employees' daily

[15]The pre-registration for this study can be found in AsPredicted #153656. The study was pre-registered separately from the main study to ensure they were treated as two distinct studies. The replication, as detailed in the pre-registration, was limited to only one treatment condition—awarded versus non-awarded branches—due to power constraints.

tasks—assessing borrowers and managing funding—are broadly similar, these tasks are allocated differently among staff, and while incentives remain largely individualized, peer support continues to play a significant role. Consequently, this smaller corporate environment provides a meaningful test of whether the results observed at the larger, more centrally managed bank can be replicated in an alternative institutional setting.

The replication followed the same "Most Supportive Employee" award procedure as in our main study. Specifically, among the 30 branches, half were randomly assigned to receive the award treatment using pair-matching randomization, while the others remained in the control group. Because each branch had only around five employees, one individual per branch was recognized as the top supporter in both rounds of awards, which took place during the first two quarters of 2024—mirroring the timeline in our primary study. As in the main experiment, employees answered a general workplace climate survey and identified their "Most Supportive Peers". After the second round of awards, they were asked to share their knowledge with their colleagues [16] Employees were informed that both their peers and the head office would be aware of their contributions. However, due to logistical constraints in this smaller organization, the key questions related to our mechanisms were only asked at the second round. Of the 215 officers invited, 204 completed the survey.

Table 1.4 presents the primary results of the replication study. Despite the smaller scale of the replication, the findings are consistent with those from the main study, reinforcing the robustness of our original results. Specifically, we observe a 5.7 percentage point increase in employees' willingness to invest additional time in helping colleagues in treated branches compared to the control group. This effect is statistically significant at the 5% level. Standard errors are clustered at the branch level, and pair-matched fixed effects are included to address the small cluster size.

Interestingly, the baseline level of willingness to help in the control group, measured as the proportion of respondents signing up to contribute was 43%. This pre-existing level mirrors the control group result in the original, larger-scale study, suggesting a

[16] Note that we administered the knowledge-sharing question after the second round of awards—rather than the first, as in the other organization, due to differences in survey timing across the two organizations, which were dictated by internal departmental schedules. In the smaller organization, the survey was conducted two weeks earlier. Still, the final phrasing of this key outcome metric had not yet been approved by top management in the original organization. At that time, researchers only had a draft version of the phrasing. Since the core goal of the replication was to re-run the exact same metrics with minor contextual variations, we decided to include the knowledge-sharing question after the two award runs.

striking consistency across organizational contexts. The magnitude of the treatment effect in the replication is also comparable to that documented in the main study.

Table 1.4: Willingness to Share - Replication

	(1)	(2)
	Willingness to Share Knowledge	
Award Branch	.057**	.111***
	(.03)	(.04)
Most Supportive		.215
		(.19)
Award Branch × Most Supportive		-.474*
		(.24)
Mean Dep. Var	.433	.405
Observations	204	204

Notes: Linear probability model (OLS) estimates the effect of the treatment on the willingness to provide productivity tips. Column (1) includes the main average effect and Column (2) adds the interaction with being selected in the previous round as most supportive. Regressions control for demographics, including tenure at the branch and firm, as well as gender. Both models include pair-matched fixed effects. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In Column (2), we observe that individuals previously selected as "Most Supportive" exhibit higher levels of helping behavior, although the coefficient is statistically insignificant. The positive and substantial magnitude of the coefficient suggests that the nomination metric effectively captures a propensity for helping behavior. Furthermore, the treatment effect remains significant and increases in magnitude when an interaction term for receiving an award is included, consistent with the main study. The interaction term is slightly negative, indicating that the observed changes in the awarded branches are not driven by the recipients of the award. Instead, the results suggest that the introduction of the award within the branch fosters a broader behavioral shift among employees, underscoring the treatment's overall positive effect.[17]

These replication findings offer two important lessons. First, for organizations, the results suggest that interventions introducing publicly recognized awards for supportive behaviors can reliably produce an effect size of approximately 13% in employees' willingness to help. This enables organizations to assess whether this effect size justifies the intervention's costs. The consistency of the results, even in an institution with a different culture and size, underscores the robustness of the intervention's impact and its relevance for companies in knowledge-intensive sectors.

[17] Due to the smaller sample size of 200 employees, the study lacks sufficient power to replicate and examine any underlying mechanisms fully.

Second, for academics, the replication strengthens confidence in the main effect of the intervention. Replication serves as a powerful tool to reduce the risk of false positives, providing greater reliability in results. However, further testing in organizations from different industries or countries is needed to understand how the intervention might perform in varying contexts. Naturally, research findings are always context-specific and depend on the unique characteristics of the setting.

1.5 Conclusion and Implications for Organizations

This study provides causal evidence on how social recognition incentivizes cooperation and tacit knowledge exchange in organizations. We show that a simple, low-cost intervention—a public award for peer support—significantly increases employees' willingness to help colleagues. The recognition incentive led to a sustained rise in voluntary expertise contributions and peer mentoring, while also improving the quality of help provided. Moreover, our findings suggest that workplace cooperation was driven by peer reputation effects and managerial signaling rather than intrinsic motivation. Notably, this study empirically unpacks the complex interpersonal dynamics of tacit knowledge exchange, demonstrating how both social image—shaped by peers—and management-related reputation and career concerns act as strong incentives for cooperative behavior.

This research contributes empirical evidence to the literature on workplace cooperation and knowledge flows by providing evidence on the supply-side barriers to helping behavior and the mechanisms through which social incentives shape cooperative norms. Additionally, our study adds to the growing body of research on awards and recognition, highlighting their potential as scalable tools for fostering collaboration in professional settings.

Future research could explore the long-term effects of social recognition on workplace dynamics, particularly whether repeated incentives sustain or diminish cooperative behaviors over time. Additionally, while our study focuses on financial institutions, testing similar interventions in other industries—such as healthcare, education, or technology—would further our understanding of how cooperation can be systematically incentivized in different organizational contexts.

Our findings suggest that organizations seeking to enhance cooperation can benefit from simple, well-designed recognition programs that acknowledge and reinforce

prosocial behavior. By aligning incentives with the firm's broader goals, such initiatives can help bridge the gap between individual and organizational objectives, ultimately fostering a more collaborative and productive work environment.

2

Fostering Psychological Safety in Teams

Experimental Evidence from a Multinational Corporation

2.1 Introduction

It has long been argued that team culture and norms are key drivers of high-performing organizations, and that leaders are fundamental to creating and supporting them (Schein, 1990; Blader et al., 2020). However, while growing, causal evidence is scarce of how management interventions succeed in changing culture and norms at the level of the team or the organization.[1]

One norm stands out as especially pertinent in innovative team-based organizations that are at the productivity frontier: psychological safety, described by Edmondson (1999) as the *shared belief held by members of a team that their team is safe for interpersonal risk taking – that others will not embarrass, reject, or punish them for speak-*

*This chapter is joint work with Florian Englmaier (LMU) and Maria Guadalupe (INSEAD).
[1]Some recent exceptions are Alan et al. (2022); Haeckl & Rege (2024).

ing up. It stems from an environment where individuals feel safe voicing ideas and are comfortable speaking up and being themselves without fear of judgment or social repercussions. By facilitating open communication and knowledge generation, psychological safety is considered an important driver of the team's learning, innovation, and performance, particularly in knowledge-intensive environments (Edmondson, 1999; Edmondson & Lei, 2014).

This is why organizations across the board – from Warner Music to the Norwegian Sovereign Wealth Fund – devote so much effort and resources to fostering it within their ranks,[2] In Project Aristotle Google identified psychological safety as key to a team's success (Duhigg, 2016). Conversely, lack of psychological safety has been cited as a major contributor to Boeing's massive quality and safety problems.[3] A workplace culture where individuals feel safe to voice ideas is an important dimension of equity and inclusion – something that matters to organizations striving to incorporate diverse demographics. However, to our knowledge, there is no field-based evidence on what causes psychological safety in organizations, the mechanisms behind it, or how to increase it.[4]

We investigate the determinants and malleability of psychological safety, as well as the role of leaders in fostering its development, through a randomized controlled trial. Partnering with Novartis-Sandoz, a large global pharmaceutical company, we conducted a pre-registered[5] trial with 544 teams, encompassing more than 4,300 employees. To evaluate its effects, we leverage company surveys measuring psychological safety and other team perceptions, along with archival data.

We examine how psychological safety relates to organizational characteristics using archival data on team leader attributes—gender, tenure, seniority, and age—and team composition at baseline, including tenure, age, seniority, gender composition, and international diversity. Only two factors strongly correlate with psychological safety: manager or team tenure and the proportion of women in the team. Teams led by senior managers and those with more women report higher psychological safety. Baseline psychological safety is also linked to higher managerial quality perceptions, a stronger

[2]Financial Times. 2024. "Psychological safety: the art of encouraging teams to be open." Accessed May 3, 2024. https://www.ft.com/content/5d544a25-d7dc-41ad-8b42-fe92406d25d7.

[3]The Wall Street Journal. 2024. "'This Has Been Going on for Years.' Inside Boeing's Manufacturing Mess." Accessed May 3, 2024. https://www.wsj.com/business/airlines/boeing-manufacturing-737-max-alaska-door-plug-spirit-18f7e233?st=wbmx4xdsja1jigr&reflink=desktopwebshare permalink.

[4]Often the managerial recommendations to increase psychological safety are rather generic, and their impact has not been adequately tested (Delizonna, 2017).

[5]See AEA RCT Registry No. AEARCTR-0008359.

belief in team innovation, and a greater sense of security in taking informed risks. Additionally, psychological safety is associated with team stability, as employees are more likely to stay and less likely to make lateral moves for the same pay. All this tells us that psychological safety is associated with key factors that matter for organizations.

Next, we evaluate how to increase psychological safety with a randomized control trial. Our intervention builds on the premise that leaders can shape this cultural dimension and aims to identify effective leader-employee interactions for building psychological safety. Our experiment nudged team leaders, i.e. middle managers, to modify their approach to one-to-one meetings with team members. We designed two treatments in addition to a control group. In both treatments, managers received an email underscoring the importance of one-to-one meetings along with reminders to hold these sessions regularly and focus on individual team members to foster psychological safety. The email included a one-page PDF guide to strategies for these meetings. In treatment T1 (*Needs*), the guide directed managers to focus on individuals by facilitating discussion of their personal needs and aspirations, thereby encouraging active listening.[6] In T2 (*Tasks*) managers were advised to concentrate on workplace-related discussion such as tackling potential barriers to team members' job effectiveness.[7] The managers in the control group were merely informed about the study without specific reference to psychological safety or to conducting one-to-one meetings.[8]

The experiment was designed to ensure scalability and integrate seamlessly with the organization's standard operations. The study was aligned as closely as possible with routine practices – the firm regularly conducted comprehensive surveys that addressed aspects of teams and organizational culture – and our experiment was timed to coincide with a six-week period between two waves of surveys in the fall of 2021.[9] Our

[6]This treatment is based on findings by Kim et al. (2020), that showed how recognizing individual 'Needs leads to increased psychological safety in a laboratory experiment with online participants.

[7]The tasks treatment aligns with traditional views of psychological safety in social psychology, emphasizing the significance of each employee's voice in their work tasks (Edmondson, 1999, 2018)

[8]Our two treatment arms also echo the themes highlighted in Bresman & Edmondson (2022), which propose that three conversational themes within teams enhance psychological safety: "Hopes and goals. What do you want to accomplish?", "Resources and skills. What do you bring to the table?", and "Concerns and obstacles. What are you up against? What are you worried about?". Our needs and tasks treatments directly address the first and third topics.

[9]The six-week duration of the intervention was jointly decided by the research team and the organization to align with the organization's regular survey schedule. This timing ensured that the intervention fit between the scheduled survey waves in the Fall of 2021. Extending the intervention beyond this period would have conflicted with subsequent survey waves, whereas a shorter duration might have been too brief to allow manager to include psychological safety on their regular meetings.

primary sample consisted of teams for which the firm had panel survey data. In addition, we conducted an endline survey to identify changes in the content and frequency of one-to-one meetings.

The primary result was that our intervention significantly improved psychological safety at the team level, particularly with treatment T1 (*Needs*) which emphasized attention to team members' individual needs, yielding a 2% rise in psychological safety compared to the baseline mean of the control group (or 14% of a standard deviation). Treatment T2 (*Tasks*) had on average a smaller (half the size) and statistically insignificant effect on team psychological safety. We also found that while the treatments modified the experiences of team members and norms at the team level, they did not translate into changes in their perceptions of psychological safety in the broader organization. Our findings establish a causal relationship between the type of interaction in one-to-one meetings and the improvement of psychological safety at the team level. They also provide evidence that norms are built locally and that the leader is a key change agent in the formation of norms.

To understand the mechanisms underpinning our core result, as well as the team dynamics at work, we further exploited the organizational and team surveys, the meeting habits survey, and other data from the firm. Exploratory analysis revealed that the effectiveness of the treatments depended on the initial psychological safety levels within teams. The benefits of focusing on personal needs (T1) were especially pronounced in teams in the lowest tercile of initial psychological safety. In contrast, the task-oriented approach (T2) was strongly and highly significantly effective in teams with intermediate (mid tercile) baseline psychological safety levels. This suggests that different strategies may be suitable for different teams, and that individual needs are crucial in teams lacking a basic sense of safety and rapport with their leader. As psychological safety within a team strengthens, shifting focus to discuss job effectiveness and tasks (while still concentrating on individuals) positively affects psychological safety. However, prioritizing task efficiency is not effective without first establishing a solid foundation of psychological safety – suggesting that a sequential strategy is needed to develop psychological safety in teams.

Moreover, our intervention was particularly beneficial for the leaders and teams who needed it most, i.e., junior managers and male-dominated teams that initially had lower psychological safety. This supports our hypothesis of a causal relationship between enhanced managerial skills and psychological safety, indicating that managerial skills can be developed with appropriate guidance. This is important given the role of

managers in implementing better managerial practices.

We also observed changes in other employee-manager dynamics as a consequence of the intervention. First, we investigated how meeting dynamics changed following the treatments, finding evidence that the treatments increased both the frequency and the content of the one-to-one meetings, suggesting that both these may have played a role in increasing psychological safety. It also served as a manipulation check and indicated that (at least some) managers had followed the email recommendations, i.e., complied with the intended treatment. Second, we observed that the needs treatment (T1) altered the dynamic between employees and their manager: employees felt more supported and more likely to view their manager as a role model worthy of recommendation. These outcomes highlight the impact of our intervention in positively reshaping team member perceptions of their leaders.

Our study contributes to the literature on differences in organizational performance, notably the role of team culture and team leaders in explaining them (Gibbons & Henderson, 2012; Bloom et al., 2012; Blader et al., 2020). Our findings are complementary to those of Sandvik et al. (2020), who establish that providing structured opportunities for one-to-one employee conversations leads to increased workplace knowledge transfers, and whose results suggest that the main barriers to communication in teams are social in nature, and can be reduced by fostering psychological safety, as documented here. Our results also relate to findings by Weidmann & Deming (2021) who show the importance of non-cognitive, social skills for teamwork. We portray psychological safety as a social and non-cognitive dimension of differences in performance across teams, explain why such differences may persist over time, and show that even within one organization there is significant dispersion in this norm.

In line with the notion that middle managers explain differences in organizational performance (Lazear et al., 2015; Hoffman & Tadelis, 2021; Friebel et al., 2022), our study documents a causal link between the behavior of managers and the norms that prevail within a team, thus contributing to the literature connecting leadership to corporate culture. We show that leaders are central to shaping team member beliefs about appropriate norms of behavior. Clearly, different leadership styles exist and are related to performance (Bertrand & Schoar, 2003; Bandiera et al., 2020b; Antonakis et al., 2021; Czura et al., 2024a). We contribute to this literature by showing that leadership skills are not fixed but malleable.

Our findings complement the RCT-based findings of Alan et al. (2022) and Haeckl & Rege (2024) that show how training employees and leaders can improve the work-

place climate and the experience of supportive leadership. However, their findings came from interventions directed at all employees (team members) but not focused on a specific cultural norm (e.g., the experiment in Alan et al. (2022), encouraged prosocial behavior and the use of professional language). In contrast, our experiment is targeted exclusively at team leaders to assess their role in a specific, highly sought-after team norm: psychological safety.

Finally, we contribute to the literature on psychological safety by being the first to provide causal, field-based evidence, using an RCT, of how to improve it and of the role of the leader-employee relationship in building it. We test two of the main mechanisms cited in existing work to explain how leaders can increase psychological safety (Bresman & Edmondson, 2022), and present cross-sectional and causal evidence of its drivers and effects.

The rest of the paper is structured as follows. Section 2.2 describes our conceptual framework, the experiment design and summarizes our outcome metrics. In Section 2.3 we present our empirical strategy and discuss the results. Section 2.4 concludes.

2.2 Conceptual Framework and Experimental Design

2.2.1 The Manager's Role in Shaping Psychological Safety

The potential of managers to act as catalysts for positive change within their team, influencing outcomes such as turnover (Friebel et al., 2022), productivity (Englmaier et al., 2024), and firm performance (Bandiera et al., 2020b), is well documented. Psychological safety has also been associated with different leadership styles (Edmondson & Bransby, 2023). Given that psychological safety is fundamentally about team members feeling secure from negative judgment by other members, we hypothesize that there are changes in managers' behavior and skills that can alter the employee's perception of psychological safety *vis-à-vis* their manager (arguably the most important dyad in a team), and that this change in perception extends to the beliefs about the team as a whole. Knowing that a manager values individuals speaking up and would prevent others from shaming them, more ideas should be voiced, significantly altering the dynamics of information sharing within the team.

Managers have two clearly defined sets of tasks in their team: interpersonal and operational. On the interpersonal level, they lead by engaging and motivating individuals

within the team. Operationally, they manage the team's functional tasks, ensuring efficient prioritization of the workload, the strategic allocation of personnel, and the matching of team members' skills with appropriate tasks.

From an interpersonal perspective, leaders can make employees feel valued for their individuality within the organization. They can convey the importance of each employee's unique contribution, ensure their individual needs are valued, and that there is room for them in the organization. Research in psychology suggests that perceiving the organization as a supportive space for self-expression and as psychologically safe can be strengthened by acknowledging the unique value of each individual (Kim et al., 2020). By fostering strong interpersonal relationships and focusing on the employee-manager dynamic, leaders can create an environment where individuals feel valued, thereby encouraging open communication of their unique ideas and thoughts and increasing psychological safety. This observation motivates our first treatment (*T1 (Needs)*).

Operationally, managers can make individuals feel valued by recognizing their contributions to the organization and team. The concept of psychological safety, as in Edmondson & Bransby (2023), focuses on this dimension of team safety, where streamlining work processes and addressing barriers to task execution reinforce the value placed on each employee's professional input. If a leader highlights the significance of each team member's contribution to the organization's objectives and actively works to unlock their potential, it can affect their willingness to share their contribution with the team. By helping them prioritize their tasks in alignment with organizational goals, the leader demonstrates to their direct reports the critical nature of their roles. This ensures they understand that their contributions are not only valued but essential to the team's success, fostering an environment where their ideas and inputs are encouraged. This observation motivates our second treatment (*T2 (Tasks)*) in the experiment.

2.2.2 Experimental Design

The study was conducted in collaboration with Novartis-Sandoz (hereafter referred to as *the firm*), a global pharmaceutical company with headquarters in Switzerland and teams spread across the world. Specifically, our pre-registered experiment involved Sandoz, then a fully-Novartis-owned generics division, with over 20,000 employees and annual revenues of 10bn CHF. Conducted from September to November 2021, the

study encompassed 544 teams comprising 4,349 employees. These were knowledge workers engaged in various non-routine analytical tasks, primarily focusing on pharmaceutical development and sales, working either remotely or in a hybrid setup. The teams were geographically dispersed across more than 65 countries, with Germany hosting the largest number of teams. They were comprised of members with diverse seniority levels and varying durations of employment with the firm.[10]

To be included in our sample, teams were required to have participated in the team-level survey during the last wave before the experiment (the firm only collected information for teams with at least five answers per team) and to have administrative records available. After applying these criteria, out of the 1,000+ teams (7,000+ employees) of non-production workers within Sandoz, our sample included 684 teams, which were randomized into three treatment conditions. The randomization unit therefore was the team. Of these 684 teams, 83% remained in the study by participating in the endline survey in November. Ultimately, 544 teams fulfilled all criteria, forming our main sample.

Table 2.1 presents descriptive statistics of the teams and shows that the randomization process successfully achieved a balance across the treatment conditions in most variables (Columns (3) and (4))[11]. We observe two statistically significant differences in characteristics at baseline for T2 (trust and team average age). While we consider these patterns to be statistical artifacts, we nonetheless control for those variables in all specifications. The control group consisted of 171 teams, while the T1 group and the T2 group comprised 192 and 181 teams, respectively.

The experiment started on September 9th, 2021, with an email to the team managers. Figure 2.1 illustrates the timeline of the experiment. Due to the firm's ongoing commitment to enhancing psychological safety through various initiatives, our study was seen as a natural extension of the organization's efforts to reinforce this cultural aspect. All communications were sent by the firm's leadership. Managers in the control group received a kick-off email from the "Meeting Habits Study Team" informing them about the ongoing study on meeting habits within the firm. This email requested permission to use their data at an aggregate level (the exact phrasing is provided in Figure B1 in the Appendix). Following this initial communication, the control group did not

[10] We do not have data on the specific business area to which each team belongs.

[11] The treatment arms are balanced when examining the characteristics of both the full sample (684 teams) and the final sample (544 teams). We include the table for the final sample, as this is the sample used for all analyses, tables, and results.

Table 2.1: Descriptive Statistics and Balance of Sample

	(1) Baseline Characteristics	(2) Univariate Correlation	(3) Control vs. T1 (Needs)	(4) Control vs. T2 (Tasks)
Team Characteristics				
Age	40.57	-0.08*	0.55	0.45
	(5.96)	(0.05)	(0.65)	(0.63)
Seniority	5.42	0.09**	0.10	
	(0.99)	(0.03)	(0.11)	(0.11)
Female ratio	0.59	0.13***	0.02	0.02
	(0.26)	(0.001)	(0.03)	(0.03)
Tenure	6.81	-0.05	-0.05	0.67
	(3.61)	(0.16)	(0.38)	(0.41)
Size	9.28	-0.01	-0.01	0.18
	(3.34)	(0.69)	(0.34)	(0.37)
Share international	0.09	-0.06	-0.03	-0.01
	(0.28)	(0.14)	(0.03)	(0.03)
Manager Characteristics				
Age	44.35	-0.07*	1.15	2.18***
	(7.48)	(0.09)	(0.80)	(0.77)
Seniority	5.52	0.15***	0.20	0.04
	(1.25)	(0.00)	(0.14)	(0.13)
Female	0.40	0.067	0.02	-0.01
	(0.49)	(0.11)	(0.05)	(0.05)
Tenure	9.22	-0.03	0.11	0.79
	(6.59)	(0.45)	(0.68)	(0.71)
Survey Metrics (0-100)				
Team PsyS	83.23		0.13	-0.44
	(11.48)		(1.20)	(1.25)
Firm PsyS	74.43	0.59***	-1.27	-1.27
	(12.64)	(0.000)	(1.33)	(1.38)
Firm Trust	80.65	0.63***	-1.26	-2.00*
	(10.85)	(0.00)	(1.16)	(1.19)
Experimenting	85.18	0.86***	-0.35	-0.63
	(10.77)	(0.00)	(1.12)	(1.15)
Take Informed Risks	81.22	0.81***	0.54	0.40
	(12.35)	(0.00)	(1.31)	(1.34)
Support of Manager	81.53	0.83***	-0.02	-0.35
	(11.71)	0.00	(1.25)	(1.26)
Quality of Manager	86.83	0.90***	0.65	0.46
	(12.24)	(0.00)	(1.30)	(1.35)
Team Turnover				
Movs within the Firm	1.32	-0.15***	0.13	0.23
	(2.00)	(0.00)	(0.19)	(0.22)
Promotions	0.44	-0.03	-0.06	-0.07
	(0.82)	(0.36)	(0.09)	(0.09)
Lateral Movs	0.88	-0.14***	0.19	0.30
	(1.86)	(0.00)	(0.18)	(0.20)
Hiring	0.26	-0.01	0.01	-0.05
	(0.58)	(0.66)	(0.06)	(0.06)

Notes: Column (1) includes the baseline summary statistics of team characteristics. The standard deviation is reported in parentheses. Column (2) reports uni-variate Pearson correlation coefficients of the variables with team level psychological safety. The differences between the control and T1 or T2 in Columns (3) and (4) show the coefficient of a simple regression of the variable on a treatment group dummy with robust standard errors. The variables *age, seniority, time at firm, and team size* are team or manager averages. The variable *gender* is a ratio of female to male. The survey metrics are the answers to the baseline levels of the team and organizational survey, answers range from 0 to 100, with being 100 the highest level of agreement. *Baseline Team PsyS* is the answer to *I feel safe sharing feedback with colleagues* and *Different perspectives are valued in my team*. *Baseline Firm PsyS* is *I feel free to speak my mind without fear of negative consequences*, *Trust* collects the answers to the question *I trust colleagues across Novartis*. *** p <0.01, ** p <0.05, * p <0.1.

receive any further messages until the conclusion of the experiment.[12]

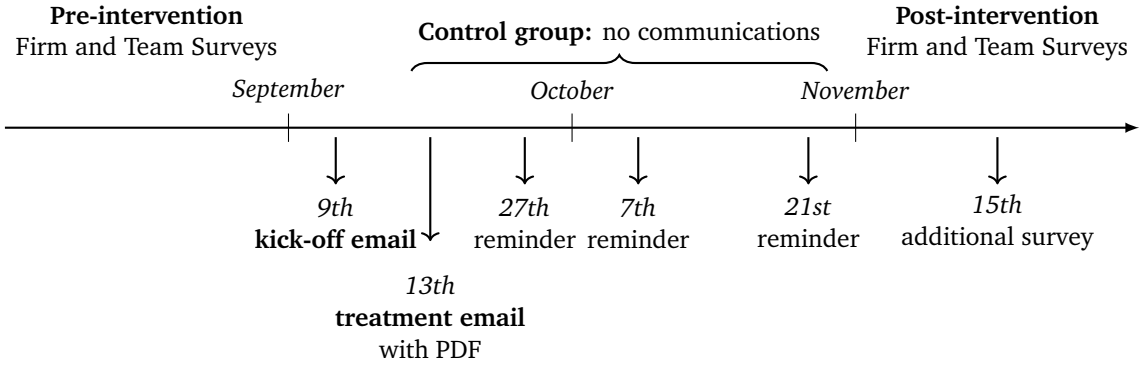


Figure 2.1: Timeline of the Experiment

Throughout the study period, the communications sent by top management to team leaders in the two treatment groups reminded them of the importance of psychological safety and encouraged them to focus on it in their one-to-one meetings. The content sent in the main body of the email to all leaders was identical. However, each treatment arm received a different attached PDF, with different instructions on how to conduct those meetings, following research in organizational psychology as discussed above (Kim et al., 2020; Bresman & Edmondson, 2022).

