SCALING UP, SPLITTING APART: THE NETWORK COST OF PRODUCTION SCALE *

Giorgia Barboni† — University of Warwick

Ayah Bohsali‡ — UPF

Elisa Giannone§ — CREI and CEPR

March 2026

Abstract

Why do organizations fail to scale, especially in low-income countries? We propose a network cost of scale: as teams grow, communication localizes into subgroups and coordination frictions rise. We test this in a randomized controlled trial in rural India that assigns 1,582 women to 206 production teams of 5, 8, or 12 members and cross-randomizes seeding of an improved technique. Using novel dyadic data of who communicates with whom within the team, we show that per-capita productivity falls with team size, seeded information diffuses non-linearly, and larger teams become sparser and more fragmented. A quantified model of team production and endogenous network implies effective interaction neighborhoods rise only from 2.09 to 2.40 as nominal size increases from 5 to 12. In a within-India top-decile organizational-capacity counterfactual, large teams move from segmented organization toward hierarchy-like coordination, cross-subgroup communication reach rises sharply, and per-capita productivity rises by 1.1%. Even under a conservative WMS-anchored US benchmark, large teams again move toward hierarchy-like coordination and per-capita productivity rises by 0.6%.

*This paper is part of a larger project on Self-Help Groups in the state of Chhattisgarh in India that entails a collaboration with the Chhattisgarh State Rural Livelihood Mission (CGSRLM). We thank Tommaso Porzio for feedback on early stages of the project. We thank Vidushi Dhawan, Sejal Luthra and Akanksha Sen at EPoD India for excellent research assistant. Funding from Initiative for What Works to Advance Women and Girls in the Economy (IWWAGE) and Warwick University (Barboni) are gratefully acknowledged.

†giorgia.barboni@warwick.ac.uk

‡ayah.bohsali@upf.edu

§egiannone@crei.cat

1 Introduction

Teamwork is a fundamental feature of the modern economy (Deloitte, 2016). From factory lines and retail floors to farms, labs, and start-ups, output is typically generated by *teams*—small groups of workers who must coordinate tasks, share know-how, and make joint decisions. This raises a basic question in organizational economics: *when does making a team larger improve performance, and when does it backfire?* The same question sits at the core of a broader puzzle about scale: despite the apparent gains from specialization, learning, most organizations remain small, and growth is especially limited in developing countries. A natural interpretation is that the constraints that keep firms small operate through the micro-organization of teamwork: as groups expand, the returns to adding members depend on whether the team can sustain the communication and coordination required for division of labor and for adopting better practices.

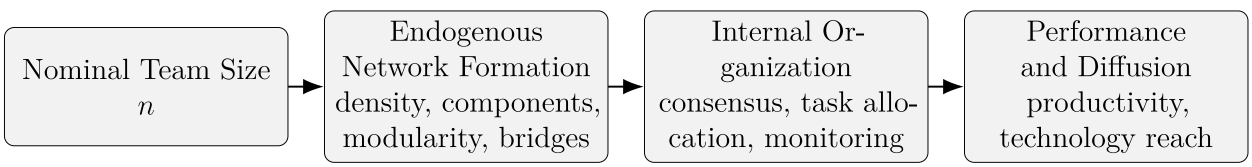
Existing theories emphasize this tradeoff but typically treat the internal structure of interaction as either chosen optimally or scaling smoothly with headcount. Classic work formalizes how the gains from division of labor are limited by coordination costs [Becker and Murphy, 1992a]. Knowledge-hierarchy models show how organizations can scale by adding layers that route problems and information efficiently, preserving effective communication despite rising complexity [Garicano, 2000a, Garicano and Rossi-Hansberg, 2006a]. In these frameworks, scale need not reduce productivity if the organization can implement an internal architecture—often hierarchical—that sustains information flow and coordinated decision-making. Yet a crucial object for both coordination and learning is typically unobserved: *who interacts with whom inside the team.*

This paper advances a complementary mechanism centered on the *endogenous internal network* of interaction within teams. When individuals face limited capacity, they may not expand their set of active relationships proportionally with team size. Instead, as teams grow, interaction can become increasingly localized: members communicate with a smaller share of potential teammates, and the team endogenously reorganizes into disconnected *subgroups*. Communication concentrates within narrower neighborhoods; cross-neighborhood links become sparse; and information and coordination must travel through fewer—and potentially fragile—bridges. This shift generates a wedge between *nominal* size and *effective* size: even if a team has many members on paper, each person's operational network can remain small. The central prediction is therefore not only that performance may decline with size, but that it does so through a measurable and systematic change in network structure.

Testing this mechanism is challenging because team size, team composition, and internal networks are usually jointly determined. Workers sort into teams; managers choose team size together with the allocation of roles and communication routines; and the interaction network itself is rarely observed. We overcome these obstacles with a randomized controlled trial that exogenously varies team size while directly measuring within-team interactions. We randomly assign 1,582 women in rural India into 206 newly formed production teams of different sizes and encourage them to engage in a standardized production process. In a cross-randomization, we seed a subset of members with information about a productivity-enhancing technique. Most importantly for our contribution, we elicit bilateral interactions within teams—who communicates with whom, who meets in person, and who coordinates around production—allowing us to reconstruct the internal network and study how it changes with scale.

This feature of the data is central. Organizational and team-production models typically take communication costs as primitives but rarely observe the *interaction architecture* that those costs generate inside a productive unit. Empirically, internal organization is usually inferred from formal roles, co-location, or partial traces of communication, and is especially difficult to measure in small-scale production settings. In contrast, we directly observe the within-team organization as a network: using roster-style questions, members list with which other members they *i)* communicate by phone, *ii)* communicate in person, *iii)* discuss production decisions, and *iv)* request support during production. These bilateral reports let us reconstruct the internal network and measure not only whether larger teams are sparser, but whether communication becomes *localized into subgroups* and whether cross-subgroup bridges emerge. This observability is what allows us to link scale to a specific mechanism—endogenous fragmentation—and to quantify the implied coordination and diffusion losses.

Figure 1: Conceptual Map of The Mechanism



Note: Conceptual map of the mechanism. The paper studies how nominal scale changes the endogenous internal network, which in turn shapes internal organization, performance, and diffusion.

Our first reduced-form finding is that *bigger teams are less productive per capita.* The main contribution of the paper, however, is to open the black box behind this negative size effect by documenting and leveraging a systematic change in internal networks. As team size increases, each member interacts with a smaller fraction of her potential teammates: networks become sparser relative to size and exhibit patterns consistent with endogenous subgroup formation. Large teams are not simply scaled-up versions of small teams; they behave more like collections of smaller interaction neighborhoods operating inside a larger nominal unit.

These network shifts line up closely with organizational outcomes that matter for production. Smaller teams are more likely to reach consensus on production choices and to organize tasks in a way consistent with efficient specialization, while larger teams more often display patterns consistent with coordination breakdown (e.g., duplicated effort or weak task differentiation). This alignment is central to our interpretation: when communication becomes localized within subgroups, team-wide agreement and coordinated task allocation become harder even if local coordination is feasible.

The size penalty is also sharply heterogeneous in teams' ex ante organizational capacity. Using the baseline survey, we build a pre-treatment team-level capacity index from member education, business exposure, phone and WhatsApp access, SHG participation, leadership, and autonomy. Higher-capacity teams fragment substantially less as they scale: in 12-member teams, a one-standard-deviation increase in baseline capacity raises the bridge index by 0.237 and lowers the number of connected components by 1.430. The amplification is statistically clearest for connected components, where the large-team slope is more negative than in small teams ($p = 0.008$) and medium teams ($p = 0.044$). This matters because it shows that the cost of scale is not mechanical: large teams can perform much more like integrated units when they begin with stronger organizational capacity.

The information-seeding intervention provides an additional test of the mechanism through diffusion. A large literature shows that diffusion depends on who is connected to whom and on network structure [Conley and Udry, 2010a, Banerjee et al., 2013a]. In our setting, the productivity impact of the new technique is *non-linear* in team size. This pattern is consistent with a central implication of the network mechanism: while the *potential* value of a new practice may rise with scale (because there is more scope for reorganizing tasks), its *reach* can fall with scale if the internal network fragments and information becomes trapped within subgroups. The diffusion results therefore reinforce that the network response to scale is not merely descriptive—it is a key determinant of how teams translate innovations into output.

To organize these findings, we develop a theory of team production with endogenous internal networks. The model combines (i) potential gains from knowledge spillovers and specialization that increase with the size of the effective interaction neighborhood [Becker and Murphy, 1992a], with (ii) endogenous communication localization that generates subgrouping as nominal size grows, raising coordination losses and limiting diffusion. We then use this framework to connect the team-level mechanism to a broader development interpretation. Maintaining cross-subgroup bridges requires organizational capabilities—leadership, routines, enforcement, record-keeping, and communication technologies—that may be scarce in low-capacity environments. As a result, the endogenous network response to scale can prevent the emergence of the efficient, connectivity-preserving hierarchies emphasized by canonical models, not because hierarchy would be unproductive, but because sustaining the necessary bridging capacity is difficult. This interpretation links our mechanism to the development literature that views management and organizational capability as scarce technologies in low-income settings [Bloom and Van Reenen, 2010, Bloom et al., 2013, Bruhn et al., 2018].

Our results contribute to several literatures. First, they provide experimental evidence on the limits of scale in team production and the organizational mechanisms behind diseconomies of scale. Second, they connect organizational economics to network formation by treating internal communication structure as an outcome that responds to scale and by directly measuring that structure inside production teams. Third, they speak to the development literature on management and organizational capability by highlighting a concrete channel through which low managerial capacity can keep organizations small: by limiting the ability to sustain cross-subgroup bridges that preserve connectivity and enable diffusion.

Related Literature This paper sits at the intersection of organizational economics, network formation, and the empirical literature on teamwork and diffusion in production settings.

Scale and internal organization. Classic accounts emphasize that the gains from division of labor are ultimately limited by coordination and communication frictions [Becker and Murphy, 1992b, Yang and Borland, 1991, Bolton and Dewatripont, 1994], and that overcoming those frictions is central to understanding why most organizations remain small [Lucas, 1978, Chandler et al., 2009]. A complementary perspective highlights that scaling production often requires standardization—otherwise "unbundling" tasks across specialized workers becomes difficult when goods and processes are customized [Piore and Sabel, 1984, Holmes and Stevens, 2014, Dessein and Santos, 2006]. Two recent contributions provide particularly useful reference points for the development and counterfactual direction of our project. First, Bassi et al. [2023] study the internal organization of small manufacturing firms and show that limits to within-firm specialization can be quantitatively important for explaining why firms remain small. Their framework is especially close to ours in spirit because it treats organization as a first-order constraint and uses a disciplined quantitative model to run counterfactuals. Our experiment complements this approach by holding constant a key confound in many production environments—heterogeneity in product characteristics and demand—and by directly observing the within-team interaction network. In our setting, even when the production task is standardized, scaling up the nominal team can reduce effective coordination because the communication network endogenously contracts and fragments. This provides a different microfoundation for why scale may fail: not only because tasks are hard to separate, but because maintaining the bridges needed for integrated coordination is itself a scarce organizational capability.