The *Needs* treatment (T1) aimed to support leaders to make employees feel valued for their individuality within the organization (Kim et al., 2020). Team leaders in T1 received a one-page PDF guide attached to the main email on how to identify and adapt to each member's individual needs. The header of the PDF stated, "*Use your one-to-one time to discover and adapt to the individual needs of your team*". It continued: "*Over the next six weeks, we encourage you to shift the focus in your one-to-one meetings from day-to-day work to what matters most to your team member in the long term, helping them feel like unique individuals within the larger organization*". The PDF then suggested specific behaviors, such as "*Invite Insight*", "*Be Empathetic*", and "*Ask Questions*" and provided examples for implementing each. For instance, under "*Invite Insight*", it recommended to "*Ask your team members to bring a meaningful topic to your next one-to-one meeting*". The content of the PDF is shown in Appendix Figure B4.

The second treatment group, *Tasks* (T2), also received instructions about conducting more frequent one-to-one meetings, but the attached PDF differed from that in T1,

[12]While we have no evidence of information spillovers from the treatments, we cannot preclude contamination of the control group if managers share the information. However, in that case any effects documented can then be viewed as a lower bound.

focusing on the execution of tasks, and how to best eliminate barriers and distractions that impede team members' most valuable contributions (Bresman & Edmondson, 2022). The PDF for T2 managers stated, "*Use your one-to-one time to remove barriers and distractions from people making their most valuable contribution*". Echoing the needs treatment, the PDF instructed, "*Over the next six weeks, we encourage you to use your one-to-one time to simplify your team's lives. This involves identifying what hinders their best work, be it competing goals, deprioritized projects, process inefficiencies, or technology issues*". It then offered concrete advice on structuring conversations around task prioritization and barrier removal, with a focus on the team member's perspective. The aim was to prompt managers to concentrate on supporting each employee in effectively performing their tasks and removing obstacles deemed significant by the team member (detailed content is available in Appendix Figure B4).

2.2.3 Outcomes and Data

We had access to survey data from two internal surveys conducted regularly by the firm. These surveys (sent to employees at different times) were designed to elicit employees' perceptions regarding two different reference groups: one asked about employees' perceptions of the team, the other about their perceptions of the organization as a whole. The data from these surveys was available for the waves immediately preceding and following our intervention.

As is common in many organizations, to ensure anonymity and promote candid responses and high participation rates these surveys were administered by an external agency who collected the data and aggregated them at the team level for teams with at least five responses before sharing it with the firm. This approach resulted in a remarkably high response rate of 75%. However, it also meant that we could not have access to individual-level responses (nor did the company). Given our study's focus on norm changes at the team level, and that the treatment was at the team level, this limitation is of minor concern for our main analysis. Therefore, in what follows all survey variables are averages at the team level for teams where at least five employees participated in the survey.

Primary Analysis: Psychological Safety Variables

Our primary outcome measure is psychological safety, assessed at both the team and organizational levels. For the team level, we measure psychological safety by the average of team responses to two specific questions in the survey run specifically to measure team attitudes: "*I feel safe sharing feedback with colleagues*" and "*Different perspectives are valued in my team*". At the organizational level, the measure of psychological safety is the team average of responses to one question in a different survey that targets employees' perceptions of the entire organization: "*I feel free to speak my mind without fear of negative consequences*". Participants rate their agreement with each statement on a scale from 0 to 100. These metrics align closely with established measures in organizational psychology (Edmondson, 1999; Baer & Frese, 2003), even though the wording is not identical. However, it is unique to have access to proxies for psychological safety relative to two relevant reference groups where norms are potentially built: the immediate team and the organization as a whole.

The final sample for our study comprised teams that completed both the baseline and endline surveys. We anticipated some level of attrition given that historical survey participation rates were typically around 75% per wave. Importantly, we observed no differential attrition rates across the treatment conditions. The attrition rates were 21% for the control group, 20% for T1, and 22% for T2, again indicating that the three conditions were comparable.

Exploring Mechanisms: Additional Outcomes

In addition to the core analysis of changes in the norms around psychological safety following the intervention, we also analyze its impact on other relevant outcomes measured by the team survey (see Table B1 in the Appendix for a detailed description of the variables). This allows us to provide a fuller picture of the impact of the intervention on perceptions and attitudes of employees and can therefore shed light on the mechanisms by which the intervention affects psychological safety. The analysis of these variables is complementary to the results and should be interpreted jointly with those. In particular, we were are able to analyze variables on the "quality" of the manager (two statements: "*My manager is a role model of our company values and behaviors*", and "*I would recommend my manager to other associates in this firm*") and the degree of supportiveness of the manager (five statements:"*I feel supported when tackling obstacles that hinder my best work", "I have support for my career development", "I*

receive ongoing coaching that helps me to constantly develop", "I receive regular feedback to improve my performance", and "*I am satisfied with the recognition I receive for my work"*). We also analyze variables that link to innovative behaviors of the team such as experimenting ("*I am encouraged to find new and better ways to get things done"*) and taking risks (*"I am encouraged to take informed risks in getting my work done"*). Please note that while this metric may appear similar to Edmondson (1999) validated survey metric, "*It is safe to take a risk*," which captures interpersonal risk-taking in the context of speaking up, it measures a different construct. Specifically, our survey metric focuses on risk-taking related to work tasks and innovation. Therefore, it cannot be automatically subsumed under psychological safety, as it addresses distinct aspects of behavior, even if sharing a similar word.

Changes in Behavior: Meeting Habits

We conducted a survey on meeting habits at the end of the study to analyze actual differential changes in behavior between treatment arms concerning the frequency and content of the one-to-one meetings. Due to confidentiality constraints, the firm did not permit identification of individuals or teams in this survey. However, we were able to identify the treatment group to which each survey respondent belonged. We received a total of 1,524 responses, with 514 from the control group, 507 from T1, and 503 from T2. The survey queried the frequency of one-to-one meetings with the question *"Reflecting on the last couple of months, how frequently do you have one-to-one meetings with your manager?"*. Respondents had to answer on a scale ranging from less than every two months to more than twice a week.

Additionally, the survey included questions on the issues discussed in these meetings, allowing respondents to select from a list that included career-related topics (such as performance, career development and team relationships), personal-related aspects (including personal life and work-life balance), and project-specific issues (such as prioritizing projects and removing barriers). These helped us assess the extent to which different topics were covered in the one-to-one meetings across various treatment groups.

In parallel to the survey, the firm made available data on one-to-one meetings at the team level from Microsoft Workplace Analytics,[13] such as the average number of meet-

[13]The research team did not have access to any other potential metrics available through Microsoft Workplace Analytics.

ings recorded on the Microsoft calendar between each manager and their direct reports in June and November 2021. For GDPR-related reasons, this information was available for only a subset of teams (217 out of the more than 1,000 teams within the firm). Within our experimental sample we had data for 144 teams, which were used as a supplementary validation of behavioral changes.

Administrative Data: Demographic Turnover

The firm provided employee-level administrative data, including variables such as gender, age group, generation, nationality, length of tenure within the company, seniority level (ranging from 1 for the most senior to 8 for the least senior), and country of employment. We also obtained team turnover data, i.e., unique data on the total number of movements by team until three months after the intervention, where movements within the firm (from one team to another) were classified as lateral transfers (if the employee stays within the same wage band and hierarchical level) or promotions. We also had data on hiring, but hiring was limited in the period of the study (entering the second winter of the COVID-19 pandemic).

2.3 Results

Correlates of Psychological Safety in Teams: Demographics, Innovative Behaviors, and Team Stability

Table 2.1 presents descriptive statistics of our sample at baseline and the correlation of team characteristics with team level psychological safety. Teams rate the psychological safety questions between 0 and 100. The average team-level psychological safety is 83 and the average organizational-level psychological safety is somewhat lower at 74. Our sample covers teams across the organization and, as Column (1) shows, employees have an average age of 40.5, with the average seniority level at 5 (out of 7 possible hierarchical levels), and have spent 6.8 years at the firm. The share of female employees is 59 percent. 1 out of 10 employees are "international" (i.e. from a country different form the country of residence). Managers are older (44 on average), more likely to be male (40 percent female), and have longer tenure at the firm (9.2 years).

Column (2) describes which characteristics of our archival demographics at the team

and manager level are correlated with psychological safety. At the team level, two variables stand out as key predictors of team psychological safety. First, psychological safety increases with manager seniority. This may indicate that as managers accumulate skills over time, they become more skilled at creating an environment where team members feel secure in expressing their opinions and providing feedback. Alternatively, managers that foster psychological safety are more likely to be promoted to more senior levels. Either way, teams with less senior managers have lower scores. This is complemented by the survey correlations: psychological safety also positively correlates with perceptions of support from the manager and the quality of the manager. All this together points out to the idea that the manager plays a role in the team's psychological safety.

Second, teams with a higher proportion of females display higher team-level psychological safety levels, suggesting that women have a greater ability to foster an environment open to self-expression. This latter correlation between gender and elevated psychological safety levels has not been documented, as far as we know, in the psychological safety literature.

We also observe that psychological safety correlates with two dimensions that can arguably be associated with team performance. Organizational psychology scholars have hypothesized that when employees feel safe to voice their ideas and take informed risks, they are more likely to engage in creative problem-solving and contribute to organizational innovation (Edmondson, 1999; Carmeli et al., 2010). We proxy innovation efforts with two variables from the team survey: *"I am encouraged to find new and better ways to get things done*" (Experimenting), and *"I am encouraged to take informed risks in getting my work done"* (Take Informed Risks). Team-level psychological safety correlates with both innovation effort measures. These results align with other studies that discuss the important role of psychological safety on innovation and willingness to take risks (Gong et al., 2012; Binyamin et al., 2018; Iqbal et al., 2022). Lastly, team psychological safety is negatively correlated with lateral movements within the team, suggesting that individuals are more likely to stay in teams with higher psychological safety.

We must emphasize that all these associations are correlational and do not allow for any causal interpretation. Nonetheless, they illustrate how psychological safety relates to key organizational outcomes. In this context, improving psychological safety in teams appears particularly important. In the following section, we apply the results of the experiment to establish causal effects.

Improving Psychological Safety: Experimental Evidence

This section presents the results of the randomized field experiment mentioned earlier. Teams were randomized into three groups: the control group, the *Needs* treatment (T1), and the *Tasks* treatment (T2). To assess the impact of this treatment assignment on psychological safety, we use the following specification:

$$Y_i = \alpha + \beta T1_i + \gamma T2_i + \delta_0 X_i + u_i, \tag{2.1}$$

where Y_i is the level of psychological safety at endline in team i. $T1_i$ and $T2_i$ are dummy variables that indicate assignment to T1 (*Needs* treatment) and T2 (*Tasks* treatment), respectively. β is the intent-to-treat estimator of the impact of *Needs* on the outcome. γ represents the equivalent for the *Tasks* treatment. $\delta_0 X_i$ includes the baseline controls, which include baseline levels of psychological safety. u_i is the error term.

Given the randomization, β and γ can be interpreted as the causal impact of assignment to treatment. Our regressions use robust standard errors, as the unit of observation corresponds to the unit of randomization, which is the team. As presented above, the balancing tests in Table 2.1 (Columns (3) and (4)) show slight imbalances in baseline age and trust. Thus, we control for these in the main specifications.

Table 2.2 is divided into two panels. Panel A shows the results with respect to team-level psychological safety, and Panel B the results with respect to organizational-level psychological safety. Column (1) includes baseline survey controls (organizational trust and baseline psychological safety). Column (2) adds team-demographic controls (share of international members, average age level, average seniority, team size, average worker tenure in firm).

Panel A reveals that team-level psychological safety increased following treatment T1 (*Needs*). This effect is statistically significant when accounting for baseline levels of the dependent variable to absorb variation in the outcome variable. In our preferred specification, as shown in Column (1), the endline psychological safety score is 1.6 points higher in *Needs* compared to the control group, and this difference is statistically significant at the 5% level. The baseline psychological safety for the control group had a mean of 82.9 points (on a scale of 1-100) with a standard deviation of 12 points. Therefore, the *Needs* treatment resulted in a psychological safety increase of 14% of a standard deviation or 2.04% of the baseline control mean. In contrast,

Table 2.2: Effect of the Intervention on Team and Organizational Psychological Safety

	(1)	(2)	(3)
Panel A	**Team Level Psychological Safety**		
			<10%
T1 (Needs)	1.694**	1.745**	1.976**
	(0.78)	(0.78)	(0.84)
T2 (Tasks)	0.844	0.886	0.966
	(0.77)	(0.78)	(0.86)
Mean Dep. Var.	82.92		81.01
Dem. Controls	No	Yes	Yes
Survey Controls	Yes	Yes	Yes
Teams	544	544	488
Employees	4,349	4,349	
Panel B	**Organizational Level Psychological Safety**		
			<10%
T1 (Needs)	0.128	0.196	0.0595
	(0.92)	(0.91)	(1.00)
T2 (Tasks)	-1.233	-0.987	-1.370
	(0.93)	(0.93)	(1.03)
Mean Dep. Var.	75.61		73.47
Dem. Controls	No	Yes	Yes
Survey Controls	Yes	Yes	Yes
Teams	500	500	444
Employees	4,121	4,121	

Notes: OLS regression of the treatment assignment on PsyS with robust standard errors. The dependent variable in Panel A is the team-level PsyS. This metric is the aggregate of the two survey questions *Different perspectives are valued in my team* and *I feel safe sharing feedback with colleagues*. The dependent variable in Panel B is the organizational level PsyS. This metric is the survey question *I feel free to speak my mind without fear of negative consequences*. All answers range from 0-100 being 100 the highest level of agreement with the statement. Column (1) reports the regression results with survey controls. Column (2) adds team-demographic controls and baseline survey controls. Column (3) shows the results when dropping the 10% highest scoring teams in PsyS at baseline. Panel B presents the same specification, with organizational-level psychological safety as the dependent variable, represented by *I feel free to speak my mind without fear of negative consequences*. *** p <0.01, ** p <0.05, * p <0.1.

the response to T2 (*Tasks*) was positive but smaller in magnitude, with an increase of 0.8 points, equivalent to 1% of the baseline control mean. These findings are robust when weighted for the number of survey responses in each team (see Table B2 in the appendix) and remain significant after adjusting for multiple hypothesis testing (see Table B3 in the appendix). The results are also stronger in Column (3) when we drop teams in the top decile of baseline psychological safety. These scores are already above 96.5 out of 100 points. This ceiling implies a limit on how much improvement can take place.

Note that all of these estimates represent intent-to-treat effects. While our emails encouraged managers to modify their behavior and the focus of discussions in the one-to-one meetings, we did not enforce compliance. This non-invasive approach allowed managers to exercise discretion regarding the extent to which they believed their teams would benefit from the recommendations.

Turning our attention to organizational-level psychological safety (Panel B), we find no statistically significant changes resulting from any of the treatments on the freedom to speak up at the organizational level, our proxy for psychological safety. The point estimates for *Needs* are economically small. For *Tasks* the coefficients are also statistically insignificant, and if anything are negative. Given the lack of significance of the impact of the intervention at the organizational level, we conclude that the intervention did not influence the average organizational psychological safety.

In summary, the experiment led to a significant increase in team-level psychological safety, particularly for teams in the T1 (*Needs*) treatment. This suggests that one-to-one meetings can serve as an effective tool for leaders to positively influence psychological safety within their teams, and hence that the leader is an important driver of team norms. Moreover, psychological safety can be impacted in the short term through focused conversations and guidance on how to conduct one-to-one meetings. However, the absence of an effect on organizational-level psychological safety implies that the enhanced team dynamics do not automatically translate to the broader organizational context. The randomization process also ensures that the drivers of these effects are not related to differences in manager personality, team structures, or other factors that have been proposed as correlates of psychological safety. Nor are they influenced by reverse causation from psychological safety to leadership; rather, the change in and around one-to-one meetings is the key driver of these results.

To further examine the impact of our intervention and better understand its mechanisms, we present additional analyses below.

Table 2.3: Heterogeneous Effect of the Intervention

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Terciles			Ratio Female		Seniority of Manager	
	Low	Middle	High	Below Median	Above Median	Junior	Senior
T1 (Needs)	4.878***	0.616	-0.361	2.741**	0.479	3.051***	0.195
	(1.64)	(1.18)	(1.10)	(1.13)	(1.06)	(0.96)	(1.16)
T2 (Tasks)	0.957	2.898**	-0.883	1.742	0.046	1.414	-0.208
	(1.79)	(1.22)	(1.05)	(1.34)	(0.96)	(1.13)	(1.12)
Mean Dep. Var.	69.28	84.89	94.60	81.89	83.57	79.21	85.56
Teams	179	177	188	248	296	233	311

Notes: OLS regression results with team level psychological safety as the dependent variable with robust standard errors. PsyS is the aggregated from responses to *Different perspectives are valued in my team* and *I feel safe sharing feedback with colleagues*. All responses are scaled from 0 to 100, where 100 indicates the highest level of agreement. Columns (1)-(3) display results for teams categorized into terciles based on baseline PsyS scores. Columns (4) and (5) differentiate teams below and above the median share of women in the firm, set at 60%. Columns (6) and (7) are segmented by the seniority of the team manager, with junior managers classified as levels 1-5 and senior managers as levels 6-7. All columns include survey and team-demographics controls. *** p <0.01, ** p <0.05, * p <0.1.

Who Benefits Most From the Treatment: Heterogeneus Effects of the Intervention

After observing a positive average effect of the intervention, we sought to understand which teams benefited most from it. Our experiment was designed to establish causal average effects on psychological safety with the full sample, so the results below can be seen as exploratory, albeit they provide a richer picture of how the treatment operated in practice.

First, we explore whether teams with different starting conditions, i.e. levels of baseline psychological safety, responded differently to our treatments. Columns (1) to (3) of Table 2.3 present the treatment effects across terciles of baseline team level psychological safety, using the same specification as in Column (2) of Table 2.2. Column (1) show that T1 (*Needs*) proved particularly effective for teams in the lowest tercile at baseline (the coefficient is statistically significant at the 1% level and substantial in magnitude: 4.8 points). Column (2) shows that this is less so for teams in the middle tercile. In contrast, while the treatment effect for T2 (*Tasks*) appears less prominent and less precisely estimated in the lower tercile (as well as when considering the entire sample), it is very effective for teams with intermediate levels of psychological safety at baseline (the coefficient here is statistically significant at the 5% level and substantial in magnitude: 2.8 points).

For the upper tercile, Column (3) confirms a minimal and imprecisely measured effect on teams that initially scored at the highest level. This outcome aligns with our

expectations, as these teams had already achieved a high psychological safety score exceeding 90 out of 100, encountering a natural ceiling. So if anything, the effect is more likely to be negative due to reversion to the mean.

A plausible interpretation of these findings is that focusing on the individual is the first required intervention for teams with low psychological safety to build a solid base. Once this is established, the value of addressing individual needs diminishes, perhaps because managers are already addressing them, while the value of focusing on other dimensions like task execution or prioritization (the focus of T2 (*Tasks*)) increases. Although our design does not allow us to examine the simultaneous effects of both treatments, these results provide guidance for intervention strategies based on a team's initial level of safety metrics.

This analysis of heterogeneity underscores the significance of considering initial conditions when assessing the effectiveness of our interventions. By excluding teams with exceptionally high baseline psychological safety scores, we can discern a clearer treatment effect and gain a better understanding of the relationship between the treatments and changes in psychological safety.

In subsection 2.3 we identified two factors that correlated with the level of psychological safety within teams: gender composition and the seniority level of managers. Building upon these observations, we investigate whether the intervention primarily benefits those who may need it most, specifically teams with a higher proportion of male members and teams led by junior managers. It is common for the gains from improved information or training to be primarily absorbed by those who are already proficient in fostering them. To explore this question, Columns (4) to (7) of Table 2.3, replicate the main specification by splitting the sample based on the median value of the female ratio (60%; Columns (4) and (5)) and the median seniority level of managers (seniors of level 6 and above; Column (6) and (7)).

Comparing Columns (4) and (5), we observe a substantial and positive effect of T1 (*Needs*) on psychological safety for teams with fewer females, while no significant effect is observed for teams with above-median shares of females (Column (5)). Similarly, in Column (6), a strong and significant positive effect of T1 (*Needs*) is evident for teams with junior managers, whereas no significant effect is observed in Column (7) for teams with senior managers.

Overall, the results suggest that the intervention was indeed successful in improving psychological safety for teams where it was needed most: helping managers of weaker

teams. These findings indicate that the treatment may help mitigate disparities in psychological safety and provide targeted support for teams facing greater challenges in this regard. Moreover, they highlight the potential for managers and teams to learn and improve psychological safety, especially teams that lag behind.[14]

Examining Mechanisms: Changes in Employee-Manager Interactions and Dynamics

Next, we explore the possible mechanisms driving the increase in team-level psychological safety induced by our interventions. Our communications encouraged managers to both increase the frequency of one-to-one meetings with each team member and modify the content of these meetings (focusing on *Needs* and *Tasks*, respectively). Table 2.5 presents the results from the survey on meeting frequency and habits conducted at the end of the study. Employees were asked how frequently they met with their manager, choosing a category from 'less than once every two months' (=1) to 'more than once a week' (=6). Survey answers are provided at the individual level and compared across treatments. Given the grouped nature of the answers, we run an ordered logit regression of the frequency (as measured by these categories) on the treatment group to which the employee belonged.[15]

Column (1) of Table 2.5 shows that employees in both treatment conditions reported having more frequent meetings compared to the control group. This indicates that, on average, managers did not disregard the intervention email and took actions based on its recommendations, increasing the frequency of one-to-one meetings in both treatment conditions.[16] We also were able to obtain Workplace Analytics (WPA) data on one-to-one meetings registered in the managers' calendar for a subset of teams in our sample (144 teams only). While this is a very different dataset, we confirm the above

[14] For completeness, we conduct the analysis for all the team and manager demographics provided by the organization available in Table 2.1. For the team, these include age, seniority, tenure, size, and the share of international members. For the manager, these include age, gender, and seniority. We find heterogeneous effects only in the variables described above, which were correlated at baseline with psychological safety.

[15] As mentioned above, while we know which treatment group they belonged to, for confidentiality reasons we were not able to identify individuals or teams in this survey so we cannot match the answers to any individual or team level characteristics. As we can match the survey responses only to the treatment arms and not to specific teams, we cannot use this survey to determine local effects of the intervention, or instrumental variables approach to understand drivers such as increased time with manager.

[16] Due to agreements with the work councils, the Workplace Analytics Data is only available for a limited subset of countries.

Table 2.4: Changes in Employee-Manager Dynamics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Meeting Frequency		Meeting Content			Perceptions of Manager	
	Survey	WPA	Career	Personal	Projects	Support	Quality
T1 (*Needs*)	0.314***	0.398**	0.036	0.067	0.147	1.58*	2.00 ***
	(0.115)	(0.180)	(0.109)	(0.109)	(0.110)	(0.84)	0.75)
T2 (*Tasks*)	0.232**	0.0364	0.033	0.177*	0.217**	0.74	0.69
	(0.109)	(0.0.173)	(0.106)	(0.107)	(0.110)	(0.87)	(0.76)
Employees	1,524		1,524	1,524	1,524	4,349	4,349
Teams		144				544	544
Survey & Dem Controls	No	Yes	No	No	No	Yes	Yes

Notes: The coefficients are estimated using an ordered logit regression with robust standard errors. They are the answers to the endline survey carried out at the end of the study period. Due to internal data regulations, we could not map the answer to the teams, and only to treatment arms. Column (1) is the answer to *Reflecting on the last couple of months, how frequently do you have 1:1 meetings with your manager?* Survey answers go from less than once every two months (=1) to more than once a week (=6). Survey answers are provided at the individual level and compared across treatments (with no team information). Column (2) is the workplace analytics data for the average number of meetings with direct subordinates in August 2021 and November 2021. Columns (3) - (5) include three blocks of question capturing *Career-related*, *Personal-related*, and *Projec-specific related* content in the meetings. The questions were framed as *Over the last couple of months, to what extent did you discuss the following topics in your 1:1 meetings*. *Career-related* questions included: Your (Current/future) performance, your career development, relationships within the team, relationships in the wider organization, training, and learning on the job. *Personal-related questions* included: personal life, work-life balance, your personal individual needs, and aspirations. *Project-specific related* included: prioritizing of projects and workload, removal of barriers or blockers to perform tasks. Columns (6) and (7) report OLS estimates from the team perspective surveys on two indices measuring management perceptions of *Support* and *Quality*, detailed survey questions can be found in Table B1. *** p <0.01, ** p <0.05, * p <0.1.

findings and observe that the number of meetings in the T1 condition increased significantly (coefficient of 0.398) compared to the control. The WPA data gives a continuous metric which allows us to easily assess magnitudes of this effect. For an average team of 8 individuals, the total number of meetings at endline increased by 3.2 per month. If we evaluate it at the control mean of 1.42, the increase would bring the average number of one-to-one meetings with the manager to almost two per month (a 28% increase). In contrast, we see no economically or statistically significant change in meetings in the T2 condition in the WPA data.

Our treatments also encouraged managers to change the nature of the conversation. While we could not directly observe the content of these meetings, we attempted to capture some of it using an endline survey. Employees were given a list of possible topics covered in their one-to-one meetings, such as career questions on performance, personal questions on work-life balance, and project-related questions on task prioritization (see Appendix Table B1 for the exact questions included in each of these broad topics).

Columns (3)-(5) of Table 2.5 show the results of regressing the likelihood of discussing

specific topics on treatment status and other relevant factors. Given the categorical and ordered nature of the answers, we use an ordered logit model. The results indicate that there are no significant differences in discussions related to career matters. There is a positive and significant effect for T2 in the probability of discussing personal topics (coefficient of 0.177 in Column (4)) as well as for the probability of having project-specific discussions, such as workload prioritization and barrier removal (coefficient of 0.217 in Column (5)). For T1, the effects are also positive but smaller in magnitude and not significantly different from 0 at conventional levels. Neither are they statistically different from T2).

These results show that meeting frequency and content clearly changed after the interventions. T1 had a clearer and stronger effect on the number of meetings; T2 on the content of meetings. This may reflect in part the fact that T1 asked the employee to raise topics they were interested in during the discussion; hence if there was more diversity/dispersion in the topics they brought up, it is harder to identify an effect on any one dimension. Unfortunately we cannot map the results to other team characteristics (such as baseline psychological safety) to explore different reactions.

The results also serve as a manipulation check that the interventions did have an effect, and that we observe effects in three unrelated set of data: our survey, calendar data from WPA, and the organizational surveys. While our intervention was designed to increase psychological safety in the latter, the body of results helps us understand what happened in each treatment within the teams. Finally, using a separate set of questions from the team survey, we show how perceptions of the manager changed with the treatment. By observing how other perceptions changed in the teams, we can provide a fuller picture of changing norms within teams as a result of the intervention.

Column (6) shows that team members' perceptions of the leader being supportive improved in *Needs*.[17] This suggests that conversations that we encouraged to increase psychological safety did increase perceptions of managers' supportive behavior. In addition, Column (7) shows that in *Needs* managers received a higher approval rating from their team: teams were more likely to perceive the manager as a role model, and more prone to recommend their manager to other associates in the firm. Consistent with the average treatment effect on psychological safety, the effect is also positive for T2 on these last two variables but smaller in size and not statistically significant. Thus, while our intervention aimed to increase psychological safety, it appears to have also been conducive to an improved perception of the managers' supportive behaviors and

[17]This is measured by an index of five questions. See Appendix Table B1 for the exact wording.

made them more appreciated overall.[18]

In contrast, we find no significant changes in organizational- level metrics. The coefficients for all variables are generally smaller in magnitude and not statistically significant. The comparison of team-level and organizational-level dynamics confirms that while the intervention had a positive effect on a number of dimensions of team-level perceptions (in particular those linked to perceptions of the manager), those effects did not translate into positive attitudes towards the firm as a whole, at least not in the short run. This suggests a mechanism through which cultural norms develop in response to the leader's behavior and take shape at the team level, where the team leader can influence beliefs. However, this shift in beliefs at the team level does not automatically extend to the organizational level. While this goes beyond the scope of the present study, we hypothesize that for this to occur employees would need to observe new behaviors and modes of interaction among others in the organization, and possibly witness new behaviors exhibited by leaders at higher levels.

2.3.1 Effects on Innovative Behaviors and Team Stability

In subsection 2.3 we identified a number of outcomes relevant to firm performance that correlate with psychological safety in the baseline, namely dimensions of innovative behavior and turnover at the team level. Table 2.5 shows the effects of the intervention on these variables. We find evidence that our experiment, in particular T1, significantly increased openness to experiment and take risks by the teams (Columns (1) and (2)). This reflects a causal impact on these reported innovative behaviors by team members as a result of the changed meeting behaviors of the manager. Moreover, given that measurable innovation outcomes in the pharmaceutical context often take years (sometimes decades) to appear, and that our teams are so diverse that no comparable hard measures across them exist, these self-reported measures are arguably a sensible substitute.

We also analyze effects on staff turnover using personnel records at endline, for December 2021 to February 2022. Unfortunately, we cannot extend the analysis beyond that date since after our experiment succeeded in increasing psychological safety, in January 2022 Novartis-Sandoz started to widely communicate (and roll out) the findings to the entire organization, and in late spring 2022 it was announced that Sandoz

[18] These are also important organizational outcomes and these findings complement the results in Alan et al. (2022) and Haeckl & Rege (2024).

Table 2.5: Effect of the Intervention on Experimenting, Taking informed Risks, and Team Stability

	(1)	(2)	(3)	(4)	(5)	(6)
			Within Firm Moves			
	Experiment	Take Risks	All	Lateral	Promotions	Hiring
T1 (Needs)	1.77**	1.12	-0.41	-0.32	-0.48	0.18
	(0.84)	(0.81)	(0.28)	(0.38)	(0.37)	(0.90)
T2 (Tasks)	0.79	0.84	-0.20	-0.21	-0.20	0.38
	(0.84)	(0.84)	(0.28)	(0.40)	(0.35)	(1.02)
Mean Dep. Var.	85.21	80.91				
Employees	4,349	4,349				
Teams	544	544	544	544	544	544
Total Movements			190	96	94	8

Notes: This table displays the findings from a Poisson regression estimation, where the dependent variable is the count of team changes across different categories within three months post-intervention (December 2021 to February 2022). Column (1) aggregates results from columns (2) and (3). Specifically, column (2) details instances where employees transitioned to another team without promotion, whereas column (3) includes both in-band promotions and those with a change in band. Column (4) addresses the recruitment and re-hiring of employees from outside the organization. The regression model controls for survey factors (such as the baseline level of trust) and characteristics pertaining to the team and manager, including average years in the firm, team size, the proportion of international team members, average team age, level of team seniority, manager's age, manager's gender, and manager's level within the organization. Significance levels are denoted as follows: *** for p <0.01, ** for p <0.05, and * for p <0. *** $p < 0.01$, ** fo $p < 0.05$, * $p < 0.10$.

would be spun-off, which created instability in the firm. Note that the experiment was not designed to reduce turnover, and we only obtained the personnel records after the study was completed.

Columns (3) to (6) of Table 6 show the estimated effects of our intervention on various measures of turnover. We find that all types of moves away from the team fell (total in Column (3), lateral moves in Column (4) and promotions in Column (5)) following the treatments relative to the control group. The point estimates are larger for T1 than T2. In contrast, the sign for the number of people joining the team (Column (6)) is positive: people were more likely to join those teams. Still, none of these effects are statistically significant.

In addition, note that while robust and statistically significant, the quantitative effects of our intervention were moderate. This implies that any plausible effects on turnover should be small. However, the consistent directional effects point to the fact that individuals are more at ease in the teams and less likely to leave. We consider these trends in team stability post-intervention noteworthy and encourage further studies to focus on longer-term effects to better understand the potential of psychological safety

to foster not just performance, but also team cohesion and stability.

2.4 Conclusion

Understanding how organizations function effectively is a central theme in economics. Recent research has focused on understanding information flows within organizations, recognizing their critical role in decision-making, innovation, and overall performance. It has also emphasized the role of norms and workplace cultures and climates in determining performance. This study complements this growing body of work by focusing on a crucial element of information exchange within teams: norms supporting psychological safety.

Our research, the first to utilize a randomized control trial design to investigate psychological safety in a real-world setting, sheds light on several key areas. First, we establish a causal link between leader behavior and the development of this crucial norm. Second, we observe differentials in the effectiveness of leader interventions depending on the initial level of psychological safety within a team; our intervention had a particularly positive impact on teams with initially low psychological safety. Moreover, we demonstrate the positive impact of enhanced psychological safety on employee perceptions of the leader and team dynamics, laying the groundwork for potential long-term performance benefits. Finally, we document a number of important patterns in cross-sectional data: teams led by more senior managers and with a higher proportion of women exhibit higher levels of psychological safety.

These results not only shed light on the malleability of psychological safety but also highlight the importance of leadership behaviors in shaping team culture. Our findings contribute to the literature on team performance, leadership, and workplace norms by demonstrating a causal relationship between leader-employee interactions and a specific team norm. Furthermore, we present a practical and scalable intervention that organizations can readily implement to enhance psychological safety within their teams.

Our findings have significant implications for organizational practices. By recognizing the importance of fostering psychological safety and equipping leaders with the necessary skills to cultivate this norm, organizations can empower teams to share information more openly, and foster creativity, innovation, inclusion and ultimately, superior performance. A final noteworthy contribution of our paper is that it shows

the potential for impactful partnerships between firms and researchers. This collaboration successfully brought together behavioral practitioners from More Than Now, data scientists and health practitioners at Sandoz, and academic researchers from LMU and INSEAD. By bridging the gap between science and industry, this research not only advanced scientific understanding but also made it more accessible to organizations. The study was recognized with the European Compliance and Ethics Conference Best Project Award in 2022, and a practitioner-focused piece was co-authored to disseminate the findings to a broader audience (Rider et al., 2023).

3

Breaking the Silence

Group Discussions and the Adoption of Menstrual Health Technologies

Journal of Development Economics 169 (2024) 103264

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular Article

Breaking the silence – Group discussions and the adoption of menstrual health technologies

Silvia Castro *, Clarissa Mang

Ludwig-Maximilians-Universität München, Munich, Germany

ARTICLE INFO

Keywords:
Group discussions
Stigma
Social norms
Menstruation
RCT at the workplace

ABSTRACT

Stigma can hinder the adoption of beneficial and affordable technologies, particularly in sensitive health areas. Menstruation is a heavily stigmatized biological process, and managing menstruation with dignity and hygiene is a challenge in low-income settings. In this study, we conducted a randomized control trial to explore the impact of discussion-based interventions on breaking the silence around menstruation and shifting practices related to menstrual products. Our findings demonstrate a significant increase in the willingness to pay for well-known menstrual products and in the adoption of novel technologies post-intervention. The key driver of these outcomes is the reduction of menstruation-related stigma at the moment of the acquisition of the technologies.

1. Introduction

Social environments influence people's health decisions - for better and for worse. In particular, social stigma can prevent people from making optimal health choices, even when the optimal choice is readily available and affordable. For example, the fear of stigma may influence an individual's decision to undergo tests for sexually transmitted diseases (Yang et al., 2023), participate in preventative health check-ups (Ghosal et al., 2022), or seek assistance for mental health issues (Shidhaye and Kermode, 2013). This phenomenon is particularly acute in low-income settings, where despite significant investments in making basic health technologies available, the adoption rates are low. It is particularly concerning if stigma prevents vulnerable populations from utilizing simple and inexpensive health technologies.

Despite its frequency and ubiquity, menstruation is a biological process subject to strong levels of stigmatization. In many low-income settings, managing menstruation with dignity and hygiene remains a challenge, and the prevalence of unhygienic practices leads to serious health consequences (Torondel et al., 2018). Previously, it was believed that unsafe menstrual management stemmed entirely from a lack of information and resources. However, despite significant investments in education and access over the past decade, sanitary menstrual strategies are still far from being universally adopted. Moreover, menstruation continues to be stigmatized. Concealment surrounding anything related to periods prevents discussions even in private settings, such as at home between mothers and daughters.

Surveys of menstrual health in low-income contexts continue to show that repurposed menstrual cloth is the most frequent menstrual health management method. This method is not inherently a health hazard if individuals adhere to hygienic maintenance practices. However, the surveys reveal that women often do not wash or dry menstrual cloth properly (Bangladesh Bureau of Statistics, 2020). Trying to conceal the existence of menstruation, individuals avoid using available water sources, washing facilities, or even their home environment to clean their absorbents. Instead, they choose locations that offer privacy but may be unhygienic, like the floors of public toilets. Subsequently, they store these absorbents without drying to avoid displaying menstrual absorbents in public. These practices lead to unsafe materials being used in menstrual health management. Alternative products such as disposable pads would alleviate the washing and drying difficulties and

We are grateful to the editor Andrew Foster, and the two anonymous referees, to Emily-Au Young, Vojtech Bartos, Kristina Czura, Laura Derksen, Florian Englmaier, Svenia Frieß, Dominik Grothe, Lena Greska, Stephen Leider, Martina Miotto, Friederike Reichel, Simeon Schudy, Joachim Winter, and to seminar participants at LMU Munich, the University of Michigan, the University of Toronto, the Technical University of Munich, and the University of Vigo for many valuable comments. Financial support from GRK1928, Reemi, Elrha's Humanitarian Innovation Fund, and by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged. This RCT was registered as AEARCTR-0007163. Ethics approval was obtained from the Ethics Commission, Department of Economics, LMU.

* Corresponding author.
E-mail address: s.castro@lmu.de (S. Castro).

https://doi.org/10.1016/j.jdeveco.2024.103264
Received 6 March 2023; Received in revised form 25 January 2024; Accepted 28 January 2024
Available online 7 February 2024

eliminate risks associated with unsanitary cleaning processes. However, the question remains why women in these settings are rarely transitioning to these alternative health technologies. The most apparent reasons might seem to be cost and awareness. Yet, even in contexts where pads are both well-known and affordable, their adoption remains low. An underexplored factor is the purchasing process itself. Unlike in higher-income countries, pharmacies and stores in these settings are almost exclusively run by men, and additionally located in public spaces like village squares or busy urban streets. In societies where discussing menstrual health with men is taboo and women seek to conceal their menstruation, this scenario poses a considerable challenge. The stigma thus acts as a barrier to the purchasing process. Women face a trade off: the benefits of improved menstrual hygiene versus the discomfort and potential social cost of buying menstrual products in public, male-dominated environments.

In this paper, we empirically investigate the presence and implications of this trade-off through a randomized controlled trial involving female workers in a garment factory in Bangladesh. Bangladesh represents an ideal context for examining this dynamic. Over the past decade, numerous initiatives by both public and private entities have aimed to improve menstrual hygiene management. These efforts have ensured widespread availability and heightened awareness of disposable pads. Nonetheless, their actual usage remains modest.[1] Moreover, unhygienic menstrual practices are often associated with cloth usage due to social stigma.[2] We hypothesize that stigma significantly hinders the broad adoption of new menstrual products and practices.

We first present observational evidence from our sample highlighting that social concerns are present in women's decisions regarding disposable pad purchases. We surveyed women who exclusively use menstrual cloth to understand their reluctance to switch to pads. A substantial majority (85%) identified the presence of men in stores as a primary reason. Furthermore, we explored prevailing social norms around purchasing practices, finding that 60% of women believe that their peer group perceives buying pads from a male vendor as *socially inappropriate*. This leads us to hypothesize that directly addressing these social concerns may enhance women's adoption of desired menstrual products, even in situations where they must be purchased from a male vendor in a publicly observable location.

To test this hypothesis, we implemented a discussion-based intervention, following recent literature by Dhar et al. (2022) and Ghosal et al. (2022). Women in the treatment group engaged in a one hour discussion session, aimed at breaking the silence surrounding menstruation. These sessions, comprising groups of 15–20 women, encouraged openly sharing and discussing personal experiences and strategies about menstruation. This format was intended to expose the participants to diverse attitudes and perceptions, shedding light on a subject typically shrouded in taboo.

The study evaluates the impact of open discourse on the expected utility of menstrual products using two primary metrics. We first examine the valuation of sanitary napkins, a familiar product already in use by part of our sample. Second, we assess the adoption of a novel product unknown to our participants: an antibacterial reusable menstrual underwear. To reflect real-world conditions, the collection of these products was arranged at a convenience store located within the factory premises, managed by a male vendor and potentially visible to other factory workers. These outcomes are compared to a control group that did not participate in discussions. Additionally, to understand underlying mechanisms, we employed a discrete choice experiment as part of the study's endline survey, varying product's price, shopkeeper's gender, and purchase privacy.

The intervention increased the expected utility of menstrual products. Individuals in the treatment group were willing to pay 25% more than the control group for pads, compared to the control group's average valuation of 90 BDT (around 1 USD). This increase, which represents about half of the market price, indicates a substantial shift in how the participants value this well-known product. Moreover, the increased valuation was not limited to familiar products and it also influenced behavioral changes. The adoption rate for the antibacterial menstrual underwear increased by 14%, from a baseline of a 71% adoption rate in the control group.

We explore two potential channels that could be driving the observed effects: an information channel and a stigma reduction channel. The information channel would suggest that the group discussions, though devoid of formal external information, could have facilitated a form of social learning. Women who used cloth might have obtained insights about specific product features or the general benefits of sanitary napkins from their peers, thereby increasing their valuation of these products. On the other hand, the stigma reduction channel would mean that the intervention helped alleviate stigma surrounding menstruation. By openly discussing menstrual health, women might recognize a shared experience, reducing the perceived and experienced stigma and normalizing the purchase and use of pads. To differentiate between these channels, we conduct a discrete choice experiment that enables us to disentangle the valuation of the product itself from the influence of contextual factors (observability and shopkeeper gender). If the information channel were predominant, we would expect minimal differences between the treatment and the control group on their distaste for specific contextual factors and a large difference in the valuation of the product itself. However, our findings suggest otherwise. The treatment group demonstrates considerably less distaste about buying from male shopkeepers and in public settings. We found no differences in the intrinsic monetary value attributed to the product itself. This suggests that the main channel is a reduction in stigma-related concerns and not the dissemination of information about menstrual products or their features.

To corroborate this further, we conduct an exploratory analysis based on participants' baseline menstrual practices. We segment our willingness to pay metric by cloth users, pad users who typically purchase the pads themselves, and pad users who send someone else (usually their husband) to buy the pads for them. We find that the strongest treatment effect is seen for women who used pads but relied on their husbands to buy them. They show a significant 45% increase in willingness to pay, contrasting with a null effect of the intervention on increasing valuation among women who already purchased pads themselves. The knowledge of the product's features and benefits is similar in both subgroups; however, women who are already buying pads themselves seem to be less elastic to social stigma concerns. The marked difference in valuation among pad-using women, depending on who purchases them, emphasizes the intervention's role in addressing stigma, rather than product awareness.

Our results therefore point towards a plausible stigma reduction channel driving our findings. It is important to acknowledge the inherent complexity in isolating the specific mechanisms behind the observed changes in our study. While our results suggest that the primary driver of these changes is a reduction in stigma, we cannot categorically rule out the influence of an informational channel. The intervention, designed to encourage open discussions about menstrual health, inherently blends informational exchange with stigma reduction. Providing information about menstrual health and products is intrinsically linked to the social stigma associated with these topics. In

[1] Within our sample, 90% of the women reported the availability of a nearby store selling sanitary pads. The Bangladesh National Hygiene Survey (2014) observed that in 2013–2014, approximately 33% of urban women used sanitary pads. The latest survey data indicates some progress among adolescent girls, but minimal change among adult women, with 64% using cloth for menstrual hygiene (Bangladesh Bureau of Statistics, 2020).

[2] As per Bangladesh Bureau of Statistics (2020), less than a third of women manage to wash and clean their menstrual cloth hygienically, and 40% store it immediately after washing, avoiding drying to prevent the display of menstrual cloth.

the absence of distinct treatment arms that could separately evaluate the effects of information provision and stigma reduction, our study cannot perfectly disentangle these intertwined channels.

With this field experiment, we contribute to the growing body of literature on three separate but closely intertwined approaches to advance health- and productivity-enhancing behavior, especially of women in low-income contexts. First, many papers have sought to directly affect the perception of social norms. This literature on social norms builds on the seminal works by Bicchieri and Dimant (2019) and Krupka and Weber (2013), who have shaped the debate by providing concise and actionable definitions and ways to measure social norms. Addressing the perception of social norms usually takes one of two forms, a norm correction strategy or a norm transformation strategy (Cislaghi and Berkowitz, 2021). Researchers using the first strategy correct misperceptions by providing factual information about others' actual behaviors and beliefs about various social norms, for example regarding female labor force participation (Bursztyn et al., 2020), savings decisions (Dur et al., 2021), energy consumption (Allcott, 2011), and salary disclosure (Cullen and Perez-Truglia, 2018, 2022). On the other hand, projects applying a norm transformation strategy often use media such as TV shows (Jensen and Oster, 2009; La Ferrara et al., 2012; Banerjee et al., 2019; Green et al., 2020) and radio shows (Paluck, 2009; Arias, 2019) to influence the perception of social norms. Our study design is more in line with the second approach, but we do not actively attempt to influence the perceived social norms in any direction.

Second, a range of interventions has sought to directly address personal attitudes toward certain (health) practices and behaviors, such as open defecation (Gauri et al., 2018) and intimate partner violence (Gupta et al., 2013; Abramsky et al., 2014; Pulerwitz et al., 2015). These studies usually use a mixture of information campaigns, direct education, and group discussions to achieve the change in personal attitudes. To address attitudes on gender equality in particular, some studies have shown that exposure to women in male-dominated areas, such as the military (Dahl et al., 2020) or local politics (Beaman et al., 2009) can successfully change attitudes toward gender equality rooted in traditional gender norms.

Lastly, we add to the literature on female (menstrual) health as an important aspect of public health provision and an important contributing factor in female labor force participation, productivity and human capital accumulation. We build on the previous literature that focuses on improving the affordability and access to pads, including Garikipati and Boudot (2017); Krenz and Strulik (2019); Czura et al. (2024), and to alternative products like menstrual cups, such as Oster and Thornton (2011). We extend the literature that looks at the role of information and social norms (Castro and Czura, 2021; Czura et al., 2024) by directly addressing the role that stigma plays in hindering access to improved menstrual products, which has limited the success of many previous projects without having been explicitly addressed.

This paper is structured as follows: In Section 2, we describe the background of the study and present survey evidence for the important role of stigma in hindering the adoption of safer menstrual health management practices. Section 3 details the experiment design and our empirical strategy. Section 4 lays out the theory of change and hypotheses. In Section 5, we present and discuss the results. Section 6 discusses the implications of the study, and Section 7 concludes. The paper closes with Section 8, which discusses various robustness checks.

2. Menstrual hygiene and stigma in Bangladesh

Menstrual hygiene is a key element for the physical, mental, and emotional well-being of women (Torondel et al., 2018; Benshaul-Tolonen et al., 2021). It aids their economic prospects by mitigating barriers to education (Agarwal et al., 2022) and employment (Krenz and Strulik, 2019).[3] Considering that menstruation affects approximately half of the global population for a significant portion of their adult lives, and that across many contexts, managing menstruation with dignity and hygiene remains a significant challenge, advancements in menstrual hygiene management yield substantial economic and humanitarian benefits. Consequently, improving menstrual hygiene has become a focal point in international development initiatives and has spurred a growing body of research dedicated to understanding and ameliorating poor menstrual hygiene. However, maintaining menstrual hygiene remains a challenge in many low-income contexts (Garg et al., 2012; Garikipati and Boudot, 2017; UNICEF, 2019; Czura et al., 2024). A major obstacle to sustainable improvements in menstrual health practices is the presence of cultural taboos and stigma surrounding menstruation (Castro and Czura, 2021).

In low-income countries, cloth is the primary material used for managing menstruation. In Bangladesh, approximately 65% of adult women rely on cloth, often repurposed from an old saree or similar materials (Bangladesh Bureau of Statistics, 2020). These women frequently lack access to private sanitation facilities, hindering their ability to change the cloth regularly, particularly during work hours. Compounding this issue, many women also do not have access to clean water or private spaces to properly wash the used cloth with soap. Instead, they often resort to using unhygienic but private spaces, such as the floors of public toilets, for washing these cloths. Commonly, the washed menstrual cloths are stored immediately without adequate drying, tucked away under mattresses or in cupboards. Such practices pose direct health risks, including urinary tract infections and inflammations, due to the unsanitary conditions of storage and use (Sumpter and Torondel, 2013; Torondel et al., 2018).