Second, Engbom et al. [2025] connect economic development to the emergence of large, managerially intensive firms in the Chandlerian sense and emphasize that development can be understood as an expansion in the capacity to reorganize production through white-collar and managerial capabilities. Our findings offer micro-level evidence for a closely related channel operating inside teams: as groups grow, interaction becomes increasingly localized and the team can split into subgroups rather than endogenously generating the connectivity-preserving hierarchies highlighted in canonical models. In this sense, our mechanism shifts the emphasis from "endogenous layers conditional on a connected organization" to "endogenous fragmentation when bridging capacity is scarce." This perspective naturally motivates the quantitative counterfactual we pursue in the paper: the output and diffusion losses attributable to limited bridging capacity, and how those losses would change if development expanded the feasibility of cross-subgroup coordination.

Hierarchy versus fragmentation. A large literature studies how organizations can scale by endogenously designing hierarchies and spans of control that route problems and information efficiently [Garicano, 2000b, Garicano and Rossi-Hansberg, 2006b]. In this view, larger scale need not reduce productivity if firms can sustain the connectivity needed to transmit knowledge and coordinate decisions; technologies that change information and communication costs can reshape delegation and organizational form [Bloom et al., 2014]. Recent work also links the prevalence of large, hierarchical firms to development through the accumulation of managerial and white-collar skills [Engbom et al., 2025]. Our mechanism is complementary but conceptually distinct: rather than focusing on endogenous *layers* conditional on a connected organization, we study an endogenous *communication network* whose connectivity itself responds to scale. When "bridging capacity" is scarce, growth need not lead to more layers—it can instead lead to *fragmentation*, with the nominal team splitting into subgroups that behave like smaller, weakly connected production units.

Networks and diffusion. Our analysis builds on the economics of networks, where agents form links by trading off the benefits of connectivity against link-maintenance costs, and disconnected or clustered networks can arise in equilibrium [Jackson and Wolinsky, 1996b, Bala and Goyal, 2000b, Jackson, 2008]. Empirically, diffusion and learning depend critically on network position and structure [Conley and Udry, 2010a, Banerjee et al., 2013a]. We complement this work by (i) randomizing team size, which cleanly shifts the set of potential within-team links, and (ii) directly measuring bilateral interactions within production teams. This allows us to document how scale changes density, centralization, and subgroup formation inside teams—and how these network changes map into coordination and technology diffusion. Relative to this literature, our distinctive contribution is that we observe the *entire within-team* interaction network in a production setting and can therefore treat internal connectivity and subgrouping as endogenous organizational outcomes rather than unobserved primitives.

Teams, peer effects, and incentives. Finally, we relate to the literature on teams and peer effects, starting with Holmstrom [1982] and a growing body of empirical work on peer effects and team production [Mas and Moretti, 2009, Bandiera et al., 2010, Agha et al., 2018, Bonhomme, 2021, Devereux, 2018]. A recent theoretical contribution by Claveria-Mayol et al. [2024] (R&R *Journal of Political Economy*) shows how optimal incentive provision depends on workers' positions in a peer network and on network centrality. Our contribution is different in focus: we treat the peer network as an endogenous outcome of team scale and organizational capacity, and we use experimental variation plus network measurement to open the black box linking size, networks, coordination, and performance.

Finally, our paper adds to a growing literature that uses field experiments to uncover and explain key characteristics of firms such as Cai and Szeidl [2018], Brooks et al. [2018] and Brooks et al. [2020]. In contrast to this set of papers that focus on connecting firms with other firms, we study the internal dynamics of production teams rather than external firm-to-firm links.

Our contributions are threefold. First, to our knowledge, this is among the first field experiments to combine exogenous variation in the size of a productive unit with direct measurement of the within-unit bilateral interaction network, allowing us to study how the interaction architecture responds to scale. Second, we show that larger teams are not simply scaled-up versions of smaller ones. As nominal size increases, effective interaction neighborhoods do not expand one-for-one and the internal network fragments into subgroups; these network shifts are mirrored in weaker consensus and less effective task allocation, and they shape technology diffusion. Third, we develop and quantify a model that merges team production with endogenous internal networks and heterogeneous organizational capacity. Using baseline survey data to discipline both organizational and productive ability, the quantification interprets development as a shift in the distribution of organizational capacity and shows that such shifts move large teams most strongly from fragmented to hierarchy-like organization.

The rest of the paper is divided as follows. In Section 2, we describe the experimental design and sample. Section 3 introduces a conceptual framework and derives the main predictions we take to the data. Section 4 presents the data, the reduced-form empirical strategy, and the main productivity results. Section 5 evaluates those predictions using reduced-form evidence on team interactions, task allocation, leadership, and technology diffusion. Section 6 reports robustness checks and alternative explanations. Section 7 quantifies the heterogeneous network mechanism and studies development counterfactuals. Section 8 concludes. Additional results are in the Online Appendix.

2 Setting and Experimental Design

We leveraged the partnership with CGSRLM to create 206 new women's collectives ("production teams" or "teams" henceforth) in rural Chhattisgarh, involving a total of 1,582 women across 120 villages. At the creation stage, all teams received business training focusing on soap production techniques to promote female entrepreneurship under the creation of small-scale joint businesses. The reason for choosing soap over other possible products is the relative simplicity of the techniques and the low cost of raw material to produce soap bars. In addition, to maximize soap bars production, women should work in groups with a clear division of labor, mirroring the functioning of more complex production systems. This holds the potential to help women set the foundations for profitable business activities. A description of the production process can be found in Appendix B. As shown in Figure 2, the experiment was initiated at the end of November 2021 and was completed at the end of April 2022.