Public and private campaigns have worked towards introducing menstrual absorbents that circumvent the limitations of cloth, such as disposable sanitary napkins. The advantage of disposable pads lies in their elimination of the health risks associated with the improper washing and drying of cloth. Despite their widespread availability in Bangladesh, however, adoption rates of disposable pads are modest. Only about 29% of adult women—and 43% of adolescents under 19—report regular use of pads (Bangladesh Bureau of Statistics, 2020).

In the sample of Bangladeshi garment factory workers recruited for this study, 40.5% of women do not use pads as their primary absorbent. Notably, availability is seldomly cited as a barrier to adoption, with 79% reporting easy access to a store selling pads. Instead, the shame, and stigma associated with purchasing pads are the predominant reported obstacles. More than 80% of surveyed women report experiencing discomfort when buying pads due to privacy concerns and the anxiety of being seen, especially when the seller is a man (Table A1 in the online appendix). Additionally, even those who use pads regularly report fear of stigma when acquiring the products. In our sample, 52% of regular pad purchasers resort to covering their faces during purchase to avoid recognition. These survey results underscore that stigma matters for women's access to for-sale menstrual hygiene products. Consequently, our study evaluates an intervention designed to alleviate these social constraints by diminishing the perceived stigma surrounding menstruation.

In recent years, paralleling our research, there has been a significant

[3] A study by the Water Supply Sanitation Collaborative Council (WSSCC) in Bangladesh suggests that an infection caused by using cloth during menstruation leads to 73 percent of women missing work for an average of 6 days a month (WSSCC, 2013).

increase in interventions aimed at addressing the stigma surrounding menstruation by public practitioners and private stackeholders. These efforts have evolved from merely providing information or subsidizing products to actively tackling the stigma and taboos associated with the topic. A robust global movement, symbolized by the hashtag #LetsTalkPERIOD(s), has emerged to normalize menstruation and promote menstrual health. This shift has witnessed considerable progress worldwide in addressing the impact of stigma and taboos on menstrual health.

Public and private entities have adopted creative strategies to break the silence around menstruation, involving collaborations with sports teams and partnerships with influencers. Notable examples include the German Women's football team partnering with o. b. tampons during the 2023 Word Cup and the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) using social media and local influencers to reshape discussions in Nepal, Albania, and the Philippines. Additionally, the MenstruAction movement, supported by the German development cooperation, aims to initiate dialogues about menstruation and women's and girls' menstrual needs and rights. This paper seeks to speak to the potential of such campaigns to effectively reduce stigma and achieve lasting change in menstrual health practices.

A distinctive aspect of our approach involves the deployment of two different types of menstrual absorbents. One is the well-established and widely available disposable pads (or sanitary napkins). The other is a novel product not yet available on the market at the time of the study — reusable menstrual underwear. The menstrual underwear, developed and supplied by our project partner Reemi, represents a culturally sensitive and innovative alternative to menstrual cloth or disposable pads. This alternative addresses several cultural, social, and health issues associated with existing methods. Its primary benefit lies in its superior absorbency compared to sanitary pads or cloth, reducing the need for frequent changes throughout the day. Furthermore, it represents a one-time investment, offering long-term use over several years and eliminating the need to repeatedly interact with male shopkeepers. While this underwear also requires washing with soap and drying — a common challenge — the material is fast-drying and anti-bacterial, facilitating easier cleaning and reducing infection risks. As a new product in Bangladesh, the menstrual underwear was unfamiliar to the women in our study. However, given its design tailored to the needs of menstruators like our participants, we anticipated a strong initial demand. The novelty of the product ensured a clear delineation of our treatment effects, as the participants had no prior access to it in the market. This exclusivity allowed us to confidently attribute any observed outcomes directly to our intervention, without the confounding effects of external access to the product.

3. Experiment design

3.1. Sample

Our field experiment was conducted in a large garment factory in Tongi, a town north of Dhaka, Bangladesh. From a pool of 6,000 workers, 600 female employees were randomly chosen to participate in our study, based on a list provided by the factory. The selected participants were contacted on their mobile phones after work hours. Upon obtaining their consent, we administered the baseline survey, continuing our outreach until 485 women agreed to participate. These participants had experienced regular menstruation in the past six months, except for 16 women who reported currently being pregnant. The baseline surveys were conducted in March and April 2021. For their participation, each respondent received 40 BDT in phone credits, approximately equivalent to 0.50 USD and corresponding to the hourly wage rate. Female enumerators conducted the phone surveys to minimize any discomfort participants might feel when discussing menstruation-related topics.

After completing the baseline survey, participants were randomly assigned to either the treatment or control group. The treatment commenced following the completion of the first 100 baseline surveys, thus reducing the time gap between the baseline assessment and the treatment sessions for the treatment group and streamlining the logistical aspects of the study. The treatment group consisted of 227 women, the control group of 258 women. All women assigned to the treatment group participated in the treatment sessions.

The endline survey was conducted with all participants in April and May 2021. Attrition rates were minimal and comparable across both groups: 1.8% in the treatment group (4 out of 227 women) and 1.9% in the control group (5 out of 258 women). The primary reason for attrition was unresponsiveness or switched-off phones during the endline survey calls. Our final sample size for the main analysis comprised 476 women, with 223 in the treatment group and 253 in the control group. Approximately six months post-treatment, in November and December 2021, we re-surveyed 339 women from our original sample (182 from the control group and 157 from the treatment group) to assess the persistence of the effects. The timeline of data collection and the specific measures collected at each stage are graphically summarized in Figures A1, A2 and A3 in the online appendix.

Our randomization process successfully resulted in balanced samples in terms of observable characteristics (see Table A1 in the online appendix). The average age of women in our sample was 26 years. Most were married and had at least one child. Their education level, averaging seven years, was slightly above the national average for women, reflecting the growing trend of young women seeking employment in garment factories after attaining higher education levels (Asadullah et al., 2021; United Nations Development Program, 2022).

The relatively young age and higher education levels in our sample are positively correlated with pad usage.[4] While 60% of our sample reported frequent pad use at baseline, surpassing the national average (Bangladesh Bureau of Statistics, 2020), half also reported frequent use of cloth, indicating that some women use both (e.g., pads on heavier flow days, cloth on lighter days), with a significant proportion not using pads at all.

3.2. Treatment intervention: open discussions about menstruation at the workplace

Our intervention is centered around a simple group discussion. Social interventions aimed at modifying individual behavior often use what Cislaghi and Berkowitz (2021) describe as norm correction strategies. They provide individuals with factual information about the actions and attitudes of others to correct misperceptions (Allcott, 2011; Bursztyn et al., 2020; Dur et al., 2021). In contrast, social psychology perceives social norms not as static beliefs but as part of a dynamic group process (Prentice and Paluck, 2020). In this view, individuals interpret social information within the context of a group, seeking validation and agreement from other group members in their responses. Instead of attempting to alter an individual's perception of group norms in isolation, our intervention seeks to provide an opportunity for real-time validation of social perceptions. This approach aligns with the methods used by Dhar et al. (2022) and Ghosal et al. (2022), who employed group discussions to directly address participants' personal attitudes and perceived (self-)stigma.

The intervention consisted of a one hour discussion session, encouraging participants to openly share their thoughts and experiences

[4] Pearson's correlation coefficient between age and cloth use: 0.18, p-value: 0.00 and between age and pad use: −0.19, p-value: 0.00; younger women tend to use pads more, older women cloth. Pearson's correlation coefficient between education and pad use: 0.19, p-value: 0.00 and between education and cloth use: −0.23, p-value: 0.00; more educated women tend to use pads more, less educated women cloth. See Appendix Figure A4 for a graphical analysis of these trends.

about menstruation. The sessions were led by trained female facilitators. These discussions were intentionally not structured as educational or training sessions. Our aim was to explore the impact of normalizing the topic, not of providing external or formal information. Thus, the discussions were geared towards sharing personal experiences, with minimal information provided by the facilitators, diverging from previous studies that focused on empowering young women and girls through external information provision or specific life skills training (Duflo et al., 2015; Ashraf et al., 2020; Bandiera et al., 2020; Buchmann et al., 2021; Castro and Czura, 2021). Instead, our intervention allowed for the endogenous influence of group feedback on the participants' perceptions. It enabled them to update their beliefs based on the responses they received from other group members, without external validation of these updated beliefs from the experimenters. By offering a positive experience of discussing menstruation, the sessions aimed to alter the women's confidence in talking about the subject, reduce the associated taboo, and lessen the stigma. In contrast, the control group did not engage in any discussions and had no opportunity to openly discuss menstruation. Their interaction with the study was limited to the baseline and endline phone surveys.

The discussion sessions were held during work hours in a conference room at the factory and were moderated by two female facilitators from the implementation partner, Change Associates Ltd.[5] The sessions occurred in March and April 2021, with a total of 15 sessions conducted. Each session had an average of 15 participants (ranging from 13 to 21) and lasted one hour. The format was hybrid, with factory workers present in the conference room and facilitators joining remotely via Google Meet. Post-session, moderators completed surveys to report any incidents, main discussion topics, questions raised, and the overall atmosphere and participation level.

Moderators reported that sessions covered similar topics, including first experiences with menstruation, discomfort during menstruation, discussions about menstruation with children, and the pros and cons of various menstrual products, especially pads. While these core topics were consistent, the emphasis varied, with some sessions focusing more on menstrual products in general, others on pads in particular, and others on how to discuss menstruation with children. The women attending the discussions exchanged personal experiences and the group *collectively* did not receive any new information, but current knowledge and experiences were shared within the group.

The post-session feedback indicated minimal technical difficulties, with only 2 of 15 sessions experiencing issues that were quickly resolved. Facilitators unanimously reported enthusiastic participation from the women, with nearly all sessions showing equal engagement from all attendees. This confirms that the intervention was implemented as intended and achieved its objective of fostering open discussions about menstruation in a supportive group setting.

3.3. Outcome variables

3.3.1. Valuation of a well-known menstrual product: disposable pads

Our primary experimental outcome is the women's willingness to pay (WTP) for a familiar menstrual product: disposable pads, also known as sanitary napkins. We specifically measured willingness to pay in a context that mirrors the real-world purchasing environment in Bangladesh, where the majority of vendors are male, and pads are sold in public locations like convenience stores or pharmacies located in busy urban streets or in villague squares. This setup is key to our study as it integrates the public nature of the purchase and the male gender of the seller, which may influence women's decisions to adopt for-sale menstrual technologies due to potential discomfort in such purchasing scenarios.

We measured the willingness to pay using a price list (Anderson et al., 2007), with enumerators describing the conditions under which the menstrual products need to be collected at the factory store. Participants were offered a choice between receiving an amount of money (in phone credits) or the product for free, with the price increasing in fixed intervals. This approach allowes us to capture the willigness to pay as an interval between the last price at which the product was chosen and the first higher price at which the money was preferred. This methodology assumes monotonically increasing preferences with a single switching point.[6] For completeness, we also measured the willigness to pay for the menstrual underwear, but the novelty of the product, participants' unfamiliarity with it, and the elicitation of the metric through a phone survey, was likely to yield noisy estimates; therefore, we focus on the willigness to pay of the known-product. Results for the underwear WTP can however be found in Table R6 in Section 8 of the robustness checks.

The WTP measurements for the disposable pads and the underwear were incentivized together. For each participant, one randomly selected choice was made payoff-relevant. Consequently, the woman received either a specific amount of money or the opportunity to collect the product (either the disposable pads or the menstrual underwear) based on her choice in the randomly selected scenario. The participants could only receive either the pads or the underwear, but not both. The women knew that only one of the choices they made between money and either of the products would be realised. To increase the power of our second outcome metric regarding the adoption of a new technology, we skewed the randomization of the payoff-relevant outcome in such a way that for 95% of the women, the choice between 0 BDT and the underwear was selected to be payoff-relevant.[7] Therefore, the majority of women were eligible to pick up the underwear for free. For seven women, a different pay-off relevant scenario was randomly selected and they received either an amount of money or a pack of pads.

3.3.2. Adoption of a novel menstrual technology: antibacterial underwear

The second outcome of our study focuses on the adoption rate of an innovative and previously unavailable product: an antibacterial reusable menstrual underwear. This product's introduction serves as a key component of our research, allowing us to measure the uptake of a novel technology in menstrual hygiene not available outside of our study.[8] During the endline survey, we carefully explained the unique characteristics and benefits of the reusable menstrual underwear to the participants. They were informed that these innovative menstrual products would be available for collection at the factory store during work hours and that they would be dispensed by the male vendor that works there. In total, 469 women from our study sample became eligible to receive the underwear. The distribution of the underwear, scheduled for June

[5] Change Associates Ltd. is a women-led organization in Bangladesh that frequently conducts trainings on health and family planning topics in garment factories.

[6] The first choice is between receiving 0 BDT or getting the product for free. Conditional on the women selecting to receive the product, the offered price is then increased in fixed intervals and the participants are asked to make the choice again between the higher amount of money and the product. This was done in steps of 20 BDT up to 140 BDT and then a jump to a maximum price of 200 BDT (around 2 EUR, or four times the market price of pads). The jump in the interval enabled us to check a very high WTP, while keeping the number of questions asked to a minimum to limit complexity. The WTP is thus recorded as an interval between a lower bound (last price at which the product was chosen) and an upper bound (first price at which the money was chosen).

[7] The women were informed that one of their decisions across both WTP exercises would be pay-off relevant, but not how this was chosen.

[8] The menstrual underwear used in this study was developed and produced by Reemi, a New Zealand-based NGO. Designed with multiple leak-proof layers on the exterior and an anti-bacterial absorbent layer on the interior, this product represents a significant advancement in menstrual hygiene technology. As of the date of our study, such reusable menstrual underwear was a novel concept and not commercially available in Bangladesh.

2021, was announced via phone call to the participants, and they had a 10-day window to collect the product from the factory store.

This outcome metric is of relevance for several reasons. First, it allows us to assess whether any observed changes in valuation or attitudes towards menstrual products also translate into tangible behavioral changes, such as picking up an available menstrual product. Second, it addresses a key limitation of measuring behavioral impact solely for existing products like sanitary napkins. As pads are readily available in the market, any influence of our intervention on participants' market behaviors, such as an increased propensity to purchase pads, could remain unmeasured within the scope of our study. For example, women might alter their behavior and start buying products at convenience stores near their homes; such changed behavior would go undetected, as these products are available outside of the study. However, by introducing a product exclusively available through the study, like the reusable menstrual underwear, we can more accurately capture the direct impact of the group discussion. In addition, it allows us to shed some light on the underlying channel, because it limits the potential of a pure information channel: if the product is new, none of the women have prior experiences they can share about it and it cannot be discussed during the treatment sessions. Any observed changes in the adoption of a novel product are therefore highly unlikely to be driven through changes in information alone.

3.3.3. Understanding the valuation dynamics through a discrete choice experiment

Discrete choice experiments (DCE) are employed to unravel the value customers assign to different product features. This method involves presenting customers with a series of choices between two sets of product characteristics (e.g., price, color, size). We adopted this approach to separate the direct monetary value attributed to the product from the various contextual factors relating to how the product is obtained. We presented participants with (hypothetical) scenarios for acquiring disposable pads, varying the purchase's visibility (publicly on the factory premises vs. an external shop), the shopkeeper's gender (male vs. female), and price levels (30, 40, 50 or 60 BDT). Women made choices between two bundles of these attributes in consecutive scenarios, providing insights into the relative utility derived from each attribute and their willingness to pay a premium for facing preferred conditions. Detailed theoretical details on how we constructed the choice sets are provided in the online appendix in Subsection B.

The rationale for integrating this discrete choice experiment into our study was to disentangle the information and stigma channels as drivers of our results. Creating separate treatment arms to isolate these factors is exceptionally complex. Discussions, information provision, or product distribution might affect both information levels and stigma. As detailed in subsection 3.2, our intervention was consciously structured to primarily address the stigma associated with menstruation without introducing external information. Despite this, participants could still exchange knowledge about product features during the discussion sessions, contributing not only to the topic's normalization but also to increased knowledge of some participants. The discrete choice experiment allows us to explicitly separate the value attributed to contextual factors, i.e. purchasing the product from a male vendor and picking it up in a more public place, from the value attributed to the product itself. While the willingness to pay encompasses both the valuation of the product itself and the context of acquisition together, the discrete choice experiment disentangles these different dimensions.

3.3.4. Supplementary survey measures as corroborative evidence

To complement the discrete choice experiment, we collected additional survey measures to support and contextualize the findings derived. These measures focus on perceived social norms and stigma and will be used as supportive evidence to underscore our main outcomes.

Social norms, the informal rules indicating socially acceptable actions, consist of empirical expectations about others' actual behavior (descriptive norms) and normative expectations about others' perceptions and beliefs (injunctive norms) (Bicchieri and Dimant, 2019). In our study, we specifically elicited an injunctive social norm related to the purchase of menstrual products from a male vendor. To capture societal perspectives rather than personal feelings towards this norm, we utilized a vignette study. Participants were presented with the scenario featuring a woman similar to themselves during menstruation and were asked about the anticipated reaction of the woman's neighbors to this woman buying pads from a male vendor. For this scenario, respondents indicated whether they expected neighbors to find the behavior very socially inappropriate, socially inappropriate, socially appropriate, or very socially appropriate.

Additionally, we assessed changes in perceived stigma. Beyond influencing second-order beliefs about others' perceptions, we anticipated that the discussions would impact feelings of shame, embarrassment, and stigma related to menstruation. To gauge perceived stigma, participants were asked about their agreement with a set of four statements, each expressing a different aspect of stigma, such as concern over being treated differently if their menstrual status was known *(i.e. i) women should hide any evidence of menstruation, ii) if someone would know that I am menstruating they might treat me or look at me differently*).[9] Rather than reporting their level of agreement with each statement individually, participants reported the total number of statements they agreed with. Our scale from 0 to 4 reflects the intensity of perceived stigma.

Alongside these measures, demographic variables were collected including age, religion, marital status, number of children, and frequency of menstrual product use at baseline. The exact survey questions for baseline and endline are available in the online appendix.

4. Theory of change and hypotheses

Our study's theory of change posits that open discussions about menstruation in peer settings can significantly alter individual perceptions and attitudes towards menstrual health and the expected utility of menstrual products, resulting in lasting behavioral change. We hypothesize that there are two potential channels driving the transformative effect.

The first channel involves reducing the social stigma surrounding menstruation. We postulate that group discussions foster comfort and openness in talking about menstruation among participants. This leads to a reshaping of both personal and collective norms regarding menstruation, enabling women to perceive menstruation as a natural process and normalizing the act of purchasing menstrual products. The stigma associated with menstruation, often deeply rooted in cultural and societal norms, extends beyond personal discomfort to reflect broader societal attitudes that treat menstruation as a shameful and secretive matter. The intervention provides participants with a safe space to openly discuss menstruation, thus normalizing the subject. This may shift perceptions about others' disapproval regarding buying and using these products, as well as reduce personal feelings of shame when doing so. This can help normalize purchasing behaviors, even if some discomfort persists in interactions with male vendors.

The discourse generated in the group discussions is therefore expected to mitigate the discomfort associated with menstruation and menstrual health management and reduce the stigma. This should manifest in reduced concerns about buying products either from male shopkeepers or in public, or both, and result in a higher expected utility of the product, as proxied by the valuation measured as the willingness to pay.

Hypothesis 1. *Participation in discussion sessions about menstruation leads to increased expected utility of period-specific menstrual products,*

[9] These statements were adapted from various surveys, as detailed in Hennegan et al. (2020).

driven by a decrease in the social stigma associated with...

Hypothesis 1a. *... purchasing menstrual products in public.*

Hypothesis 1b. *... purchasing menstrual products from a male vendor.*

To compare the level of discomfort related to the shopkeeper's gender and the public nature of the purchase between control and treatment groups, we use the results from the discrete choice experiment. To confirm this hypothesis and its sub-hypotheses, we would need to observe reduced discomfort for the treatment group associated with purchasing the pads in a more public setting (inside the factory) (to confirm Hypothesis 1a), and a reduced discomfort associated with purchasing the pads from a male vendor (to confirm Hypothesis 1b). In other words, the control group needs to be willing to pay a higher premium to avoid a male vendor and to avoid a more public place compared to the treatment group.

The second channel through which the group discussion could increase the valuation of menstrual products is through an information and awareness channel. Participants could share and learn about various aspects of menstrual products from their peers. The discussions might reveal less-known information about product varieties, usage techniques, disposal methods, and even cost-effective purchasing options. Such detailed knowledge can shift the participants' valuation from a purely cost-based perspective to a value-based one, where they appreciate the full spectrum of benefits offered by these products. Consequently, we would observe an increase in the value attributed to this product, completely unrelated to stigma or the social dimension of the purchase, purely driven by this enriched awareness and understanding.

Hypothesis 2. *Participation in discussion sessions about menstruation leads to an increased expected utility of period-specific menstrual products, driven by an enhanced understanding of the products' features and benefits, acquired through interactive peer discussions. This elevated awareness influences the perceived monetary value of menstrual products, independently of any changes to the social dimension of the purchase.*

The DCE allows us to compare the monetary valuation attributed to disposable pads independently of the contextual factors by the treatment and control group. To confirm this hypothesis, we would need to observe a greater intrinsic value attributed to the diposable pads by the treatment group, such that the treatment group would be willing to pay a higher price for the pads regardless of contextual factors, i.e. holding the vendor's gender and the pickup location constant.

Hypotheses 1 and 2 are not mutually exclusive, and may in fact very well be true at the same time. The intervention can lead to a decrease in stigma surrounding menstruation, and also increased awareness of the characteristics of menstrual products simultaneously. This outcome would be evident from observing an increase in the monetary value of menstrual products and a decrease in the social stigma surrounding menstruation as a result of the intervention. We would expect statistically significant differences in all measured attributes of the discrete choice experiment between the treatment and control groups. This would confirm both Hypotheses 1 and 2, demonstrating that the intervention operates effectively through two simultaneous channels.

5. Results: the effects of discussing menstruation openly

In this section, we present our experimental findings. The results show a notable increase in the valuation of period-specific menstrual products among women who participated in the discussion sessions. Alongside this, we observe a tangible behavioral change: women in the treatment group have a larger propensity to adopt a new menstrual technology. Additional metrics shed light on the underlying mechanisms, suggesting that the changes observed are predominantly influenced by a decrease in purchasing-related stigma and not by increased awareness of the products gained through the discussions. Notably, women in the treatment group show less apprehension regarding the public nature of their purchases and the male gender of the vendor.

5.1. Collective discussions: a pathway to greater menstrual health technology adoption

In this section, we show how group discussions influence participants' valuation of menstrual health products and their adoption rates. We compare the willigness to pay for sanitary napkins and the collection rates for menstrual underwear between the control and treatment groups. We employ an interval regression of the willigness to pay on the intervention dummy, and use a linear probability model to regress the pick-up rates of menstrual underwear on the intervention dummy. Table 1 presents these findings, with Column (1) detailing WTP results and Column (2) showing pick-up rates.

The results in Column (1) of Table 1 show that women in the control group have an average willigness to pay of approximately 91 BDT for a pack of sanitary napkins. In comparison, the treatment group exhibits an average increase of about 23 BDT, exceeding in 25% the control mean. This significant difference at the 5% level, combined with the substantial magnitude of the increase, indicates a considerable treatment effect.[10] These results remain robust when including demographic controls, controls regarding menstrual practices, and also stigma and social norm controls (see Table R1 in Section 8 on Robustness Checks).

The observed average increase in WTP of over 20 BDT implies a significant shift in the treatment group's valuation, moving up to the next WTP interval. An analysis of the WTP distribution shows how women in different valuation intervals reacted to the intervention. Furthermore, it provides insights into whether the responses of women with valuations above the market price differed from those with valuations below it. As illustrated in Fig. 1, the impact of the intervention was consistent across the entire distribution. For every lower bound value of WTP (the maximum amount where a woman preferred pads over money), the cumulative distribution function for the treatment group consistently falls below that of the control group, indicating a comprehensive shift in valuation. The treatment group's WTP distribution exhibits first-order stochastic dominance over the control group. Up until the 80–100 BDT range, the control group sees larger increments, signifying a greater proportion of women at each lower valuation interval. The analysis reveals no significant difference in responses between women whose valuation of pads was above or below the market price.

Column (2) of Table 1 reveals a powerful behavioral impact of the

Table 1
Effect on valuation and adoption of menstrual products.

	(1)	(2)
	WTP sanitary napkins	Pick-up reusable underwear
Intervention	22.982**	0.099**
	(8.98)	(0.04)
Mean of dep. var	90.620	0.713
Observations	476	469

Notes: Column (1) reports interval regression estimates for WTP (in BDT) for sanitary napkins. Column (2) presents linear probability regression (OLS) results for the collection of underwear at the factory store from a male shopkeeper. *Mean of dep.var.* indicates the control group's results. Robust standard errors in parentheses. Stars denote significance levels: **p < 0.05.

[10] This effect is noteworthy, especially given the already high baseline WTP in our sample, which is 50–100% higher than the market price of pads. This might stem from several factors, including perceived quality differences in pads provided in the market compared to our study, the experimental framing of our WTP elicitation, and the unique financial autonomy provided to the women in our study compared to their typical household budgetary constraints.