The RCT leverages two interventions that promote interactions and knowledge sharing among co-workers. The first intervention consists in exogenously varying the size of the production teams – either 5 ("Small-size teams" treatment), or 8 ("Medium-size teams" treatment), or 12 ("Large-size teams" treatment) women per team. This is used to understand how the size of teams of coworkers affects learning and productivity. The second intervention consists in randomly "seeding" additional technology-related information ("Info Seed" treatment) in the teams to test how coworkers learn from each other, change the internal organization of labor to adopt it and whether their performance improves. This is used to study how a new technology might affect the performance of groups of varying size and their internal organization.

Sampling The RCT target population consists of women living in rural villages in the Mahasamund district of the Indian State of Chhattisgarh. The main inclusion criteria for women to be part of this study and to receive the business training are: i) they should be members of a Self-Help Group (SHG) affiliated to CGSRLM at the time of the recruitment; ii) they should be between 18 and 65 years of age; iii) they should own or have access to a smartphone (this is because the business training was administered via smartphones).

Randomization Randomization first occurred at the village level. Study villages were randomly assigned to three treatment arms, corresponding to creating production teams of 5, 8 and 12 women, respectively. A total of 218 groups were created, of which 47 consist of 12 members, 63 of 8 members and 96 of 5 members, respectively.[1] This yielded a sample size of 1,582 women in 120 villages. We then selected women to be part of the group(s) in a given village by randomly drawing from the lists of female SHG members in that village. Figure 3 shows the design of the experiment and how many women and villages were assigned to each treatment and cross-treatment arm.

Interventions After being administered a baseline survey individually, women were invited to attend a business training event to be held on the same day with other women from the same village. The training was held in central locations in the village and lasted approximately two hours. The main objectives of the training were to: *i)* introduce women to soap bars' production and *ii)* encourage them to start working with the other women who also attended the training event. The training was designed in partnership with CGSRLM. The training followed a "blended" approach: at the event, one of our enumerators explained to participants the aims of the business training, including the importance of acquiring business skills to increase their agency. A video was shared with women via their smartphones, wherever available, providing a detailed guide on how to produce soap bars. Importantly, the video stressed the group dimension of the production process, and encouraged trainees to engage in a joint soap-business activity. Every woman who attended the training event was offered the same business skill training, irrespective of the randomly-assigned size of the production group they ended up being assigned to.

In addition to the initial set of information women received as part of the business training, half of the women in our study sample were also randomly exposed to additional information about a new technology that were delivered to them via phone calls. Specifically, we provided information on how to use sari cloth to wrap the soap bars instead of plastic foil. Overall, the aim of this second intervention was to enhance the quality of soap bars –especially in terms of innovation and environmental sustainability of the production process.

[1]A village could be randomized to have at most two groups

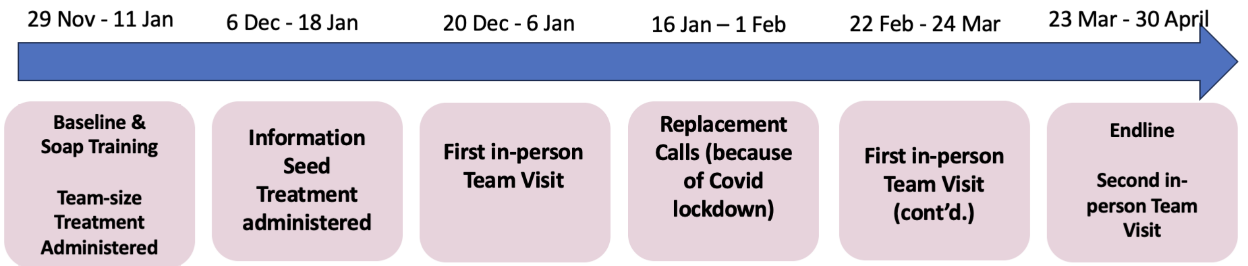
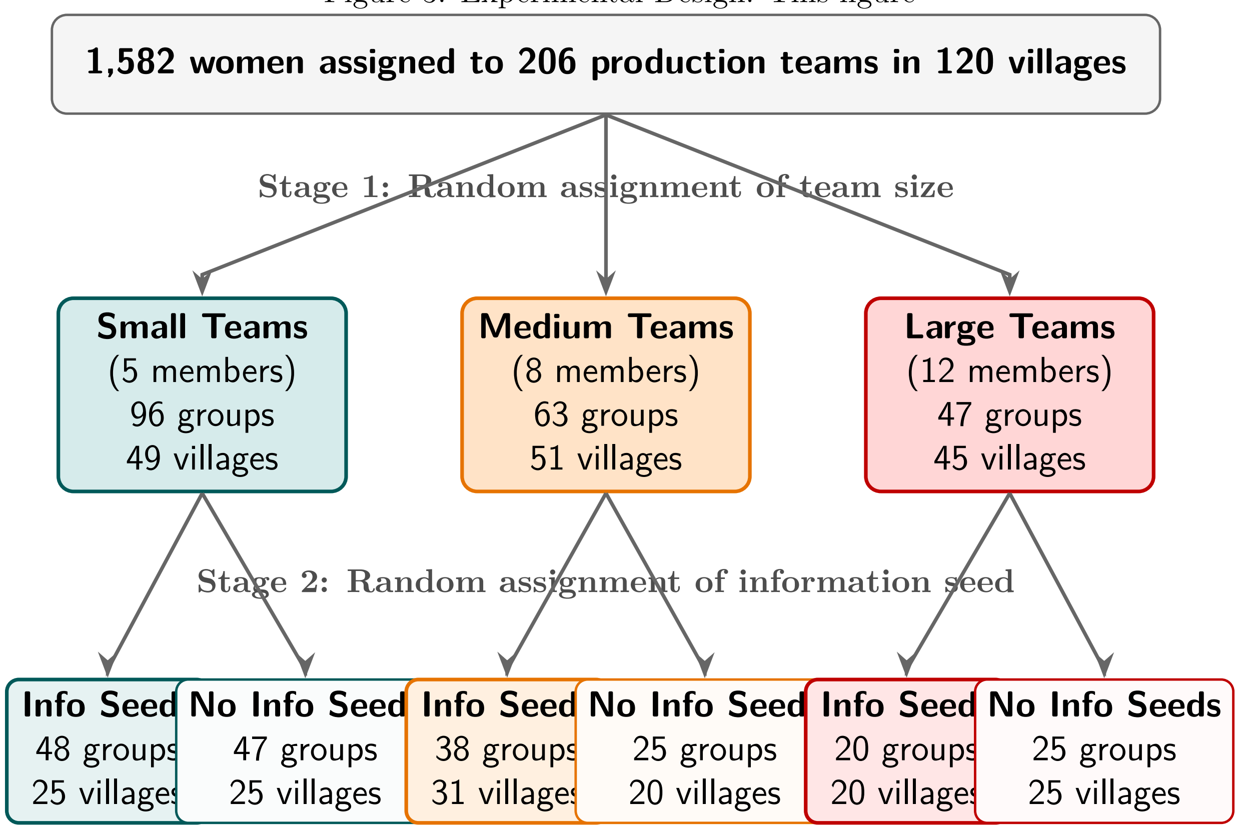


Figure 2: RCT and Data Collection Timeline

Note: Figure 2 plots the full RCT and data-collection timeline, including group formation, business training, monitoring rounds, replacement calls during the Covid interruption, and endline follow-up.

Figure 3: Experimental Design: This figure



Note: Figure 3 shows the experimental design and the type of groups, number of subjects involved, and number of members and villages in each treatment arm. It specifies how many groups were in each treatment and in how many villages were distributed.

3 Theory: Endogenous Fragmentation as a Network Cost of Scale

This section formalizes a simple mechanism: as nominal team size n increases, individuals do not scale their interactions one-for-one, so the team endogenously splits into smaller interaction neighborhoods ("subgroups"). This shrinkage of *effective networks* generates coordination losses and diffusion bottlenecks that can dominate the potential benefits of scale.

The purpose of the section is to interpret the reduced-form results already shown in Table 1 and to organize the mechanism evidence that follows. The model is therefore deliberately parsimonious. It is not yet a quantified structural exercise. Instead, it isolates the comparative statics that the paper can test directly with the experimental data: how size changes effective network reach, when subgrouping emerges, when hierarchy can reconnect those subgroups, and how those organizational responses affect productivity and diffusion.

We use three standard ingredients from the literature. First, specialization and knowledge pooling generate productivity gains but require coordination, so scale faces a coordination tradeoff [Becker and Murphy, 1992b]. Second, communication ties are costly/limited, a central premise in network formation models with costly links [Jackson and Wolinsky, 1996a, Bala and Goyal, 2000a]. Third, diffusion of a new practice occurs through interaction paths, as in the diffusion and learning-on-networks literature [Conley and Udry, 2010b, Banerjee et al., 2013b].

Novel component. The new element is to treat *endogenous subgroup formation* as the central organizational response to scale, and to derive sharp predictions for measurable within-team network moments (density/degree normalized by n, number of components, centralization patterns), and for diffusion and performance. The hierarchy/managerial-capital extension is introduced only at the end as a complement: it shows when a connected organization (a GRH-like regime) is feasible and efficient.

3.1 A baseline model of endogenous fragmentation

Environment and timing A production team has exogenous size $n \in \mathbb{N}$. The team produces output Y which is shared equally across members, so per-member income is $y \equiv Y/n$. We consider a one-shot production episode:

1. The team chooses an internal interaction structure (organization).
2. Production occurs. Output depends on the chosen structure via spillovers and coordination losses.

We model the organization as a partition into k subgroups of equal size s, so that

$$n = ks, \qquad k \in \mathbb{N}, \quad s \in \{1, \dots, n\}. \tag{1}$$

The partition is a reduced-form representation of equilibrium localization that arises in many settings with limited attention and costly ties [Jackson and Wolinsky, 1996a, Bala and Goyal, 2000a].

Spillovers within the effective network (standard). Let s denote the size of a worker's effective interaction neighborhood (her subgroup). We assume spillovers/specialization gains are increasing and concave in s:

$$B(s) = \alpha \ln(s), \qquad \alpha > 0. \tag{2}$$

This concavity captures diminishing marginal returns to additional collaborators and is consistent with the specialization/knowledge logic in Becker and Murphy [1992b].

Local communication/interaction cost (standard reduced form). Coordinating within a subgroup of size s consumes time and attention (meetings, messages, conflict resolution), so local coordination costs are increasing in s:

$$C_{\text{local}}(s) = c\,(s-1), \qquad c > 0. \tag{3}$$

This is the standard "communication is costly" premise that underlies both coordination theories [Becker and Murphy, 1992b] and costly-link network formation [Jackson and Wolinsky, 1996a, Bala and Goyal, 2000a]. (We interpret this as the cost of sustaining an interaction neighborhood of size s, not literally the cost of a single graph edge.)

Cross-subgroup coordination losses (novel mechanism). Fragmentation into many subgroups creates additional costs: inconsistent decisions, duplicated tasks, and failure to reach consensus for group-level choices. We model this as a per-member loss increasing in the number of subgroups $k = n/s$:

$$C_{\text{frag}}(n,s) = \delta \left(\frac{n}{s} - 1\right), \qquad \delta > 0. \tag{4}$$

This term is the model's mechanism object: it makes *subgrouping* costly and links scale to coordination failures.