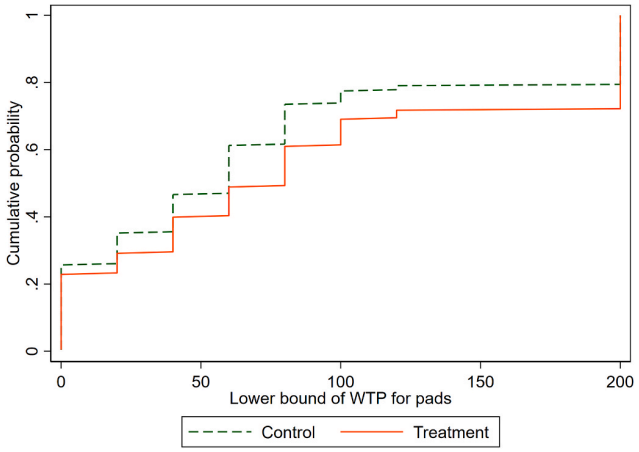


Fig. 1. Cumulative Distribution of the Willingness to Pay.
Notes: The figure shows the cumulative distribution function of the share of participants indicating each lower bound for the willingness to pay for sanitary pads. WTP was elicited in 20 BDT intervals, ranging from 0 to 120 BDT and at 200 BDT.

intervention. The treatment led to a 14% increase in the collection rates of antibacterial menstrual underwear. This increase in collection rates is both significant and sizeable, with about 71% of women in the control group and 81% in the treatment group picking up the product.

Fig. 2 graphically represents the proportion of women in the treatment and control groups who collected the underwear. The cumulative distribution function illustrates that the treatment group's collection rates were consistently higher than the control group's.

Our results are not driven by a specific combination of group characteristics or by some group discussions where the intervention was exceptionally effective. To substantiate this, we conduct additional analysis of group composition and treatment effects, detailed in subsection 8.2 on Robustness Checks. Our findings indicate that the precise composition and characteristics of the discussion groups do not play a pivotal role in determining the effectiveness of the treatment. Furthermore, we demonstrate the robustness of our results by excluding potential outliers (Table R3) and by incorporating enumerator fixed effects (Table R2).

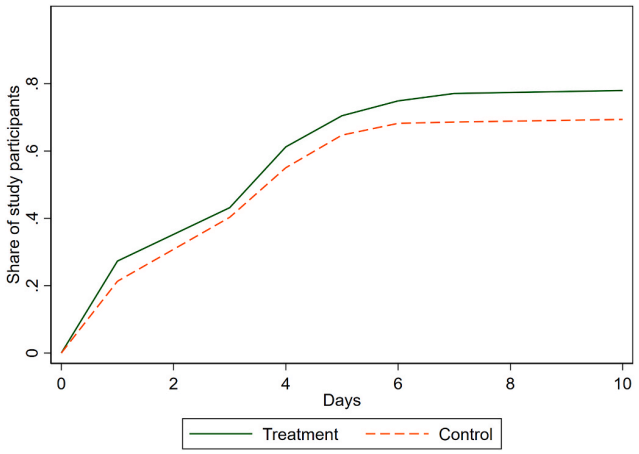


Fig. 2. Collection of underwear over time.
Notes: Proportion of participants in the treatment and control groups collecting the menstrual underwear at the factory store from a male shopkeeper. Availability period: 10th-19th June 2021.

5.2. Understanding the drivers: a reduction in purchase-related concerns

In this section, we explore the mechanisms driving the observed changes in product valuation and adoption. Specifically, our analysis seeks to differentiate between two key mechanisms: information transmission and stigma reduction. The first mechanism, information transmission, hypothesizes that the group discussions provided a platform for women to share knowledge about menstrual products, particularly sanitary napkins. This mechanism suggests that women who were previously unaware or had limited knowledge about the benefits and features of these products gained new information through peer interactions. This enhanced understanding, in turn, could lead to a higher expected utility from these menstrual products.

Conversely, the second mechanism, stigma reduction, posits that the discussions helped in breaking down the social barriers and stigmas associated with menstruation. This mechanism is grounded in the idea that open conversations can normalize the topic, reduce feelings of embarrassment or shame, and in turn, influence women's behaviors towards accessing period-specif products, especially in public or male-dominated spaces.

The information channel is a plausible explanation for the increase in valuation of disposable pads. Although all women in the factory know about the existence of disposable pads, those who use them regularly might have enhanced knowledge about their features, a level of understanding that women who exclusively use cloth may not possess and this information could flow from one to another in the discussions. In the case of menstrual underwear, an information channel explaining the differences in adoption rates is less plausible. None of the women had any prior knowledge of this product; it is a novel item they had neither seen nor used before. The underwear was only introduced and mentioned to the participants after the discussion sessions, meaning its features and attributes were not part of the discussion.

However, this argument alone cannot completely rule out an information channel. Participants might have acquired new information about menstruation in general, such as the negative health consequences of using unhygienic cloth, which may influence their decision to switch from cloth to other products. Given that the adoption of menstrual underwear presents a less clear avenue with limited scope for detecting information channels, we focus our analysis on the increased valuation of sanitary napkins and turn to the results of the discrete choice experiment.

5.2.1. Discrete choice experiment: a shift in attitudes towards public purchasing and male interactions

In our study, the discrete choice experiment plays a key role in quantitatively disentangling the multifaceted aspects of product valuation. Specifically, the DCE allows to independently measure three key dimensions: the role of the vendor's gender, the role of the public visibility of the purchase, and the intrinsic valuation of the menstrual products on itself. Our previous measure of willingness to pay inherently combined these factors, making it challenging to isolate the direct value of the product from the conditions under which it was purchased. The DCE, by providing exogenous variation in these attributes, enables us to parse out the individual contribution of each aspect to the overall product valuation. The DCE thus serves as a tool to test the hypotheses and theories of change outlined in Section 4. If we observe a change in the value attributed to shopkeeper gender and/or purchase location, this provides compelling evidence for hypothesis 1. If we observe a change in the value attributed to the product itself, this provides compelling evidence for hypothesis 2. If we observe a change in all three dimensions, then both 1 and 2 would be correct at the same time. Therefore, the DCE will allow us to determine if the intervention's impact was predominantly driven by alleviating social stigmas, fostering informational exchanges among participants, or a combination of both.

Table 2 presents the findings from the conditional logit model. In this model, the coefficients signify the change in the log odds of selecting a

Table 2
Discrete choice experiment - conditional logit model.

	(1)	(2)
	Utility level	
Location inside	−0.384***	−0.592***
	(0.09)	(0.14)
Male vendor	−1.452***	−1.753***
	(0.07)	(0.10)
Price	−0.154***	−0.168***
	(0.01)	(0.01)
Intervention*Location inside		0.396**
		(0.18)
Intervention*Male vendor		0.590***
		(0.14)
Intervention*Price		0.023
		(0.02)
Observations	3,808	3,808
Participants	476	476

Notes. This table presents the results of a conditional logit regression analysis from a discrete choice experiment, focusing on the utility of sanitary pads. Column (1) displays the basic model, assessing the utility change associated with the product's location of collection (inside the factory vs. outside), the gender of the shopkeeper (male vs. female), and the price. Column (2) extends the model to include interaction effects between these attributes and the intervention. The dependent variable is the stated utility for each choice set, captured in a binary outcome (choice/non-choice) for each option. Standard errors, reported in parentheses and clustered at the individual level. Stars denote significance levels: *p < 0.10, **p < 0.05, ***p < 0.01.

particular option, which in our case relates to the utility derived from various conditions of menstrual product acquisition. The coefficients, while not directly interpretable in their magnitudes, provide essential insights through their signs and relative sizes. Column (1) shows that attributes like higher prices, male vendors, and less private collection locations are associated with a decrease in utility, with the effect being most substantial for the vendor's gender. This indicates a notable general aversion among participants to purchasing from male shopkeepers and in less private settings. Column (2) incorporates interaction terms with the treatment and reveals a significant shift in how participants in the treatment group derive utility from the specific conditions. The positive signs on the interaction terms with shopkeeper gender and purchase location suggest that the intervention has made these previously aversive conditions less so. Women in the treatment group derive a substantially lower disutility from collecting pads from a male shopkeeper in a more public location than women in the control group. In contrast, the treatment does not significantly alter the utility associated with the price. This means women in the treatment group are deriving the same amount of disutility from a higher price of the pads as women in the control group. They are not willing to pay a higher price for a pack of pads *per se*, when purchasing conditions are held constant.

To allow for better interpretation of the coefficients, in Table 3 we quantify the marginal willingness to pay for each attribute. This expresses the monetary value participants assign to avoiding specific purchasing conditions, i.e. the premium they are willing to pay in order to buy from a female vendor and a more private location. The willingness to pay to avoid male vendor or more public collection locations is substantially lower in the treatment group, by about 23% (from 10.44 to 8.04 BDT) and 62% (from 3.52 to 1.34 BDT) respectively. Women in the treatment group are therefore willing to pay a lower premium than women in the control group to buy pads from a woman or in a location that provides privacy.

Taken together, our evidence suggest that hypothesis 1 is the most plausible explanation for our results and that we can rule out hypothesis 2 as the main driver. Both parts of hypothesis 1 — reduced stigma associated with purchasing pads in public and from a male vendor — are affected by the discussion and seem to be mechanisms affecting our results, but we do not observe that the product is becoming more

Table 3
Discrete choice experiment - willingness to pay.

	Willingness to pay to avoid the attribute (in BDT)
Location inside	
- Control	3.523***
	(0.63)
- Treatment	1.349*
	(0.72)
Male vendor	
- Control	10.442***
	(0.73)
- Treatment	8.024***
	(0.82)
Observations	3,808
Participants	476

Notes: This table presents the estimated WTP to avoid specific attributes in the purchase of menstrual products, measured in BDT. The WTP is calculated based on the results of a conditional logit model from a discrete choice experiment. The model evaluates the utility trade-offs associated with various attributes of the product's acquisition process, including the location of collection (inside the factory vs. outside) and the gender of the shopkeeper (male vs. female). The WTP estimates are derived by dividing the coefficients for 'location inside' and 'male shopkeeper' by the coefficient for 'price' from the conditional logit model. Standard errors, reported in parentheses, are clustered at the individual level. Stars denote significance levels: *p < 0.10, **p < 0.05, ***p < 0.01.

desirable *per se* after the discussion once we control for the social dimension of the purchase.

This analysis underscores a broader attitudinal shift facilitated by the intervention. The reduction in social apprehensions regarding the gender of the shopkeeper and the privacy of the transaction suggests a decline in stigma surrounding menstrual product purchases.

5.2.2. Exploratory analysis: differential impact on women's valuation based on purchasing autonomy and baseline social dynamics

This subsection focuses on the nuanced impacts of our intervention on distinct groups of women, differentiated by their initial menstrual product usage, purchasing habits, and baseline levels of stigma. Our objective is to offer exploratory insights that complement the results from the discrete choice experiment. Specifically, we focus on examining how the intervention affected differently the valuation women place on menstrual products depending on their autonomy in purchasing these products and their initial feelings surrounding menstrual health. Table 4 presents interval regression results for the willingness to pay for disposable pads, disaggregated by three distinct categories of participants. The first category, *Cloth Users*, includes those participants who exclusively use cloth as menstrual absorbents. The second category, *Pad Users who Do Not Buy Themselves*, comprises participants who use

Table 4
Willingness to pay by baseline use and access of menstrual absorbent.

	(1)	(2)	(3)
	Cloth User	Sanitary Napkin User	
		Do Not Buy Herself	Buy Herself
Intervention	26.22**	41.24***	5.38
	(10.62)	(11.23)	(14.41)
Mean dep. var	84.55	90.80	107.51
Observations	168	157	112

Notes: This table presents interval regression results for willingness to pay for sanitary napkins, split into three categories based on baseline product usage. Column (1) shows results for women who only use cloth as an absorbent at baseline, Column (2) for women who use pads but do not purchase them themselves, and Column (3) for women who both use and purchase pads themselves. The intervention coefficients, with robust standard errors in parentheses, measure the change in WTP due to the intervention. Mean dependent variable values indicate the average WTP in the control group for each category. Stars denote significance levels: *p < 0.10, **p < 0.05, ***p < 0.01.

sanitary napkins but report not purchasing them themselves. In the majority of these cases (81.47%), it is the husband who goes to the store to buy these products. The final category, *Pad Users who Buy Themselves*, consists of participants who not only use pads but also personally go to the store and make the purchase. This categorization enables an assessment of how the intervention's impact varies among women with different levels of exposure and autonomy in purchasing menstrual products.

Column (1) reports the results for *Cloth Users*. Post-intervention, this group shows an increased willingness to pay for sanitary napkins, suggesting a growing appreciation for a product they previously did not use. This shift in valuation among participants who previously did not use menstrual products does not conclusively support or discredit any of the hypotheses, as it could be influenced both by increased knowledge and awareness of the product and/or a reduction in the social stigma associated with its purchase.

The most insightful finding comes from comparing the results in Column (2) and (3). Both columns include women already using sanitary napkins, who are therefore familiar with their features and functionality. They differ only in the exposure they have had to the social aspects of purchasing the product, because women in column (2) do not buy the pads themselves, while women in column (3) do. It can be seen in Column (2) that for *Pad Users who Do Not Buy Themselves*, willigness to pay increases by 45% (41.24***) after the treatment compared to the control group mean. In contrast, for *Pad Users who Buy Themselves*, there is no significant change in valuation after the intervention. Since both groups are equally familiar with pads, the intervention's significant effect for women not buying pads themselves suggests that the reduction in social stigma around the purchase is a more plausible channel than changes in information, in line with Hypothesis 1 and in contrast to Hypothesis 2. The increase in valuation likely stems from either a direct normalization of the topic or an awareness that a significant portion of their colleagues personally purchase pads.

Finally, we examine survey metrics of social norms and stigma to gain deeper insights into the different treatment effects for subgroups of participants. These metrics provide additional suggestive evidence to support our findings. To assess social concerns, we used two key survey metrics: one measuring stigma levels and another eliciting second-order beliefs about the social appropriateness of *buying sanitary napkins from male vendors*. These metrics help us understand how our intervention might have affected our sample differently depending on their starting stigma and social norm perceptions.

First, in Table 5 Columns (2)–(4) we split the sample into terciles by their baseline level of stigma - measured as the number of statements expressing stigma they agreed to at baseline (out of four). We find that the increase in valuation is mainly driven by women with higher baseline levels of stigma. Participants with more intense feelings of shame about menstruation — seeing it as a matter to be hidden, regarding it as unclean, and expressing significant concern about others becoming aware of their menstrual status — are the ones who show the most pronounced increase in their willingness to pay for sanitary pads after participating in the discussion sessions. This pattern suggests that the open discussion had a meaningful impact in addressing the deeper layers of psychological and social discomfort associated with menstruation. For the participants in the high stigma tercile, the increase in willingness to pay is statistically significant at the 1% level and represents an increase of 67% compared to the control group mean. In contrast, for the low and middle terciles, the willigness to pay after treatment is not statistically significantly different from the control group, and the magnitude of increase is smaller, at 15% and 9% respectively.

Secondly, in Columns (5)–(6) of Table 5 we look at the effect of the intervention based on whether participants believed at baseline that others viewed purchasing pads from male vendors as socially appropriate versus socially inappropriate. Here, the differential responses based on women's baseline perceptions do not offer a clear direction. A stigma channel would suggest that the increase in valuation would be driven largely by those participants who believed that their peer group sanctions buying disposable pads from men as socially inappropriate. We observe in our data that actually both subgroups experienced an increase in valuation. Women who believed others viewed buying from male shopkeepers as socially inappropriate exhibited a notable, albeit marginal, increase of 21% in valuation compared to the control group. The increase in valuation among women who initially considered buying pads from men to be considered appropriate is of a larger magnitude (31%). Overall, the treatment was effective regardless of initial perceptions of social norms.

These last results should be interpreted within the context of their exploratory nature. They offer suggestive evidence on the interplay between social norms, stigma, and menstrual product valuation, but are not conclusive. The variations in responses, influenced by both stigma and social norm perceptions, highlight the complexity of these factors in shaping women's health-related decisions.

6. Discussion

This discussion section is dedicated to a critical examination of the nuances and complexities inherent to any attempts to disentangle the mechanisms underlying our observed effects, as well as an exploration into any spillover effects and the lasting impacts of this intervention.

It is important to acknowledge the inherent complexity in isolating the specific mechanisms behind the observed changes in our study. While our results suggest that the primary driver of these changes is a reduction in stigma and social concerns, we cannot categorically rule out the influence of an information channel. The intervention, designed to encourage open discussions about menstrual health, inherently blends informational exchange with stigma reduction. Providing information about menstrual health and products is intrinsically linked to the social stigma associated with these topics. In the absence of distinct treatment

Table 5
Willigness to pay by stigma and social norm baseline levels.

	(1)	(2)	(3)	(4)	(5)	(6)
		Terciles of Stigma			Buying Pads from Male Vendor	
	Full Sample	Low	Middle	High	Inappropriate	Appropriate
Intervention	22.98***	15.51	7.90	55.09***	17.28*	34.76**
	(6.34)	(14.04)	(14.87)	(18.87)	(10.05)	(17.70)
Mean dep. var	90.62	99.15	87.78	81.17	80.64	109.23
Observations	476	199	151	124	288	186

Notes: This table presents interval regression results for willingness to pay for sanitary napkins. The analysis is divided into subgroups based on terciles of the stigma index (Low, Middle, High) and perceptions of the social appropriateness of buying pads from male clerks (Inappropriate, Appropriate). Columns (2)–(4) present the results for the stigma index terciles, this index is derived from participants' agreement with statements about menstruation-related stigma, ranging from 0 (no agreement) to 4 (agreement with all statements). Columns (5)–(6) report the perceived social appropriatedness of buying from male clerks, the social norm measure reflects second-order beliefs and not own opinions. Robust standard errors are in parentheses. The mean dependent variable indicates the average WTP for the control in each subgroup. Stars denote significance levels: *p < 0.10, **p < 0.05, ***p < 0.01.

arms that could separately evaluate the effects of information provision and stigma reduction, our study could not completely disentangle these intertwined channels. This limitation highlights a key area for future research, where studies with a more segmented intervention design could offer deeper insights into the relative contributions of information dissemination and stigma alleviation in changing menstrual health practices and perceptions.

A second matter of particular interest in this discussion concerns the outcomes derived from the additional metrics included in the survey to assess changes in stigma and social norms. We presented participants with four scenarios exemplifying typical situations of stigma, and asked them to specify the number of scenarios they agreed with. For example, the scenarios include statements such as *"Women should hide any evidence of menstruation"* and *"I worry about stains or odor during menstruation, because others might realize I am menstruating"*.

In Table 6, we present the regression results derived from the measure of stigma, utilizing a difference-in-differences regression framework. The findings indicate that the intervention significantly reduced the perceived stigma associated with menstruation. Initially, women agreed on average with 1.8 out of 4 stigma-related statements. Post-treatment, this agreement dropped to approximately 0.9 in the treatment group, a significant change at the 1% level. Interestingly, we also noted a change in the control group, with an endline agreement of around 1.3.

The observed changes in the control group could stem from two primary factors. The first is the potential spillover effect from the treatment group to their peers in the factory, thereby influencing the control group as well. The second is an indirect effect caused by the factory management making menstrual underwear available in the factory, which might have normalized access to menstrual products and the overall stigma around menstruation.

To gain further insights into these dynamics, we capitalized on an opportunity six months post-intervention when our research partner conducted a follow-up survey on those study participants that had collected the underwear. This survey aimed to understand the usage and opinions of the participants regarding the products. We included in the follow-up survey a *pure control group* of 59 women who had not been part of the original survey.[11]

The first observation we made was regarding the menstrual absorbents used by the pure control group, which closely mirrored the baseline usage. Table 7 shows that approximately 60% of them use sanitary napkins, while 53% rely on cloth. This pattern of usage is statistically significantly different from that of the study participants at endline and similar to their baseline level. Consequently, we do not observe any direct spillover effect from the treated participants across the factory in terms of changes in the menstrual products women use.

Secondly, the levels of stigma and social norms (appropriatedness of buying pads from a male vendor) over time areare depicted in Fig. 3. The graphical depiction shows that the metrics for the pure-control group closely mirror the baseline values. This consistency offers another piece of suggestive evidence, pointing to the lack of widespread spillover effects from the treatment across the factory to the rest of the female garment workers. Therefore, a more likely explanation for the changes noted in the control group in Table 6 is the direct impact of the garment factory leadership's initiative to make menstrual products available within the factory store and notifying the women directly on their phone (due to the survey) of their availability. This move presumably shifted the women's perceptions about the social acceptability of obtaining these products.

As of January 2024, the outcomes of our research project are evidenced by the pick-up of over 4000 pairs of reusable underwear by women in the collaborating factories provided by the partner NGO. The factories, previously silent on the complexities of menstrual management, have adopted proactive measures to facilitate discussion on this topic. Regular announcements over the factory's loudspeakers now promote a culture of open dialogue about menstrual health, effectively challenging the stigma traditionally associated with menstruation. This strategic communication initiative has helped ensuring widespread awareness and understanding of menstrual health, thereby fostering a workplace environment that acknowledges and supports the menstrual well-being of its predominantly female workforce. This achievement reflects a crucial advancement in making menstrual health options more accessible and aligns with the aims of improving overall menstrual health management.

7. Conclusion

In this paper, we report findings from a field experiment involving 476 women in a garment factory. Our results demonstrate that engaging in open discussions about menstruation, a stigmatized topic, significantly enhances the expected utility of both familiar and innovative menstrual products. Participation in these dialogues led to a notable increase in the willingness to pay for sanitary pads, which needed to be collected from a male vendor in a public store. The average valuation increased by more than 25% (from approximately 91 BDT to about 113

Table 6
Perceived stigma levels.

	Stigma
Endline	−0.493***
	(0.09)
Intervention	0.008
	(0.11)
Endline*Intervention	−0.394***
	(0.13)
Mean of dep. var	1.758
Observations	475

Notes: Difference-in-differences estimation (OLS) of the treatment effect on perceived stigma. *Mean of dep. var* represents the control group mean before the discussion session. *Endline* is a dummy equal to 1 at endline and 0 at baseline. *Intervention* is a dummy equal to 1 if the respondent belongs to the treatment group and 0 to the control. Standard errors clustered at the individual level are reported in parentheses. Stars denote significance levels: *p < 0.10, **p < 0.05, ***p < 0.01.

Table 7
Reported material used at six-month follow-up.

	(1)	(2)	(3)	(4)	(5)
	Sample *(Share)*			Difference	
	Pure-control	Control	Treatment	T-Pure Control	T-Control
Cloth or fabric	0.53	0.23	0.22	−0.31**	−0.01
	(0.50)	(0.42)	(0.42)	(0.07)	(0.05)
Disposable pads	0.58	0.68	0.68	0.10	0.00
	(0.50)	(0.47)	(0.47)	(0.08)	(0.05)
Observations	59	150	141		

Notes. Share of women reporting to use each material frequently at the six-month follow-up, conditional on having collected the product. For the pure control group, the menstrual underwear had not been made available. For columns (1), (2), and (3), standard deviations are reported in parentheses. Columns.
(4) and (5) reports the coefficient of a simple regression of the variable on the treatment status comparing the treatment group to both control groups, the pure control group and the experiment control group. Robust standard errors reported in parentheses. Stars denote significance levels: *p < 0.10, **p < 0.05, ***p < 0.01.

[11] These workers were randomly selected from a complete list of the remaining factory workers.

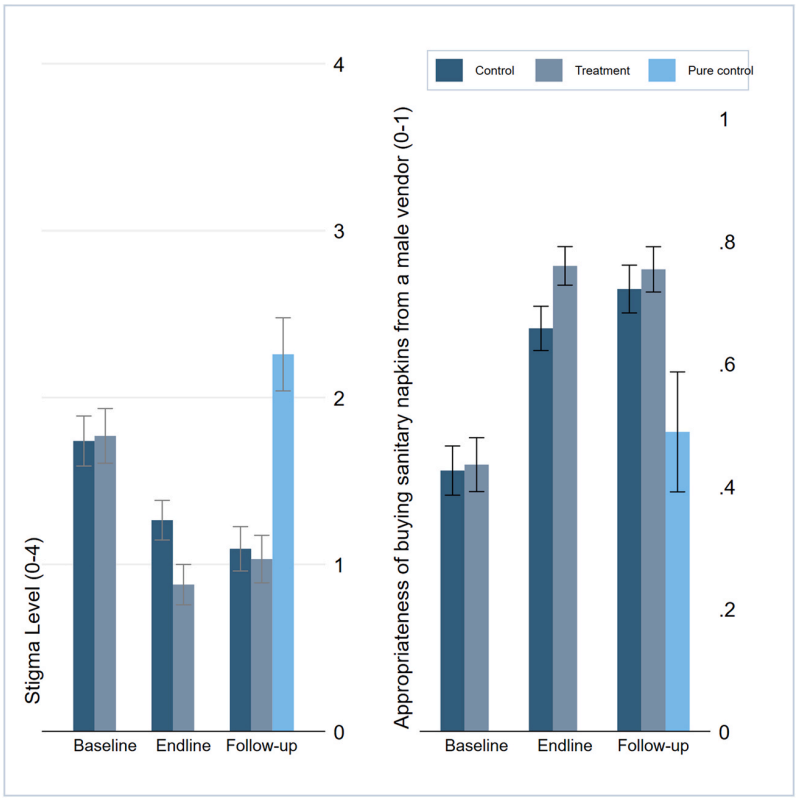


Fig. 3. Perceived Stigma and Social Norms over time.
Notes: The left panel shows the average stigma level for the treatment and control group and the pure control group at baseline, endline and the 6-month follow-up measured on a scale from 0 to 4. The right panel shows the perceived appropriateness of purchasing pads from a male vendor, on a scale from 0 (very socially inappropriate) to 1 (very socially appropriate). Error bars represent 95% confidence intervals.