Per-member income (baseline). Let baseline per-member productivity absent spillovers and coordination losses be y_0. Then per-member income is

$$u(n,s) = y_0 + \alpha \ln(s) \;-\; c(s-1) \;-\; \delta \left(\frac{n}{s} - 1\right). \tag{5}$$

3.1.1 Endogenous subgroup size and effective networks

The team chooses s to maximize per-member income:

$$s^*(n) \in \arg\max_{s \in [1,n]} u(n,s), \tag{6}$$

treating s as continuous for characterization and then taking the nearest integer.

Lemma 1 (Characterization of optimal subgroup size)**.** *The objective $u(n, s)$ is strictly concave in s on $(1, n)$. The unique interior maximizer satisfies*

$$\frac{\alpha}{s} - c + \delta \frac{n}{s^2} = 0, \tag{7}$$

with closed-form solution

$$s^*(n) = \frac{\alpha + \sqrt{\alpha^2 + 4c\delta n}}{2c}. \tag{8}$$

Proof of Lemma 1. Fix $n \geq 1$. The per-member objective is

$$u(n, s) = y_0 + \alpha \ln(s) - c(s - 1) - \delta \left(\frac{n}{s} - 1 \right), \qquad s \in [1, n].$$

Since $u(n, \cdot)$ is continuous on the compact interval $[1, n]$, a maximizer exists.

For $s \in (1, n)$,

$$\frac{\partial u}{\partial s}(n, s) = \frac{\alpha}{s} - c + \delta \frac{n}{s^2}, \qquad \frac{\partial^2 u}{\partial s^2}(n, s) = -\frac{\alpha}{s^2} - 2\delta \frac{n}{s^3} < 0.$$

Hence $u(n, \cdot)$ is strictly concave on $(1, n)$, so there is at most one interior maximizer. Any interior maximizer $s^*(n) \in (1, n)$ must satisfy the first-order condition

$$\frac{\alpha}{s} - c + \delta \frac{n}{s^2} = 0.$$

Multiplying by s^2 gives the quadratic

$$cs^2 - \alpha s - \delta n = 0,$$

whose roots are

$$s = \frac{\alpha \pm \sqrt{\alpha^2 + 4c\delta n}}{2c}.$$

Because $\sqrt{\alpha^2 + 4c\delta n} > \alpha$, the "minus" root is negative. Therefore the unique positive (and thus economically relevant) root is

$$s^*(n) = \frac{\alpha + \sqrt{\alpha^2 + 4c\delta n}}{2c},$$

which is the unique interior candidate. If this $s^*(n)$ falls outside $[1, n]$, the (unique) maximizer is at the boundary; otherwise it is the interior solution above. □

Proposition 1 (Core fragmentation prediction: effective networks grow sublinearly)**.** *The optimal subgroup size $s^*(n)$ is increasing in n, but the* share *of the team contained in a worker's effective network shrinks:*

$$\frac{ds^*(n)}{dn} > 0 \qquad \textit{and} \qquad \frac{d}{dn}\left(\frac{s^*(n)}{n} \right) < 0 \textit{ for all sufficiently large } n.$$

Hence, as teams become larger, individuals interact with a smaller fraction of potential teammates.

Proof of Proposition 1. Let $\Delta(n) \equiv \alpha^2 + 4c\delta n$ and recall the interior candidate

$$s^*(n) = \frac{\alpha + \sqrt{\Delta(n)}}{2c}.$$

Whenever the optimum is interior (so $s^*(n) \in (1, n)$), we can differentiate:

$$\frac{ds^*(n)}{dn} = \frac{1}{2c} \cdot \frac{1}{2\sqrt{\Delta(n)}} \cdot (4c\delta) = \frac{\delta}{\sqrt{\Delta(n)}} > 0,$$

so $s^*(n)$ is increasing in n.

Next consider the fraction of teammates contained in the effective neighborhood:

$$p(n) \equiv \frac{s^*(n)}{n}.$$

For interior $s^*(n)$,

$$p'(n) = \frac{n\, s^{*\prime}(n) - s^*(n)}{n^2}.$$

Compute the numerator:

$$\begin{aligned}
n\, s^{*\prime}(n) - s^*(n) &= \frac{n\delta}{\sqrt{\Delta(n)}} - \frac{\alpha + \sqrt{\Delta(n)}}{2c} \\
&= \frac{2cn\delta - \alpha\sqrt{\Delta(n)} - \Delta(n)}{2c\sqrt{\Delta(n)}} \\
&= \frac{2cn\delta - \alpha\sqrt{\Delta(n)} - \alpha^2 - 4c\delta n}{2c\sqrt{\Delta(n)}} \\
&= -\frac{\alpha\sqrt{\Delta(n)} + \alpha^2 + 2c\delta n}{2c\sqrt{\Delta(n)}} < 0.
\end{aligned}$$

Therefore $p'(n) < 0$ for every n in the region where the interior solution applies. In particular, for all sufficiently large n, the interior condition holds and $s^*(n)/n$ is strictly decreasing.

(For completeness: a sufficient condition for $s^*(n) < n$ is $n > (\alpha + \delta)/c$, obtained by solving $s^*(n) < n$; thus for n beyond this threshold the interior formula applies and the comparative statics above are exact.) □

3.2 Network implications (density, components, and centralization)

To map the model to measurable network moments, interpret each subgroup as a dense interaction neighborhood. For transparency, we use the complete-graph proxy within subgroups (any sufficiently dense proxy yields the same scaling for density).