BDT). Additionally, we observed an increase of around 14% in the take-up rates of the anti-bacterial menstrual underwear (from approximately 71% to about 81%).

We investigated two potential factors influencing the observed effects: an information channel and a stigma channel. The information channel hypothesized that group discussions enabled social learning, where women learned about the benefits of disposable pads from peers, thereby increasing product valuation. Conversely, the stigma reduction channel posited that open discussions about menstruation reduced perceived stigma and normalized the purchase of menstrual products. To discern between these channels, we conducted a discrete choice experiment focusing on product valuation and the impact of contextual factors like observability and vendor's gender. Our findings indicate that the stigma reduction channel was more influential. The treated participants showed less distaste about purchasing from male vendors or in less private settings, while the control group valued more privacy and female shopkeepers. Importantly, we observed no significant difference in the intrinsic value placed on the product itself between the groups. This implies that the primary effect of our intervention was reducing stigma-related concerns, rather than disseminating specific information about menstrual products.

This study offers important insights for policymakers. We propose an effective, light-touch intervention that capitalizes on women's own knowledge and their exchange of ideas and experiences, thereby eliminating the need for external skills or knowledge. We observed significant interest and eagerness among the women to actively engage in discussions and share their personal experiences. Our results suggest that the suboptimal equilibrium limiting women's opportunities to discuss menstruation is fragile and can be changed. This indicates a promising potential for large-scale implementation of such interventions, akin to recent initiatives by public practitioners. Efforts like those made by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), using social media and local influencers to reshape discussions around menstruation and destigmatize the topic in countries like Nepal, Albania, and the Philippines, could prove very effective. These approaches do not necessitate the involvement of every menstruator in a formal discussion group. Simple encouragements to openly discuss this topic and the provision of a safe space for doing so may be sufficient, potentially leading to significant positive impacts on the adoption of health- and productivity-enhancing technologies. Furthermore, our findings highlight that the male gender of vendors is a major barrier; thus, exploring alternative distribution channels that bypass this issue could be highly effective. Such options could include installing vending machines in women's restrooms or offering menstrual products in the factory's health center, a more private setting typically staffed by female nurses.

8. Robustness checks

This section summarizes a series of robustness checks designed to test the resilience of our primary results against different model specifications and assumptions. We first examine the effects of including a comprehensive set of control variables in our primary specifications. Such controls account for potential confounders that might otherwise bias our estimates. We then extend the analysis by considering the composition and characteristics of the discussion groups, which allows to explore whether and how group dynamics might influence the treatment effects observed. Lastly, we include the results for the reusable

underwear WTP. Thought this additional tests, we document that our estimated effects remain robust and unchanged.

8.1. Main specification with controls and enumerator FE

Table R1
Valuation with different type of controls.

	(1)	(2)	(3)	(4)
	Willigness to pay for sanitary napkins			
Intervention	22.32***	22.48*	23.49***	22.60*
	(6.36)	(8.94)	(8.93)	(8.96)
Demographic Controls	Yes	No	No	Yes
Behavior Controls	No	Yes	No	Yes
Stigma Controls	No	No	Yes	Yes
Observations	476	476	476	476

Notes: This table reports the interval regression coefficients of willingness to pay (in BDT) for disposable menstrual pads. The estimation is presented with four different set of controls according to baseline levels of the control variables: (1) Demographic controls (age, religion, years of educ, marital status, and number of children), (2) Menstruation-related variables (use of cloth, use of pads, pregnancy), (3) Stigma index and social norm from buying pads from a male vendor, and (4) All combined. Robust standard errors are shown in parentheses. Stars denote significance levels: *p < 0.10, **p < 0.05, ***p < 0.01.

Table R2
Willingness to pay and collection rates - enumerator fixed effects.

	(1)	(2)	(3)	(4)
	WTP pads		Pick-up rates	
Intervention	22.76**	21.72**	0.08**	0.10**
	(9.34)	(9.12)	(0.04)	(0.04)
Demographic Controls	Yes	Yes	Yes	Yes
Enumerator Fixed Effects	No	Yes	No	Yes
Observations	476	476	469	469

Notes. Columns (1) and (2) report the regression coefficients (OLS) of the intervention on the WTP for pads, with and without enumerator fixed effects. Columns (3) and (4) report the linear probability regression of the collection of the underwear with column (4) adding enumerator fixed effects. Differences in the number of observations between WTP and collection rates are due to six participants winning money or pads in the WTP lottery instead of the underwear. Robust standard errors are reported in parentheses. Stars denote significance levels: *p < 0.1, **p < 0.05, ***p < 0.01.

8.2. Discussion group composition

To learn more about how the discussions affected the women in the treatment group, we explore the treatment effect for each discussion group separately. This allows us to check whether the treatment worked in a similar manner for all women in all treatment groups and to rule out effects driven by outliers. Second, we can evaluate if the size of the treatment effect depends on specific discussion group characteristics. We look at differences in group size, the share of pad users and cloth users, the average age and education level, and the average stigma level at baseline for each group. This allows to examine if any characteristics of the discussion groups are more predictive of success than others to provide lessons for designing discussion groups in future studies or program implementations.

We first regress the WTP for sanitary pads and the probability of product collection on a set of 15 dummy variables, one for each of the 15 discussion groups. The base category consists of the women in the control group. Figure R1 plots the regression coefficients by group for the WTP for sanitary pads (left) and the probability of product collection (right). The figure shows a positive treatment effect on WTP in the majority of treatment groups (though given the small sample sizes of around 15 participants per group, the confidence intervals are wide and the treatment effects not statistically significant for each individual group). The effect of the treatment on the collection of the menstrual underwear is more consistently positive, with most groups showing a higher average collection rate than the control group. Figure R1 also shows that two groups experienced a very large treatment effect on the WTP, groups 11 and 15. To ensure that our results are not only driven by these two groups, we re-run our main regression excluding these groups as a robustness check. The results can be seen in Table R3. This does not

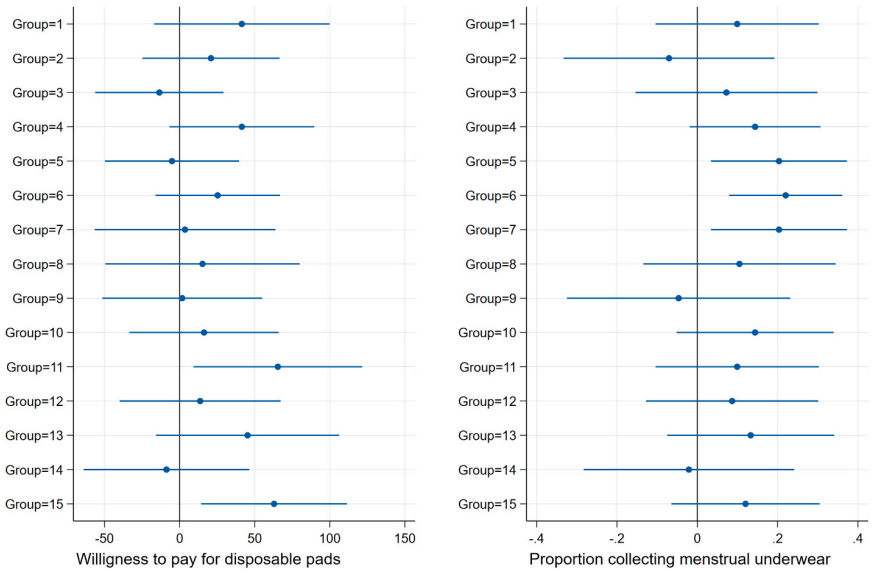


Figure R1. Treatment Effects by Discussion Group
Notes: The left panel plots the regression coefficients obtained from the interval regression of the WTP for sanitary pads on a set of 15 dummy variables indicating participation in the discussion groups (including demographic controls). The right panel plots the regression coefficients obtained from the linear probability regression of the collection probability on the same set of dummy variables. The dots represent the mean effect of being assigned to a given discussion group on the WTP (left) and product collection (right). The bars represent 95% confidence intervals. The base category is the control group.

greatly change the magnitude of the treatment effects or the interpretation of our results.

Table R3
Willingness to pay and collection rates - reduced group sample.

	(1)	(2)	(3)	(4)
	WTP for pads		Pickup of underwear	
Intervention	18.817**	18.455**	0.088**	0.087**
	(9.39)	(9.42)	(0.04)	(0.04)
Excluded Group	*11*	*15*	*11*	*15*
Demographic Controls	Yes	Yes	Yes	Yes
Observations	443	445	438	439

Notes. Columns (1) and (2) report the willingness to pay (in BDT) for disposable menstrual pads. Columns (3) and (4) report the linear probability of the collection of the underwear. Even columns exclude discussion group 15 from the analysis, odd columns exclude discussion group 11 from the analysis. Demographic controls include age, years of education, marital status, number of children and baseline use of pads and cloth (as dummies). Robust standard errors reported in parentheses. Stars denote significance levels: *p < 0.1, **p < 0.05, ***p < 0.01.

Looking at the composition of groups 11 and 15, it is interesting to note that both groups were among the largest groups, with 20 and 17 participants, respectively. Moreover, in group 15 all women were using pads already at baseline. Table R4 provides a general overview of the average characteristics of each group in comparison to each other and the control group.

Table R4
Group summary characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Size	Cloth users	Pad users	Age	Education	Knowledge	Stima
Control	258	0.50	0.61	26.59	7.06	0.77	1.74
Group 1	16	0.69	0.38	24.75	6.56	0.72	1.56
Group 2	15	0.67	0.53	27.13	6.60	0.69	2.21
Group 3	14	0.57	0.43	30.07	5.00	0.77	1.21
Group 4	21	0.43	0.71	27.71	7.23	0.78	2.14
Group 5	13	0.23	0.92	25.08	8.69	0.78	1.92
Group 6	16	0.63	0.50	25.94	7.43	0.76	1.63
Group 7	12	0.50	0.50	26.92	8.50	0.81	1.50
Group 8	11	0.64	0.45	25.81	5.91	0.77	2.00
Group 9	14	0.43	0.50	24.71	9.64	0.82	1.43
Group 10	14	0.57	0.50	26.56	6.14	0.82	1.29
Group 11	17	0.41	0.65	27.41	5.88	0.78	2.06
Group 12	15	0.60	0.40	28.60	7.07	0.74	1.87
Group 13	14	0.21	0.86	23.64	8.64	0.77	2.36
Group 14	13	0.42	0.50	25.85	7.15	0.82	1.69
Group 15	20	0.13	1.00	24.40	8.00	0.83	1.55

Notes. Arithmetic mean and proportions of group characteristics for different demographic and survey measures at baseline. *Size* includes the number of participants in the specified group. *Cloth users* and *Pad users* reports the proportion of respondents that reported to use said absorbent at baseline, *Age* reports the average age, *Education* reports the average years of schooling, *Knowledge* represents the proportion of questions that participants answered accurately regarding biological functions of menstruation, *Stigma* reports the group average on perceived stigma, measured on a scale from 0 to 4 (being 0 the lowest level of perceived stigma).

To determine whether these and other factors of the group composition played a role, we regress the average WTP for sanitary pads of each discussion group (average lower bound) and the average probability of product collection of each discussion group on some of the group characteristics. Given the small number of groups, this analysis lacks statistical power and should be interpreted as only indicative of directional effects.

Table R5
Group composition effects.

	(1)	(2)
	WTP for pads	Pickup of underwear
Share cloth users in group	−34.610	0.422
	(73.11)	(0.64)
Share pad users in group	−19.245	0.539
	(66.00)	(0.53)
Number of group members	4.388**	−0.002
	(1.71)	(0.01)
Average age	−6.319**	0.009
	(2.66)	(0.02)
Average education	−5.999	0.011
	(4.47)	(0.03)
Average stigma at baseline	23.876	−0.136
	(16.56)	(0.14)
Average taboo at baseline	−11.522	0.097
	(17.60)	(0.18)
Constant	235.687	0.085
	(150.69)	(0.80)
Observations	15	15

Notes. Column (1) reports results from the regression of the average (lower bound of the) willingness to pay for pads per group on the different group characteristics. Column (2) reports results from the linear probability regression of the average underwear pick-up rate per group on the group characteristics. Share of cloth and pad users is measured between 0 and 1. Standard errors reported in parentheses. Stars denote significance levels: *p < 0.1, **p < 0.05, ***p < 0.01.

The results are shown in Table R5. Being in a discussion group with a higher share of cloth users appears to have a negative effect on the WTP, a more negative effect than being in a group with a higher share of pad users. The scatter plot in Figure R2 suggests that, if anything, there is a weakly positive relationship between the share of pad users and WTP and a weakly negative relationship between the share of cloth users and WTP. Neither share has an effect on collection rates. Second, being in larger groups with on average younger colleagues seems to increase WTP, though these coefficients are of a very low magnitude. Moreover, the WTP of women in a discussion group with a higher average level of perceived stigma at baseline is higher after the treatment. This could indicate that the treatment is effective in the face of higher stigma levels and has more bite when women are initially constrained. The scatter plot in Figure R3 shows this relationship in more detail.

For the probability of product collection, in contrast, there seems to be no difference between having many cloth users or many pad users in the group. Group size, age and education also have no effect. While lower baseline stigma appears to be positively related with higher collection rates, the scatter plots in Figures R2 and R3 suggest that these effects are not statistically or economically significant.

Overall, these results suggest that the exact group composition and characteristics of the discussion groups do not play a decisive role in determining the treatment effectiveness. We will need to leave it to future research to explore the marginal benefits of further design elements of the discussion groups, such as reducing or extending the time of the discussion or varying the exact content.

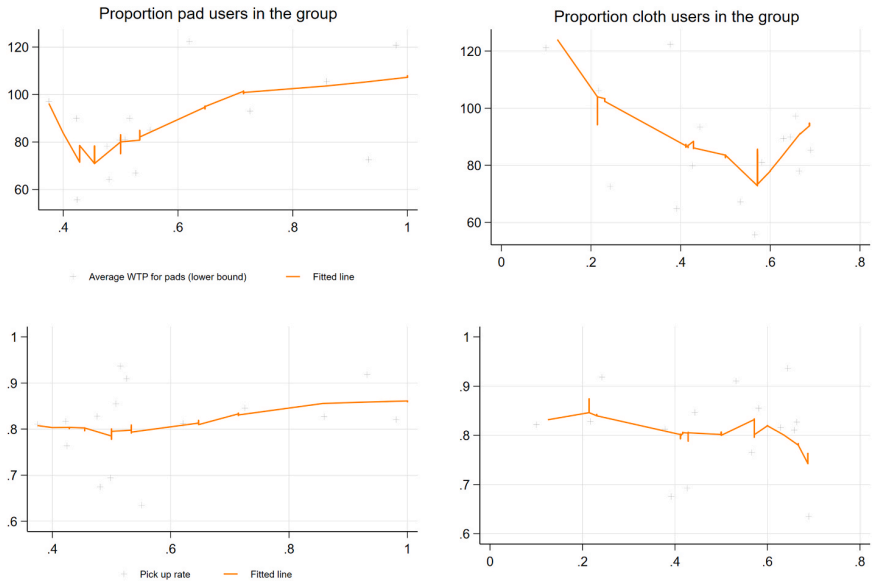


Figure R2. Relationship of the Share of Pad and Cloth Users and Group-Level Outcomes
Notes: The four plots show the average lower bound of the WTP (top panels) and average pickup rates (bottom panels) for each of the 15 discussion groups, plotted against the share of pad users in each group (left-hand panels) and the share of cloth users in each group (right-hand panels). Pad users are defined as women reporting using pads frequently (2 days or more during a period) at baseline, cloth users are defined as women reporting using cloth frequently (2 days or more during a period) at baseline. The lower bound of the WTP is the last value at which a woman preferred the product over the money. The line of best fit is drawn as smoothed locally weighted regression line.

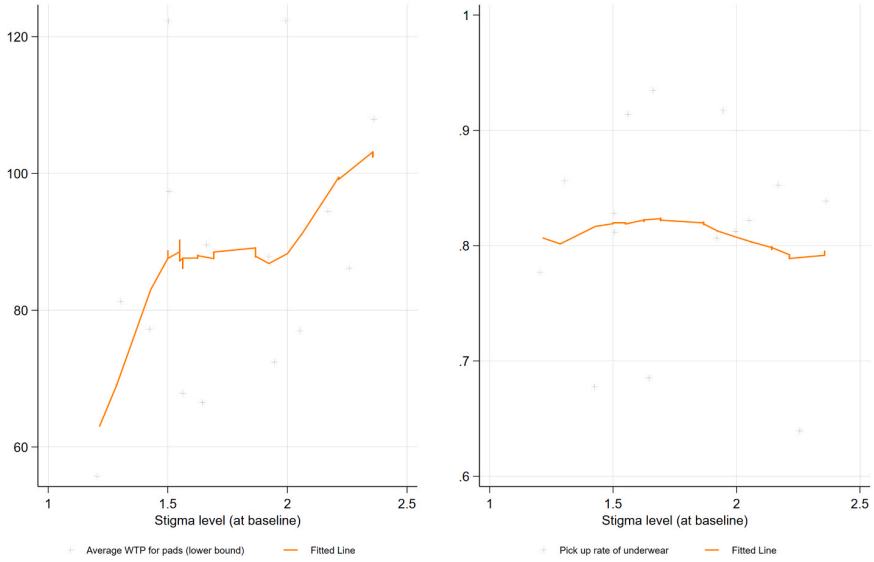


Figure R3. Relationship of Stigma and Group-Level Outcomes
Notes: The plots show the average lower bound of the WTP (left panel) and average pickup rates (right side panel) for each of the 15 discussion groups, plotted against the share of pad users in each group and the share of cloth users in each group. Pad users are defined as women reporting using pads frequently (2 days or more during a period) at baseline, cloth users are defined as women reporting using cloth frequently (2 days or more during a period) at baseline. The lower bound of the WTP is the last value at which a woman preferred the product over the money. The line of best fit is drawn as smoothed locally weighted regression line.

8.3. Willingness to pay for menstrual underwear

Table R6
Valuation of underwear at endline.

	(1)	(2)	(3)	(4)
	WTP underwear			
	Full sample		Without always takers	
Intervention	68.200	50.122	77.469**	71.525**
	(67.76)	(68.54)	(30.47)	(29.89)
Mean Dep. Var	873.187	349.941	93.590	10.775
Demographic Controls	No	Yes	No	Yes
Observations	476	460	106	102

Notes. Interval regression of the WTP (in BDT) at endline for the reusable menstrual underwear from a male shopkeeper at the factory store. Demographic controls in columns (2) and (4) include age, years of education, marital status, number of children and baseline use of pads and cloth. Columns (3) and (4) exclude from the regression participants with a perfectly inelastic demand (i.e. who still preferred the underwear at the maximum price of 500 BDT). Robust standard errors reported in parentheses. Stars denote significance levels: **p < 0.05.

CRediT authorship contribution statement

Silvia Castro: Conceptualization, Data curation, Funding acquisition, Investigation, Software, Supervision, Validation, Visualization, Writing – review & editing. **Clarissa Mang:** Formal analysis, Project administration, Resources, Software, Writing – original draft.

Data availability

The data has been deposited at the LMU-ifo Economics & Business Data Center.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jdeveco.2024.103264.

References

Abramsky, T., Devries, K., Kiss, L., Nakuti, J., Kyegombe, N., Starmann, E., Cundill, B., Francisco, L., Kaye, D., Musuya, T., Michau, L., Watts, C., 2014. Findings from the SASA! Study: a cluster randomized controlled trial to assess the impact of a community mobilization intervention to prevent violence against women and reduce HIV risk in Kampala, Uganda. BMC Med. 12.
Agarwal, S., Chia, L.E., Ghosh, P., 2022. Do sanitary pads alleviate period poverty and improve girls' educational outcomes? SSRN Working Paper.
Allcott, H., 2011. Social norms and energy Conservation. J. Publ. Econ. 95, 1082–1095.
Anderson, S., Harrison, G.W., Lau, M.I., Elisabet, R.E., 2007. Valuation using multiple price list formats. Appl. Econ. 39, 675–682.
Arias, E., 2019. How does media influence social norms? Experimental evidence on the role of common knowledge. Political Science Research and Methods 7, 561–578.
Asadullah, M.N., Mansoor, N., Randazzo, T., Wahhaj, Z., 2021. Is son preference disappearing from Bangladesh? World Dev. 140, 105353.
Ashraf, N., Bau, N., Low, C., McGinn, K., 2020. Negotiating a better future: how interpersonal skills facilitate intergenerational investment. Q. J. Econ. 135, 1095–1151.
Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., Sulaiman, M., 2020. Women's empowerment in action: evidence from a randomized control trial in Africa. Am. Econ. J. Appl. Econ. 12, 210–259.
Banerjee, A., La Ferrara, E., Orozco-Olvera, V.H., 2019. The Entertaining Way to Behavioral Change: Fighting HIV with MTV. National Bureau of Economic Research. Working Paper 26096.
Bangladesh Bureau of Statistics, 2020. National Hygiene Survey, vol. 122.
Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., Topalova, P., 2009. Powerful women: does exposure reduce bias? Q. J. Econ. 124, 1497–1540.
Benshaul-Tolonen, A., Zulaika, G., Nyothach, E., Oduor, C., Mason, L., Obor, D., Alexander, K., Laserson, K., Phillips-Howard, P., 2021. "Sanitary Products, Absenteeism and Psycho-Social Well-Being: Evidence from a Three-Arm Cluster Ran Domized Controlled Feasibility Study in Western Kenya," CDEP–CGEC Working Paper Series 93. Columbia SIPA Center for Developement Economics and Policy.
Bicchieri, C., Dimant, E., 2019. Nudging with care: the risks and benefits of social information. Publ. Choice 1–30.
Buchmann, N., Field, E.M., Glennerster, R., Nazneen, S., Wang, X.Y., 2021. A Signal to End Child Marriage: Theory and Experimental Evidence from Bangladesh. National Bureau of Economic Research. Working Paper 29052.
Bursztyn, L., Gonzalez, A.L., Yanagizawa-Drott, D., 2020. Misperceived social norms: women working outside the home in Saudi Arabia. Am. Econ. Rev. 110, 2997–3029.
Castro, S., Czura, K., 2021. Social Norms and Misinformation: Experimental Evidence on Learning about Menstrual Health Management in Rural Bangladesh. CESifo Working Paper No. 9081.
Cislaghi, B., Berkowitz, A.D., 2021. The evolution of social norms interventions for health promotion: Distinguishing norms correction and norms transformation. J. Glob. Health 11.
Cullen, Z., Perez-Truglia, R., 2022. How Much does Your Boss make? The effects of salary comparisons. J. Polit. Econ. 130, 766–822.
Cullen, Z.B., Perez-Truglia, R., 2018. The Salary Taboo: Privacy Norms and the Diffusion of Information, Working Paper 25145. National Bureau of Economic Research.
Czura, K., Menzel, A., Miotto, M., 2024. Improved menstrual health and the workplace: an RCT with female Bangladeshi garment workers. J. Dev. Econ. 166, 103174.
Dahl, G.B., Kotsadam, A., Rooth, D.-O., 2020. Does integration change gender Attitudes? The effect of randomly assigning women to traditionally male teams. Q. J. Econ. 136, 987–1030.
Dhar, D., Jain, T., Jayachandran, S., 2022. Reshaping adolescents' gender attitudes: evidence from a school-based experiment in India. Am. Econ. Rev. 112, 899–927.
Duflo, E., Dupas, P., Kremer, M., 2015. Education, HIV, and early Fertility: experimental evidence from Kenya. Am. Econ. Rev. 105, 2757–2797.
Dur, R., Fleming, D., van Garderen, M., van Lent, M., 2021. "A Social Norm Nudge to Save More: A Field Experiment at a Retail Bank," CESifo Working Papers 8894. CESifo.
Garg, R., Goyal, S., Gupta, S., 2012. India moves towards menstrual hygiene: Subsidized sanitary napkins for rural adolescent girls - issues and challenges. Matern. Child Health J. 16, 767–774.
Garikipati, S., Boudot, C., 2017. To pad or not to pad: towards better sanitary care for women in Indian slums. J. Int. Dev. 29, 32–51.
Gauri, V., Rahman, T., Sen, I.K., 2018. Shifting social norms to reduce open defecation in rural India. In: Policy Research Working Paper Series 8684. The World Bank.
Ghosal, S., Jana, S., Mani, A., Mitra, S., Roy, S., 2022. Sex workers, stigma, and self-image: evidence from Kolkata brothels. Rev. Econ. Stat. 104, 431–448.
Green, D.P., Wilke, A.M., Cooper, J., 2020. Countering violence against women by encouraging disclosure: a Mass media experiment in rural Uganda. Comp. Polit. Stud. 53, 2283–2320.
Gupta, J., Falb, K.L., Lehmann, H., Kpebo, D., Xuan, Z., Hossain, M., Zimmerman, C., Watts, C., Annan, J., 2013. Gender norms and economic empowerment intervention to reduce intimate partner violence against women in rural Cote d'Ivoire: a randomized controlled pilot study. BMC Int. Health Hum. Right 13.
Hennegan, J., Nansubuga, A., Smith, C., Redshaw, M., Akullo, A., Schwab, K.J., 2020. Measuring menstrual hygiene experience: development and validation of the menstrual practice needs scale (MPNS-36) in Soroti, Uganda. BMJ Open 10.
Jensen, R., Oster, E., 2009. The power of TV: Cable Television and women's status in India. Q. J. Econ. 124, 1057–1094.
Krenz, A., Strulik, H., 2019. Menstruation Hygiene Management and Work Attendance in a Developing Country. CEGE Discussion Papers Number 364.
Krupka, E.L., Weber, R.A., 2013. Identifying social norms using coordination games: why does dictator game sharing vary? J. Eur. Econ. Assoc. 11, 495–524.
La Ferrara, E., Chong, A., Duryea, S., 2012. Soap Operas and fertility: evidence from Brazil. Am. Econ. J. Appl. Econ. 4, 1–31.
Oster, E., Thornton, R., 2011. Menstruation, sanitary products, and school attendance: evidence from a randomized evaluation. Am. Econ. J. Appl. Econ. 3, 91–100.
Paluck, E.L., 2009. Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda. J. Pers. Soc. Psychol. 96 (3), 574–587.
Prentice, D., Paluck, E.L., 2020. Engineering social change using social norms: lessons from the study of collective action. Current Opinion in Psychology 35, 138–142.
Pulerwitz, J., Hughes, L., Mehta, M., Kidanu, A., Verani, F., Tewolde, S., 2015. Changing gender norms and reducing intimate partner violence: results from a quasi-experimental intervention study with young men in Ethiopia. Am. J. Public Health 105, 132–137.
Shidhaye, R., Kermode, M., 2013. Stigma and discrimination as a barrier to mental health service utilization in India. International Health 5, 6–8.
Sumpter, C., Torondel, B., 2013. A systematic review of the health and social effects of menstrual hygiene management. PLoS One 8, 1–15.
Torondel, B., Sinha, S., Mohanty, J.R., Swain, T., Sahoo, P., Panda, B., Nayak, A., Bara, M., Bilung, B., Cumming, O., Panigrahi, P., Das, P., 2018. Association between unhygienic menstrual management practices and prevalence of lower reproductive tract infections: a hospital-based cross-sectional study in Odisha, India. BMC Infect. Dis. 18.
UNICEF, 2019. Guidance on Menstrual Health and Hygiene. UNICEF Reports.
United Nations Development Program, 2022. Human Development Reports: Bangladesh. Accessed at. https://hdr.undp.org/en/countries/proles/BGD.
WSSCC, 2013. "Celebrating Womanhood. How Better Menstrual Hygiene Management Is the Path to Better Health, Dignity and Business," Technical Report. Water Supply Sanitation Collaborative Council.
Yang, D., Allen, J., Mahumane, A., Riddell, J., Yu, H., 2023. Knowledge, stigma, and HIV testing: an analysis of a widespread HIV/AIDS program. J. Dev. Econ. 160, 102958.