Network density. If each subgroup is a clique, then the number of within-subgroup edges is

$$E_{\text{in}}(n, s) = k \cdot \frac{s(s-1)}{2} = \frac{n(s-1)}{2}.$$

Total possible edges are $n(n-1)/2$, so density is

$$D(n) = \frac{E_{\text{in}}(n, s^*(n))}{n(n-1)/2} = \frac{s^*(n) - 1}{n - 1}. \tag{9}$$

Proposition 2 (Density and normalized degree decline with n)**.** *Under* (9) *and Proposition 1, network density $D(n)$ is decreasing in n for large n. Equivalently, average degree grows sublinearly while degree normalized by* $(n-1)$ *declines with n.*

Proof of Proposition 2. Under the clique-within-subgroup proxy,

$$D(n) = \frac{s^*(n) - 1}{n - 1}.$$

Treat n as continuous. For $n > 1$ in the interior region,

$$D'(n) = \frac{(n-1)s^{*\prime}(n) - (s^*(n) - 1)}{(n-1)^2}.$$

We show that the numerator is negative for all sufficiently large n. Using $\Delta(n) = \alpha^2 + 4c\delta n$ and $\sqrt{\Delta(n)} \geq 2\sqrt{c\delta n}$,

$$s^{*\prime}(n) = \frac{\delta}{\sqrt{\Delta(n)}} \leq \frac{\delta}{2\sqrt{c\delta n}} = \frac{1}{2}\sqrt{\frac{\delta}{cn}}.$$

Hence

$$(n-1)s^{*\prime}(n) \leq ns^{*\prime}(n) \leq \frac{1}{2}\sqrt{\frac{\delta n}{c}}.$$

Also, since $s^*(n) = \dfrac{\alpha + \sqrt{\Delta(n)}}{2c} \geq \dfrac{\sqrt{\Delta(n)}}{2c} \geq \sqrt{\dfrac{\delta n}{c}}$, we have $s^*(n) - 1 \geq \sqrt{\dfrac{\delta n}{c}} - 1$. Therefore the numerator satisfies

$$(n-1)s^{*\prime}(n) - (s^*(n) - 1) \leq \frac{1}{2}\sqrt{\frac{\delta n}{c}} - \left(\sqrt{\frac{\delta n}{c}} - 1\right) = 1 - \frac{1}{2}\sqrt{\frac{\delta n}{c}}.$$

The right-hand side is strictly negative whenever $\sqrt{\dfrac{\delta n}{c}} > 2$, i.e. for $n > 4c/\delta$. Thus there exists $\bar{n}$ (e.g. $\bar{n} = \max\{(\alpha + \delta)/c,\ 4c/\delta,\ 2\}$) such that $D'(n) < 0$ for all $n \geq \bar{n}$. This proves that density declines with team size for large n. Equivalently, average degree equals $D(n)(n-1) = s^*(n) - 1$, which grows sublinearly (on the order of $\sqrt{n}$), and degree normalized by $(n-1)$ declines. $\square$

Fragmentation / components. In the baseline, subgroups are disconnected, so the number of components equals the number of subgroups:

$$\mathrm{Comp}(n) = k(n) = \frac{n}{s^*(n)}. \tag{10}$$

Proposition 3 (Endogenous fragmentation rises with size). *The number of components Comp(n) = $n/s^*(n)$ is increasing in n for large n.*

Proof of Proposition 3. In the baseline (no bridges), the number of connected components equals the number of subgroups:

$$\mathrm{Comp}(n) = \frac{n}{s^*(n)}.$$

In the interior region, differentiate:

$$\mathrm{Comp}'(n) = \frac{s^*(n) - n s^{*\prime}(n)}{(s^*(n))^2}.$$

The denominator is positive. From the calculation in Proposition 1,

$$n s^{*\prime}(n) - s^*(n) = -\frac{\alpha\sqrt{\Delta(n)} + \alpha^2 + 2c\delta n}{2c\sqrt{\Delta(n)}} < 0,$$

so $s^*(n) - n s^{*\prime}(n) > 0$. Hence $\mathrm{Comp}'(n) > 0$ wherever the interior characterization applies. In particular, for all sufficiently large n, $\mathrm{Comp}(n)$ is strictly increasing. Moreover, using $s^*(n) \sim \sqrt{\delta n/c}$ implies $\mathrm{Comp}(n) \sim \sqrt{cn/\delta}$. □

Centralization. If within each subgroup interaction is mediated by a local coordinator (a star rather than a clique), then Freeman degree centralization declines in the number of subgroups k (and hence declines with n), because the organization has more local hubs rather than a single team-wide hub.

3.3 Performance and diffusion under fragmentation

3.3.1 Per-capita performance

Let $u^*(n) \equiv u(n, s^*(n))$ denote equilibrium per-capita performance when a team of size n chooses the optimal effective interaction neighborhood $s^*(n)$.

Proposition 4 (Net effect: scale can turn negative under endogenous fragmentation). *Per-capita income $u^*(n)$ eventually decreases with n. In particular, $u^*(n) \to -\infty$ as $n \to \infty$, so there exists $\bar{n}$ such that for all $n \geq \bar{n}$, $u^*(n)$ is strictly decreasing in n.*

Moreover, in the counterfactual without fragmentation costs ($\delta = 0$), the organization would optimally set $s = n$ and realize spillovers $\alpha \ln(n)$, highlighting that the negative size effect arises from the endogenous network response.

Proof of Proposition 4. Let $u^*(n) \equiv u(n, s^*(n))$ denote equilibrium per-capita income.

Step 1: eventual monotonic decline. For n large enough that the optimum is interior, the envelope theorem applies:

$$\frac{du^*(n)}{dn} = \frac{\partial u}{\partial n}(n, s^*(n)).$$

From $u(n,s) = y_0 + \alpha \ln s - c(s-1) - \delta(n/s - 1)$ we have

$$\frac{\partial u}{\partial n}(n,s) = -\frac{\delta}{s},$$

so in the interior region

$$u^{*\prime}(n) = -\frac{\delta}{s^*(n)} < 0.$$

Thus $u^*(n)$ is strictly decreasing for all n beyond the point where $s^*(n) < n$ (which holds for all sufficiently large n).