A

Appendix to Chapter 1

Tables and Figures

A.1 Tables and Figures

Table A1: Sign-Ups for Cooperative Tasks: Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Endline			*Follow-up*		
	Expertise	*Contribution I*	*Expertise*	*Contribution II*	*Volunteer*	*Mentor*
Award Branch	0.086***	0.089***	0.112***	0.112***	0.068**	0.067**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Most Supportive R1		0.101*				
		(0.05)				
Award Branch × Most Supportive R1		-0.026				
		(0.07)				
Most Supportive R2				0.041		0.088
				(0.05)		(0.06)
Award Branch × Most Supportive R2				0.002		0.006
				(0.08)		(0.08)
Mean Dep. Var	.418		.475		.534	
Branch FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	908	908	908	908	908	908

Notes: The table presents linear probability model estimates of the treatment effect on sign-ups for different cooperation tasks: willingness to provide expertise to the repository (Columns (1) and (2)), willingness to meet with a consultant to improve customer management processes (Columns (3) and (4)), and willingness to sign up as a junior mentor (Columns (5) and (6)). The dependent variable is binary, coded as 1 if the person chose to sign up and 0 otherwise. Columns (2), (4), and (6) include an interaction with being selected as the most supportive employee in the previous survey. If a person was selected and was in the award branch, they received the award. All models include pair-matched fixed effects. Standard errors are clustered at the branch level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Sign-Ups for Cooperative Tasks: Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	Endline		*Follow-up*			
	Expertise	*Contribution I*	*Expertise*	*Contribution II*	*Volunteer*	*Mentor*
Award Branch	0.072**	0.073**	0.124***	0.123***	0.069**	0.061**
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Most Supportive R1		0.068				
		(0.05)				
Award Branch × Most Supportive R1		-0.012				
		(0.07)				
Most Supportive R2				0.043		0.083
				(0.05)		(0.06)
Award Branch × Most Supportive R2				0.005		0.012
				(0.08)		(0.08)
Mean Dep. Var	.4355		.470		.531	
Observations	1016	1016	1004	1004	1004	1004

Notes: The table presents linear probability model estimates of the treatment effect on sign-ups for different cooperation tasks: willingness to provide expertise to the repository (Columns (1) and (2)), willingness to meet with a consultant to improve customer management processes (Columns (3) and (4)), and willingness to sign up as a junior mentor (Columns (5) and (6)). The dependent variable is binary, coded as 1 if the person chose to sign up and 0 otherwise. Columns (2), (4), and (6) include an interaction with being selected as the most supportive employee in the previous survey. If a person was selected and was in the award branch, they received the award. The table includes participants that have been transferred. Standard errors are clustered at the branch level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Sign-Ups for Cooperative Tasks: Logit

	(1)	(2)	(3)	(4)	(5)	(6)
	Endline		*Follow-up*			
	Expertise	*Contribution I*	*Expertise*	*Contribution II*	*Volunteer*	*Mentor*
Award Branch	0.359***	0.391***	0.476***	0.474***	0.268**	0.265*
	(0.14)	(0.15)	(0.14)	(0.15)	(0.13)	(0.14)
Most Supportive R1		0.460**				
		(0.22)				
Award Branch × Most Supportive R1		-0.182				
		(0.30)				
Most Supportive R2				0.175		0.384
				(0.21)		(0.25)
Award Branch × Most Supportive R2				0.012		0.032
				(0.33)		(0.35)
Observations	908	908	908	908	908	908

Notes: Logit estimates of the treatment effect on sign-ups for different cooperation tasks: willingness to provide expertise to the repository (Columns (1) and (2)), willingness to meet with a consultant to improve customer management processes (Columns (3) and (4)), and willingness to sign up as a junior mentor (Columns (5) and (6)). The dependent variable is binary, coded as 1 if the person chose to sign up and 0 otherwise. Columns (2), (4), and (6) include an interaction with being selected as the most supportive employee in the previous survey. If a person was selected and was in the award branch, they received the award. Standard errors are clustered at the branch level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Correlation of Survey Cooperativeness Rankings with Helping Measurement

	(1)	(2)	(3)
	Signing up to Share Knowlege		
Supervisor rating	0.000		
	(0.00)		
Peer rating		0.002**	
		(0.00)	
In top 20% Most Supportive			0.095*
			(0.06)
Mean Dep Var	0.42		
Observations	471	471	471

Notes: The linear probability model estimates the predictive power of various rankings on the willingness to provide productivity tips. The dependent variable takes the value of 1 if the employee signed up to share knowledge and 0 if they chose not to. Column (1) shows the predictive power of the supervisor ranking, which ranges from 0 (employee very rarely helps) to 100 (employee helps very frequently). Column (2) presents the linear relationship with peer ratings, which range from 0 (if no one chose the person as supportive) to 100 (if everyone in the branch chose the person as frequently helping). Column (3) uses a binary variable equal to 1 if the person is in the top 20% most supportive employees in the branch (and would therefore have received the award if in the treatment group), and 0 otherwise. Results are shown only for the control group. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

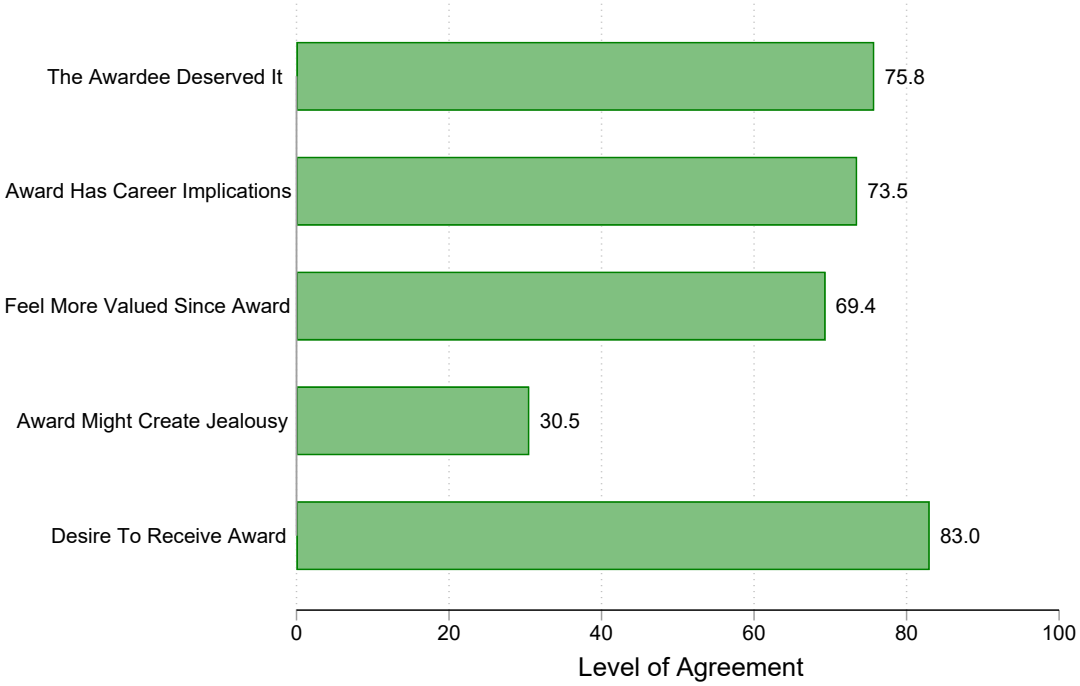


Figure A1: Perceptions of the Intervention at Six Months

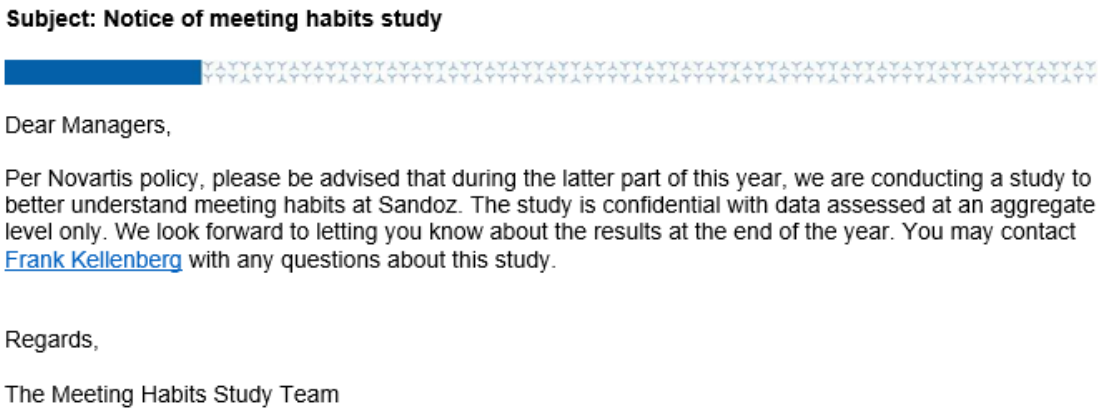
B

Appendix to Chapter 2

Tables and Figures

B.1 Figures

Figure B1: Email to Control Group



Subject: Notice of meeting habits study

Dear Managers,

Per Novartis policy, please be advised that during the latter part of this year, we are conducting a study to better understand meeting habits at Sandoz. The study is confidential with data assessed at an aggregate level only. We look forward to letting you know about the results at the end of the year. You may contact Frank Kellenberg with any questions about this study.

Regards,

The Meeting Habits Study Team

Figure B2: Email to Treatment Groups (1)

Subject: You're personally invited!

Psychological safety? Yes, please!

We would like to ask your support in a joint study to explore the effectiveness of 1:1 conversations with your team members. As we continue to progress in creating psychological safety, our aim is to develop a deeper understanding of how our conversations can impact team dynamics, performance, and well-being. **The study will be conducted between September 13-November 5, 2021.**

1:1 conversations help create the safe space all of us need to dig deeper into understanding what drives and inspires us to perform at our best and create impact that moves the business forward. Each 1:1 conversation is an opportunity to ensure a psychologically safe space for associates to collaborate effectively, be bold in sharing big ideas as well as perceived barriers, and provide the tools needed to enhance performance, well-being, and curiosity.

During this study, you will receive emails from Global Communications (sandoz_global.communications@sandoz.com) with guidance on how to participate and all the support you need. Your involvement will be seamless and easy, we promise! As we're sure you already conduct 1:1 meetings with your team members, incorporating the new tools and techniques into your schedule will be a light lift. Please review these emails carefully and contact Frank Kellenberg with any questions about this study.

Psychological safety is a prerequisite for our continued growth. Your invaluable feedback will therefore help inform our progress whilst also providing you with new insights that will drive success within your teams.

You may opt out of this study by clicking **here** to send an email saying 'opt out.'

Regards,

Inge, Steffen, and Martin
Sponsors of our Psychological Safety Priority

Inge Maes	Steffen Kurzawa	Martin Bischof
Head, Global P&O	Chief Communications Officer	Head, Ethics, Risk, & Compliance

Figure B3: Email to Treatment Groups (2)

Subject line: 😎 Welcome to the psychological safety study at Sandoz!

Thank you for joining our study on psychological safety! Our aim is to develop a deeper understanding of how the quality and frequency of 1:1 conversations impact team dynamics.

During the next eight weeks (Sep 13 – Nov 5), we ask you to set up recurring weekly 1:1s with your team members as your schedule allows. Most managers already practice this approach, so this will be an easy way to get started. The ***attached instructions*** will guide you in how to reimagine your 1:1s so you can have more successful conversations and outcomes. Topics for discussion will also be suggested.

Be sure to use this opportunity to reset how you engage with your teams. Think about how you will actively listen, ask questions, allow the space for organic conversations to happen, and how you will offer support to your team members. Then, consider how you will gauge impact and collaboration within the context of how we go for growth as an organization.

Your continuous participation throughout the entire study is critical to informing our progress on psychological safety as it is a prerequisite for our continued growth. With a little mindfulness, we can seamlessly gather the information we are seeking to explore to ensure Sandoz associates have the tools and space they need be their best at work and at home. We look forward to working with you, and hope you gain insights that will drive success within your teams.

-Your Psychological Safety Study Team

Figure B4: Meeting Guidance for the *Needs* Treatment

THE 1:1 STUDY

Use your 1:1 time to discover and adapt to the individual needs of your team.

Over the next six weeks we want you make space in your 1:1 meetings to focus away from day to day work and towards what matters most to your team members in the long-term, so they feel like unique individuals within the larger organization. Understanding their individual needs and perspectives will open up untapped potential.

We are all unique individuals with unique aspirations. Encouraging your team members to bring more of who they are to work will help develop better team dynamics. Here, we recommend some behaviours to role-model to provide the most valuable support. Follow the below approach:

Invite Insight

Ask your team members to come to your next 1:1 meeting with a topic that is meaningful to them. This could be anything from maintaining work-life balance to career aspirations to relationships within the team. Plan to work on it together over the six week period.

Be Empathetic

Give your team member time to talk about their chosen topic, and make sure you are an active, empathetic listener. Use it as an opportunity to really get to know their drivers, concerns and long-term hopes.

Ask Questions

Continue to let your team member lead the conversation and ask questions in return. They may dedicate their check-in to one topic or combine topics. What's important is that they feel they 'own' the chat and can use it for their benefit, development and well-being.

SANDOZ A Novartis Division

Figure B5: Meeting Guidance for the *Tasks* Treatment

THE 1:1 STUDY

Use your 1:1 time to remove barriers & distractions from people making their most valuable contribution.

Over the next six weeks, we want you to focus your 1:1 time on making your team's lives simpler. This means exploring what is getting in the way of their best work, whether competing goals, unnecessary or deprioritised projects or process, systems or technology issues. Be sure to concentrate on the things within your and your colleagues control.

Removing barriers, distractions and time-sapping tasks is often the best way to unlock potential. Explore what your colleagues feel is getting in the way of their best work, and offer your support and guidance in clearing that path.

Using the Urgency/Importance Matrix to discuss tasks with your team members can help them reach their highest potential. Focus on what isn't important but is taking up unnecessary time and energy, and work together to reduce that friction.

	URGENT	NOT URGENT
IMPORTANT	QUADRANT 1 **REDUCE** Dealing with crisis management. Encourage your team member to spend less time here, and focus on Quadrant 2.	QUADRANT 2 **SCHEDULE** Spending time on strategically important tasks. Discuss the blockers that could prevent time spent on these tasks.
NOT IMPORTANT	QUADRANT 3 **DELEGATE** Enable independent decision making. Empower your team member by assigning tasks in this quadrant.	QUADRANT 4 **DECLUTTER** Help your team member eliminate tasks that do not align with their goals. Allow them to say no to these.

Based on feedback from Sandoz teams, goal conflicts can often be a barrier to completing projects and tasks with the best possible effort.

During your 1:1 sessions, make sure you focus on asking questions that help your team members open up. Ask about their needs and how you can help so that they can bring their best selves to their most important goals.

SANDOZ A Novartis Division

B.2 Tables

Table B1: Description of Variables in the Survey

Variable	Question
	Team Level Metrics
Team PsyS	- I feel safe sharing feedback with colleagues - Different perspectives are valued in my team
Experiment	- I am encouraged to find new and better ways to get things done
Take Risks	- I am encouraged to take informed risks in getting my work done
Quality of Manager	- My manager is a role model of our company values and behaviors - I would recommend my manager to other associates in this firm
Support of Manager	- I feel supported when tackling obstacles that hinder my best work - I have support for my career development - I receive ongoing coaching that helps me to constantly develop - I receive regular feedback to improve my performance - I am satisfied with the recognition I receive for my work
	Organizational Level Metrics
Firm PsyS	- I feel free to speak my mind without fear of negative consequences
Trust	- I trust colleagues across the firm
	Meeting Survey
Meeting Frequency	Reflecting on the last couple of months, how frequently do you have 1:1 meetings with your manager? Answers in scale of: Less than every two months, once every two months, once a month, 2-3 times a month, nce a week, more than once a week
Meeting Content	Over the last couple of months, to what extent did you discuss the following topics in your 1:1 meetings. Answers in a scale of: never, rarely, often, always
Career (related)	Your (Current/future) performance Your career development Relationships within the team Relationships in the wider organization Training and learning on the job
Personal (related)	Personal life Work-life balance Your personal individual needs and aspirations
Poject (related)	Prioritizing of projects and workload Removal of barriers or blockers to perform tasks

Notes: The variables under Team Level Metrics belong to the survey that elicits team-level perceptions. The variables under Organizational Level Metrics include variables that are elicited to measure employee perceptions of the organization. These questions can be answered on a scale from 0-100. The variables under Meeting Survey are categorical and coming from the additional endline survey.

Table B2: Effect of the Intervention on Team and Organizational Psychological Safety: Weighted

	(1)	(2)	(3)
Panel A	**Team Level Psychological Safety**		
			<10%
T1 (Needs)	1.377*	1.431*	1.655*
	(0.79)	(0.79)	(0.85)
T2 (Tasks)	0.598	0.659	0.754
	(0.75)	(0.75)	(0.82)
Dem. Controls	No	Yes	Yes
Survey Controls	Yes	Yes	Yes
Observations	544	544	488
Panel B	**Organizational Level**		
			<10%
T1 (Needs)	0.130	0.101	-0.0264
	(0.87)	(0.86)	(0.95)
T2 (Tasks)	-0.924	-0.688	-0.893
	(0.87)	(0.87)	(0.96)
Dem. Controls	No	Yes	Yes
Survey Controls	Yes	Yes	Yes
Observations	559	559	498

Notes: Weighted OLS regression of the treatment assignment on PsyS with robust standard errors. The dependent variable in Panel A is the team-level PsyS. This metric is the aggregate of the two survey questions *Different perspectives are valued in my team* and *I feel safe sharing feedback with colleagues*. The dependent variable in Panel B is the organizational level PsyS. This metric is the survey question *I feel free to speak my mind without fear of negative consequences*. All answers range from 0-100 being 100 the highest level of agreement with the statement. Column (1) reports the regression results with baseline survey controls. Column (2) adds team-demographic controls. Column (3) shows the results when dropping the 10% highest scoring teams in PsyS at baseline. Panel B presents the same specification, with organizational-level psychological safety as the dependent variable, represented by *I feel free to speak my mind without fear of negative consequences*.

Table B3: Effect of Intervention on Team and Organizational PsyS: Multiple Hypothesis Test

	(1)	(2)	(3)
Panel A	**Team Level Psychological Safety**		
			<10%
T1 (Needs)	1.694**	1.745**	1.976**
	(0.78)	(0.78)	(0.84)
FWER p-value	*.077*	*.073*	
T2 (Tasks)	0.844	0.886	0.966
	(0.77)	(0.78)	(0.86)
FWER p-value	*.286*	*.282*	
Mean Dep. Var.	82.92		81.01
Dem. Controls	No	Yes	Yes
Survey Controls	Yes	Yes	Yes
Teams	544	544	488
Employees	4,349	4,349	
Panel B	**Organizational Level Psychological Safety**		
			<10%
T1 (Needs)	0.128	0.196	0.0595
	(0.92)	(0.91)	(1.00)
FWER p-value	*.922*	*.900*	
T2 (Tasks)	-1.233	-0.987	-1.370
	(0.93)	(0.93)	(1.03)
FWER p-value	*.388*	*.415*	
Mean Dep. Var.	75.61		73.47
Dem. Controls	No	Yes	Yes
Survey Controls	Yes	Yes	Yes
Teams	500	500	444
Employees	4,121	4,121	

Notes: OLS regression of the treatment assignment on PsyS with robust standard errors. The dependent variable in Panel A is the team-level PsyS. This metric is the aggregate of the two survey questions *Different perspectives are valued in my team* and *I feel safe sharing feedback with colleagues*. The dependent variable in Panel B is the organizational level PsyS. This metric is the survey question *I feel free to speak my mind without fear of negative consequences*. All answers range from 0-100 being 100 the highest level of agreement with the statement. Column (1) reports the regression results with baseline survey controls. Column (2) adds team-demographic controls. Column (3) shows the results when dropping the 10% highest scoring teams in PsyS at baseline. Panel B presents the same specification, with organizational-level psychological safety as the dependent variable, represented by 'I feel free to speak my mind without fear of negative consequences'. The p-values reported in this table have been adjusted for multiple hypothesis testing using the Family-Wise Error Rate method as outlined in List et al. (2023). *** p <0.01, ** p <0.05, * p <0.1.

C

Appendix to Chapter 3

DCE, Tables, and Figures

C.1 Construction of Choice Set for Discrete Choice Experiment

The method of the DCE is based on random utility theory (Perez-Troncoso, 2020). The assumption is that individuals receive utility not from the product itself, but from the characteristics, such that the total utility received depends on the combination of characteristics and a random additional term. The utility is thus given by

$$V_{isj} = A'_{isj}\delta + \epsilon_i$$

where V_{isj} is the utility of individual i gained by choosing alternative j in scenario s. A_{isj} is a vector of the attributes and δ is the vector of coefficients. Assuming a linear relationship, the total utility is a linear combination of the utility obtained from each individual characteristic plus the random utility term ϵ_i.

To construct the choice set, using a full-factorial design was not feasible. With three different attributes that have either two or four levels each, there are $2 \times 2 \times 4 = 16$ possible scenarios. This results in $(16 \times 15)/2 = 120$ different comparison scenarios. This is clearly too many to test them all. Instead, we follow the standard procedure as discussed in Mangham-Jefferies et al. (2009) and construct a fractional factorial design that is orthogonal, balanced and maximizes the D-efficiency.[1] We use the existing features of SPSS to construct the choice set fulfilling all of these criteria: Using the inbuilt SPSS orthogonal design feature, we determine that a minimum of eight choice sets is needed to achieve an efficient design. Subsequently, we let SPSS generate eight choice scenarios using the inbuilt “choice design" feature, which fulfil the above criteria. This results in eight scenarios in which the participants need to choose between two alternatives.

In our analysis of the DCE data, we closely follow Lancsar et al. (2017). The coefficients of interest are estimated using the following model:

$$V_{isj} = \alpha_j + A'_{isj}\delta + Z'_i\gamma_j$$

A_{isj} is the vector of characteristics, where price is estimated as continuous variable and location and shopkeeper gender as dummy variables. Z_i is a vector of case-specific variables that are included as controls (age, education, marital status, and baseline

[1] Orthogonal means that the linear parameter estimates are uncorrelated, so the different attributes are independent of each other. A balanced design means each attribute level occurs equally often. A D-efficient design minimizes the size of the variance-covariance matrix given a prior for δ (**?**).

material used). We use a conditional logit model (McFadden's Choice Model, McFadden (1974)) to estimate the coefficients of interest. Our DCE design uses unlabelled alternatives, i.e. the options are defined entirely by the different characteristics and there is no additional name or label to the set of characteristics containing any additional information. Therefore, we estimate the model without alternative-specific constants, since we would expect that there is no difference in the utility obtained from Option 1 or Option 2 if they have the same characteristics (there is no constant utility obtained from choosing either Option 1 or Option 2 independent of the characteristics). In order to determine the effect of the treatment on the evaluation, we add interaction effects of the treatment with each characteristic. These steps allow us to finally determine the willingness to pay (in BDT) of the participants in the treatment and control group to avoid having a male shopkeeper (as opposed to a hypothetical female one) and to avoid collecting the underwear on the factory premises (as opposed to a more anonymous external corner store).

C.2 Tables and Figures

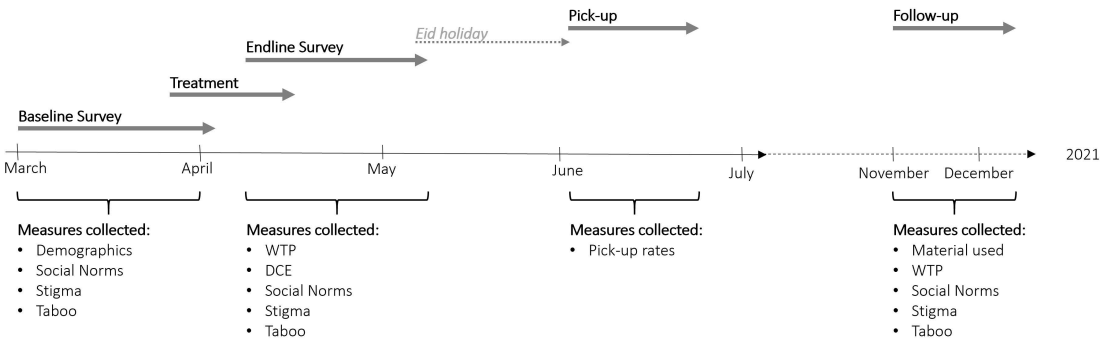


Figure C1: Timeline

Table C1: Descriptive Statistics and Balance of Sample

	(1)	(2) Mean	(3)	(4) Difference
	Full Sample	Control	Treatment	T-C
Age	26.48	26.60	26.34	-0.25
	(4.68)	(4.63)	(4.75)	(0.43)
Muslim religion	0.98	0.98	0.98	-0.00
	(0.13)	(0.12)	(0.13)	(0.01)
Married	0.85	0.87	0.82	-0.05
	(0.35)	(0.33)	(0.38)	(0.03)
Total number of children	1.01	1.04	0.98	-0.07
	(0.84)	(0.87)	(0.80)	(0.08)
Years of education	7.11	7.05	7.17	0.11
	(2.87)	(2.92)	(2.82)	(0.26)
Non-pregnant	0.97	0.96	0.98	0.02
	(0.18)	(0.20)	(0.15)	(0.02)
Menstrual absorbent				
Cloth or fabric	0.48	0.49	0.48	-0.01
	(0.50)	(0.50)	(0.50)	(0.05)
Disposable pads	0.60	0.60	0.59	-0.01
	(0.49)	(0.49)	(0.49)	(0.05)
Reasons to not take-up pads				
Uncomfortable in a store due to a lack of privacy	0.85	0.86	0.83	-0.03
	(0.36)	(0.34)	(0.38)	(0.05)
There is no store nearby	0.21	0.22	0.20	-0.01
	(0.41)	(0.41)	(0.40)	(0.06)
Behavior when buying pads				
Cover face for anonymity	0.52	0.48	0.57	0.10
	(0.50)	(0.50)	(0.50)	(0.09)
Visit store far away to avoid recognition	0.16	0.15	0.18	0.02
	(0.37)	(0.36)	(0.38)	(0.07)
Discomfort if men present in store	0.74	0.71	0.79	0.08
	(0.44)	(0.46)	(0.41)	(0.08)
Discomfort if women present in store	0.03	0.02	0.04	0.02
	(0.16)	(0.12)	(0.20)	(0.03)
Husband buys the pads	0.47	0.46	0.47	-0.01
	(0.50)	(0.50)	(0.50)	(0.06)

Notes: This table presents baseline summary statistics of participant characteristics. Columns (1) through (3) display means for the full sample, control group, and treatment group, respectively, with standard deviations shown in parentheses. Column (4), labeled "T-C", represents the difference between the treatment and control groups. This difference is calculated as the coefficient from a simple regression of the variable on a treatment group dummy, with robust standard errors, also presented in parentheses. Key observations include demographic details like age, religious affiliation, marital status, number of children, and education level. Additionally, the table covers attitudes and behaviors related to menstrual absorbent usage, reasons for not using pads, and behavior patterns when purchasing pads. For interpretation purposes, all variables, except for "age", "total number of children", and "years of education", are coded as binary (0 to 1). *** p <0.01, ** p <0.05, * p <0.1.

Bibliography

ABRAMSKY, T., K. DEVRIES, L. KISS, J. NAKUTI, N. KYEGOMBE, E. STARMANN, B. CUNDILL, L. FRANCISCO, D. KAYE, T. MUSUYA, L. MICHAU, & C. WATTS (2014): "Findings from the SASA! Study: A cluster randomized controlled trial to assess the impact of a community mobilization intervention to prevent violence against women and reduce HIV risk in Kampala, Uganda," BMC Medicine, 12.

AGARWAL, S., L. E. CHIA, & P. GHOSH (2024): "Do Sanitary Pads Improve Girls' Educational Outcomes?" The Review of Economics and Statistics, 1–54.

ALAN, S., G. COREKCIOGLU, & M. SUTTER (2022): "Improving Workplace Climate in Large Corporations: A Clustered Randomized Intervention," The Quarterly Journal of Economics, 138, 151–203.

ALFITIAN, J., D. SLIWKA, & T. VOGELSANG (2023): "When bonuses backfire: Evidence from the workplace," Management Science.

ALLCOTT, H. (2011): "Social norms and energy conservation," Journal of Public Economics, 95, 1082–1095.

ANDERSON, S., G. W. HARRISON, M. I. LAU, & R. E. ELISABET (2007): "Valuation using multiple price list formats," Applied Economics, 39, 675–682.

ANTONAKIS, J., G. D'ADDA, R. A. WEBER, & C. ZEHNDER (2021): "Just Words? Just Speeches? On the Economic Value of Charismatic Leadership," Management Science, 138, 151–203.

ARIAS, E. (2019): "How does media influence social norms? Experimental evidence on the role of common knowledge," Political Science Research and Methods, 7, 561–578.

ASADULLAH, M. N., N. MANSOOR, T. RANDAZZO, & Z. WAHHAJ (2021): "Is son preference disappearing from Bangladesh?" World Development, 140, 105353.

ASHRAF, N., O. BANDIERA, E. DAVENPORT, & S. S. LEE (2020a): "Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services," American Economic Review, 110, 1355–1394.

ASHRAF, N., N. BAU, C. LOW, & K. MCGINN (2020b): "Negotiating a better future: How interpersonal skills facilitate intergenerational investment," Quarterly Journal of Economics, 135, 1095–1151.

AURIOL, E., G. FRIEBEL, & L. PECHLIVANOS (2002): "Career Concerns in Teams," Journal of Labor Economics, 20, 289–307.

AUTOR, D. H., F. LEVY, & R. J. MURNANE (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," The Quarterly Journal of Economics, 118, 1279–1333.

BAER, M. & M. FRESE (2003): "Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance," Journal of Organizational Behavior, 24, 45–68.

BAKER, M. (2016): "1,500 Scientists Lift the Lid on Reproducibility," Nature, 533, 452–454.

BANDIERA, O., N. BUEHREN, R. BURGESS, M. GOLDSTEIN, S. GULESCI, I. RASUL, & M. SULAIMAN (2020a): "Women's empowerment in action: Evidence from a randomized control trial in Africa," American Economic Journal: Applied Economics, 12, 210–259.

BANDIERA, O., S. HANSEN, R. SADUN, & S. BLADER (2020b): "CEO Behavior and Firm Performance," Journal of Political Economy, 128, 1325–1369.

BANERJEE, A., E. LA FERRARA, & V. H. OROZCO-OLVERA (2019): "The Entertaining Way to Behavioral Change: Fighting HIV with MTV," National Bureau of Economic Research Working Paper 26096.

BANGLADESH BUREAU OF STATISTICS (2020): "National Hygiene Survey," Tech. rep.

BATTISTON, D., J. B. I VIDAL, & T. KIRCHMAIER (2021): "Face-to-Face Communication in Organizations," Review of Economic Studies, 88, 574–609.

BEAMAN, L., R. CHATTOPADHYAY, E. DUFLO, R. PANDE, & P. TOPALOVA (2009): "Powerful women: Does exposure reduce bias?" Quarterly Journal of Economics, 124, 1497–1540.

BÉNABOU, R. & J. TIROLE (2003): "Intrinsic and extrinsic motivation," The Review of Economic Studies, 70, 489–520.

BENSHAUL-TOLONEN, A., G. ZULAIKA, E. NYOTHACH, C. ODUOR, L. MASON, D. OBOR, K. ALEXANDER, K. LASERSON, & P. PHILLIPS-HOWARD (2021): "Sanitary Products, Absenteeism and Psycho-Social Well-Being: Evidence from a Three-Arm Cluster Randomized Controlled Feasibility Study in Western Kenya," CDEP–CGEC Working Paper Series 93, Columbia SIPA Center for Development Economics and Policy.

BERTRAND, M. & A. SCHOAR (2003): "Managing with Style: The Effect of Managers on Firm Policies," The Quarterly Journal of Economics, 118, 1169–1208.

BICCHIERI, C. & E. DIMANT (2019): "Nudging with care: The risks and benefits of social information," Public Choice, 1–30.

BINYAMIN, G., A. FRIEDMAN, & A. CARMELI (2018): "Reciprocal care in hierarchical exchange: Implications for psychological safety and innovative behaviors at work," Psychology of Aesthetics, Creativity, and the Arts, 12, 79–88.

BLADER, S. L., C. GARTENBERG, & A. PRAT (2020): "The Contingent Effect of Managerial Practices," The Review of Economic Studies, 87, 721–749.

BLOOM, N., R. SADUN, & J. VAN REENEN (2012): "The Organization of Firms Across Countries," The Quarterly Journal of Economics, 127, 1663–1705.

BRESMAN, H. & A. C. EDMONDSON (2022): "Research: To Excel, Diverse Teams Need Psychological Safety," Harvard Business Review.

BRUHN, M. & D. MCKENZIE (2009): "In pursuit of balance: Randomization in practice in development field experiments," American Economic Journal: Applied Economics, 1, 200–232.

BUCHMANN, N., E. M. FIELD, R. GLENNERSTER, S. NAZNEEN, & X. WANG (2021): "A Signal to End Child Marriage: Theory and Experimental Evidence from Bangladesh," National Bureau of Economic Research Working Paper 29052.

BURSZTYN, L., G. EGOROV, I. HAALAND, A. RAO, & C. ROTH (2023): "Justifying Dissent," The Quarterly Journal of Economics, 138, 1403–1451.

BURSZTYN, L., A. L. GONZALEZ, & D. YANAGIZAWA-DROTT (2020): "Misperceived social norms: Women working outside the home in Saudi Arabia," American Economic Review, 110, 2997–3029.

CAMERER, C. F., A. DREBER, F. HOLZMEISTER, & ET AL. (2018): "Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015," Nature Human Behaviour, 2, 637–644.

CARMELI, A., R. REITER-PALMON, & E. ZIV (2010): "Inclusive Leadership and Employee Involvement in Creative Tasks in the Workplace: The Mediating Role of Psychological Safety," Creativity Research Journal, 22, 250–260.

CASTRO, S. & K. CZURA (2025): "Cultural taboos and misinformation about menstrual health management in rural Bangladesh," World Development, 188, 106871.

CASTRO, S., F. ENGLMAIER, & M. GUADALUPE (2022): "Fostering Psychological Safety in Teams: Evidence from an RCT," SSRN Electronic Journal.

CHATTERJI, A. K., M. FINDLEY, N. M. JENSEN, S. MEIER, & D. NIELSON (2016): "Field Experiments in Strategy Research," Strategic Management Journal, 37, 116–132.

CISLAGHI, B. & A. D. BERKOWITZ (2021): "The evolution of social norms interventions for health promotion: Distinguishing norms correction and norms transformation," Journal of Global Health, 11.

CULLEN, Z. & R. PEREZ-TRUGLIA (2022): "How much does your boss make? The effects of salary comparisons," Journal of Political Economy, 130, 766–822.

CULLEN, Z. B. & R. PEREZ-TRUGLIA (2018): "The Salary Taboo: Privacy Norms and the Diffusion of Information," National Bureau of Economic Research Working Paper 25145.

CZURA, K., F. ENGLMAIER, H. HO, & L. SPANTIG (2024a): "Employee performance and mental well-being: The mitigating effects of transformational leadership during crisis," Management Science, forthcoming.

CZURA, K., A. MENZEL, & M. MIOTTO (2024b): "Improved menstrual health and the workplace: An RCT with female Bangladeshi garment workers," Journal of Development Economics, 166, 103174.

DAHL, G. B., A. KOTSADAM, & D.-O. ROOTH (2020): "Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams," Quarterly Journal of Economics, 136, 987–1030.

DELIZONNA, L. (2017): "High-performing teams need psychological safety. Here's how to create it," Harvard Business Review, 8.

DEVERSI, M., M. G. KOCHER, & C. SCHWIEREN (2020): "Cooperation in a Company: A Large-Scale Experiment," CESifo Working Paper.

DHAR, D., T. JAIN, & S. JAYACHANDRAN (2022): "Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India," American Economic Review, 112, 899–927.

DUFLO, E., P. DUPAS, & M. KREMER (2015): "Education, HIV, and early fertility: Experimental evidence from Kenya," American Economic Review, 105, 2757–2797.

DUHIGG, C. (2016): "What Google Learned From Its Quest to Build the Perfect Team," The New York Times.

DUR, R., D. FLEMING, M. VAN GARDEREN, & M. VAN LENT (2021): "A social norm nudge to save more: A field experiment at a retail bank," Journal of Public Economics, 200, 104443.

EDMONDSON, A. (1999): "Psychological Safety and Learning Behavior in Work Teams," Administrative Science Quarterly, 44, 350–383.

EDMONDSON, A. C. (2018): The Fearless Organization, John Wiley Sons.

EDMONDSON, A. C. & D. P. BRANSBY (2023): "Psychological Safety Comes of Age: Observed Themes in an Established Literature," Annual Review of Organizational Psychology and Organizational Behavior, 10, 55–78.

EDMONDSON, A. C. & Z. LEI (2014): "Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct," Annual Review of Organizational Psychology and Organizational Behavior, 1, 23–43.

ENGLMAIER, F., S. GRIMM, D. GROTHE, D. SCHINDLER, & S. SCHUDY (2024): "The Effect of Incentives in Nonroutine Analytical Team Tasks," Journal of Political Economy, 132.

FRIEBEL, G., M. HEINZ, & N. ZUBANOV (2022): "Middle Managers, Personnel Turnover, and Performance: A Long-Term Field Experiment in a Retail Chain," Management Science, 68, 211–229.

GALLUS, J., S. CAMPBELL, & U. GNEEZY (2023): "Awards: Tangibility, self-signaling and signaling to others," Self-Signaling and Signaling to Others.

GALLUS, J. & B. S. FREY (2017): "Awards as Strategic Signals," Journal of Management Inquiry, 26, 76–85.

GARG, R., S. GOYAL, & S. GUPTA (2012): "India moves towards menstrual hygiene: Subsidized sanitary napkins for rural adolescent girls - Issues and challenges," Maternal and Child Health Journal, 16, 767–774.

GARIKIPATI, S. & C. BOUDOT (2017): "To pad or not to pad: Towards better sanitary care for women in Indian slums," Journal of International Development, 29, 32–51.

GAURI, V., T. RAHMAN, & I. K. SEN (2018): "Shifting social norms to reduce open defecation in rural India," Tech. rep., The World Bank, policy Research Working Paper Series 8684.

GHOSAL, S., S. JANA, A. MANI, S. MITRA, & S. ROY (2022): "Sex workers, stigma, and self-image: Evidence from Kolkata brothels," Review of Economics and Statistics, 104, 431–448.

GIBBONS, R. & R. HENDERSON (2012): "Relational Contracts and Organizational Capabilities," Organization Science, 23, 1350–1364.

GNEEZY, U., S. MEIER, & P. REY-BIEL (2011): "When and Why Incentives (Don't) Work to Modify Behavior," Journal of Economic Perspectives, 25, 191–210.

GONG, Y., S. Y. CHEUNG, M. WANG, & J. C. HUANG (2012): "Unfolding the Proactive Process for Creativity: Integration of the Employee Proactivity, Information Exchange, and Psychological Safety Perspectives," Journal of Management, 38, 1611–1633.

GREEN, D. P., A. M. WILKE, & J. COOPER (2020): "Countering violence against women by encouraging disclosure: A mass media experiment in rural Uganda," Comparative Political Studies, 53, 2283–2320.

GUILLOUET, L., A. K. KHANDELWAL, R. MACCHIAVELLO, M. MALHOTRA, & M. TEACHOUT (2024): "Language Barriers in Multinationals and Knowledge Transfers," Review of Economics and Statistics, 1–56, forthcoming.

GUPTA, J., K. L. FALB, H. LEHMANN, D. KPEBO, Z. XUAN, M. HOSSAIN, C. ZIMMERMAN, C. WATTS, & J. ANNAN (2013): "Gender norms and economic empowerment intervention to reduce intimate partner violence against women in rural Cote d'Ivoire: A randomized controlled pilot study," BMC International Health and Human Rights, 13.

HAECKL, S. & M. REGE (2024): "Effects of Supportive Leadership Behaviors on Employee Satisfaction, Engagement, and Performance: An Experimental Field Investigation," Management Science, 71, 347–365.

HAMILTON, B. H., J. A. NICKERSON, & H. OWAN (2003): "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," Journal of Political Economy, 111, 465–497.

HENNEGAN, J., A. NANSUBUGA, C. SMITH, M. REDSHAW, A. AKULLO, & K. J. SCHWAB (2020): "Measuring menstrual hygiene experience: Development and validation of the menstrual practice needs scale (MPNS-36) in Soroti, Uganda," BMJ Open, 10.

HEURSEN, L., S. FRIESS, & M. CHUGUNOVA (2024): "Reputational Concerns and Advice-Seeking at Work," Research Paper 23-17, Max Planck Institute for Innovation & Competition, available at SSRN: https://ssrn.com/abstract=4538216.

HOFFMAN, M. & S. TADELIS (2021): "People management skills, employee attrition, and manager rewards: An empirical analysis," Journal of Political Economy, 129, 243–285.

HOLMSTROM, B. & P. MILGROM (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," Journal of Law, Economics, & Organization, 7, 24–52, special Issue: Papers from the Conference on the New Science of Organization, January 1991.

IQBAL, Z., F. GHAZANFAR, F. HAMEED, G. MUJTABA, & M. A. SWATI (2022): "Ambidextrous leadership and change-oriented organizational citizenship behavior: Mediating role of psychological safety," Journal of Public Affairs, 22, e2279.

JENSEN, R. & E. OSTER (2009): "The power of TV: Cable television and women's status in India," Quarterly Journal of Economics, 124, 1057–1094.

KIM, S., H. LEE, & T. P. CONNERTON (2020): "How Psychological Safety Affects Team Performance: Mediating Role of Efficacy and Learning Behavior," Frontiers in Psychology, 11, 1581.

KING, G., E. GAKIDOU, N. RAVISHANKAR, R. T. MOORE, J. LAKIN, M. VARGAS, M. M. TÉLLEZ-ROJO, J. E. HERNÁNDEZ ÁVILA, M. H. ÁVILA, & H. H. LLAMAS (2007): "A "politically robust" experimental design for public policy evaluation, with application to the Mexican universal health insurance program," Journal of Policy Analysis and Management, 26, 479–506.

KRENZ, A. & H. STRULIK (2019): "Menstruation Hygiene Management and Work Attendance in a Developing Country," Tech. rep., CEGE, discussion Papers Number 364.

KRUPKA, E. L. & R. A. WEBER (2013): "Identifying social norms using coordination games: Why does dictator game sharing vary?" Journal of the European Economic Association, 11, 495–524.

LA FERRARA, E., A. CHONG, & S. DURYEA (2012): "Soap operas and fertility: Evidence from Brazil," American Economic Journal: Applied Economics, 4, 1–31.

LANCSAR, E., D. FIEBIG, & A. HOLE (2017): "Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software," PharmacoEconomics, 35.

LAZEAR, E. P., K. L. SHAW, & C. T. STANTON (2015): "The Value of Bosses," Journal of Labor Economics, 33, 823–861.

LEVITT, S. D. & J. A. LIST (2009): "Field Experiments in Economics: The Past, the Present, and the Future," European Economic Review, 53, 1–18.

LIST, J. A., A. M. SHAIKH, & A. VAYALINKAL (2023): "Multiple Testing with Covariate Adjustment in Experimental Economics," Journal of Applied Econometrics, 38, 920–939.

MANGHAM-JEFFERIES, L., K. HANSON, & B. MCPAKE (2009): "How to do (or not to do) ... Designing a Discrete Choice Experiment for Application in a Low-Income Country," Health Policy and Planning, 24, 151–158.

MCFADDEN, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in Frontiers in Econometrics.

MICKELER, M., P. KHASHABI, M. KLEINE, & T. KRETSCHMER (2023): "Knowledge seeking and anonymity in digital work settings," Strategic Management Journal.

OSTER, E. & R. THORNTON (2011): "Menstruation, sanitary products, and school attendance: Evidence from a randomized evaluation," American Economic Journal: Applied Economics, 3, 91–100.

PALUCK, E. L. (2009): "Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda," Journal of Personality and Social Psychology, 96, 574–587.

PEREZ-TRONCOSO, D. (2020): "A step-by-step guide to design, implement, and analyze a discrete choice experiment," Tech. rep., arXiv.

PRENTICE, D. & E. L. PALUCK (2020): "Engineering social change using social norms: Lessons from the study of collective action," Current Opinion in Psychology, 35, 138–142.

PULERWITZ, J., L. HUGHES, M. MEHTA, A. KIDANU, F. VERANI, & S. TEWOLDE (2015): "Changing gender norms and reducing intimate partner violence: Results from a quasi-experimental intervention study with young men in Ethiopia," American Journal of Public Health, 105, 132–137.

RIDER, C., A. FERRÈRE, Z. BELOVAI, M. GUADALUPE, & F. ENGLMAIER (2023): "Proven tactics for improving teams' psychological safety: An experiment reveals interventions that managers can use to increase employees' comfort with speaking up and raising concerns," MIT Sloan Management Review.

SANDVIK, J. J., R. E. SAOUMA, N. T. SEEGERT, & C. T. STANTON (2020): "Workplace Knowledge Flows," The Quarterly Journal of Economics, 135, 1635–1680.

SCHEIN, E. H. (1990): "Organizational Culture," American Psychologist, 45, 109–119.

SHIDHAYE, R. & M. KERMODE (2013): "Stigma and discrimination as a barrier to mental health service utilization in India," International Health, 5, 6–8.

SUMPTER, C. & B. TORONDEL (2013): "A systematic review of the health and social effects of menstrual hygiene management," PLoS One, 8.

TORONDEL, B., S. SINHA, J. R. MOHANTY, T. SWAIN, P. SAHOO, B. PANDA, A. NAYAK, M. BARA, B. BILUNG, O. CUMMING, P. PANIGRAHI, & P. DAS (2018): "Association

between unhygienic menstrual management practices and prevalence of lower reproductive tract infections: A hospital-based cross-sectional study in Odisha, India," BMC Infectious Diseases, 18.

UNICEF (2019): "Guidance on Menstrual Health and Hygiene," Tech. rep., uNICEF Reports.

UNITED NATIONS DEVELOPMENT PROGRAM (2022): "Human Development Reports: Bangladesh," Tech. rep., accessed at: https://hdr.undp.org/en/countries/profiles/BGD.

WAGNER, Z., J. B. ASIIMWE, & D. I. LEVINE (2020): "When financial incentives backfire: evidence from a community health worker experiment in Uganda," Journal of Development Economics, 144, 102437.

WATER SUPPLY SANITATION COLLABORATIVE COUNCIL (2013): "Celebrating Womanhood: How Better Menstrual Hygiene Management Is the Path to Better Health, Dignity and Business," Tech. rep., technical Report.

WEIDMANN, B. & D. J. DEMING (2021): "Team players: How social skills improve team performance," Econometrica, 89, 2637–2657.

YANG, D., J. ALLEN, A. MAHUMANE, J. RIDDELL, & H. YU (2023): "Knowledge, stigma, and HIV testing: An analysis of a widespread HIV/AIDS program," Journal of Development Economics, 160, 102958.