Step 2: $u^*(n) \to -\infty$. Use the explicit form $s^*(n) = \dfrac{\alpha + \sqrt{\alpha^2 + 4c\delta n}}{2c} \leq \dfrac{\alpha}{c} + \sqrt{\dfrac{\delta n}{c}}$, so for large n there exists a constant $K > 0$ such that $s^*(n) \leq K\sqrt{n}$. Then for n in that region,

$$u^{*\prime}(n) = -\frac{\delta}{s^*(n)} \leq -\frac{\delta}{K\sqrt{n}}.$$

Integrating from some n_0 in the interior region to n gives

$$u^*(n) \leq u^*(n_0) - \frac{\delta}{K}\int_{n_0}^{n} t^{-1/2}\, dt = u^*(n_0) - \frac{2\delta}{K}\left(\sqrt{n} - \sqrt{n_0}\right) \xrightarrow[n\to\infty]{} -\infty.$$

Therefore there exists $\bar{n}$ such that $u^*(n)$ is strictly decreasing for all $n \geq \bar{n}$.

Counterfactual $\delta = 0$. If $\delta = 0$, then $u(n,s) = y_0 + \alpha \ln s - c(s-1)$ depends on n only through the constraint $s \leq n$. Whenever $u(n,s)$ is increasing in s over $[1,n]$ (e.g. if $\alpha/s > c$ for all $s \leq n$), the optimum sets $s = n$, delivering spillovers $\alpha \ln n$ and showing that the negative size effect in the baseline arises from the endogenous fragmentation channel induced by $\delta > 0$. □

3.3.2 Technology diffusion

We now incorporate the information-seeding arm. Following the diffusion literature [Conley and Udry, 2010b, Banerjee et al., 2013b], assume that information diffuses fully within a subgroup but does not cross subgroups in the baseline (no bridges). Let m seeds be selected uniformly at random.

A given subgroup receives at least one seed with probability $1 - (1 - s/n)^m$, so expected adoption share is

$$A(n) = 1 - \left(1 - \frac{s^*(n)}{n}\right)^m. \tag{11}$$

Proposition 5 (Adoption falls with fragmentation; treatment effects can be non-linear)**.** *Since $s^*(n)/n$ declines with n (Proposition 1), adoption $A(n)$ is decreasing in n for large n.*

Let the per-adopter productivity gain be $\tau(n)$, potentially increasing in n if the technology complements scale/specialization [Becker and Murphy, 1992b]. Then the per-capita treatment effect

$$TE(n) = \tau(n) \cdot A(n)$$

can be non-monotone in n (e.g., hump-shaped) when $\tau(n)$ rises with n while $A(n)$ falls with n.

Proof of Proposition 5. Let $p(n) \equiv s^*(n)/n$ and recall the adoption share under m uniformly random seeds:

$$A(n) = 1 - (1 - p(n))^m .$$

Since

$$\frac{dA}{dp} = m(1-p)^{m-1} > 0,$$

$A(n)$ is strictly increasing in $p(n)$. By Proposition 1, $p(n)$ is decreasing in n for all sufficiently large n (indeed for all n in the interior region), hence $A(n)$ is decreasing in n for large n.

To see the asymptotics, Proposition 1 implies $p(n) \to 0$ and in fact $p(n) \sim \sqrt{\delta/(cn)}$. Thus for large n,

$$A(n) = 1 - (1 - p(n))^m \sim mp(n) \sim m\sqrt{\frac{\delta}{cn}},$$

so adoption falls to 0 at rate $n^{-1/2}$.

For treatment effects, $TE(n) = \tau(n)\, A(n)$. Differentiating yields

$$TE'(n) = \tau'(n)A(n) + \tau(n)A'(n),$$

where for large n we have $\tau'(n) \geq 0$ by assumption while $A'(n) \leq 0$ as shown above. Hence the sign of $TE'(n)$ is in general ambiguous, so $TE(n)$ can be non-monotone. A constructive example is $\tau(n) = \log n$, which is increasing; combined with $A(n) \asymp n^{-1/2}$ we obtain $TE(n) \asymp (\log n)/\sqrt{n}$, which is hump-shaped (increasing for small n and decreasing for large n). □

3.4 Extension: heterogeneous organizational capacity, hierarchy, and development

The threshold bridge-cost extension is useful for intuition, but it is too coarse for the quantitative exercise. Empirically, teams of the same randomized size exhibit substantial dispersion in connected components, bridge reach, leadership concentration, monitoring, and productivity. That dispersion is not random noise: the reduced-form evidence in Section 5 shows that baseline organizational capacity strongly predicts which teams remain integrated, and that this heterogeneity matters most in large teams. The quantitative model therefore extends the baseline framework in two directions. First, teams differ in both *organizational capacity* and *productive ability*. Second, bridging is continuous rather than all-or-nothing, so development can move a team gradually from a fragmented organization toward a hierarchy-like one.

Team heterogeneity. Each team g is characterized by two state variables. The first, z_g, is organizational capacity: the ability to keep communication from localizing too sharply, to sustain bridges across subgroups, and to make leadership and monitoring legible inside the team. The second, a_g, is productive ability: baseline output potential that is conceptually distinct from organization. The separation matters empirically. A team can be made up of members with good productive skills and still fail to coordinate at scale, or it can be organizationally strong without having unusually high baseline productive ability. The quantification therefore does not force one latent variable to explain both network structure and output.

Capacity-dependent subgrouping before bridges. Let $\mathrm{Comp}(n) = n/s^*(n)$ denote the baseline number of components implied by the common-parameter model in (10). In the heterogeneous model, that object becomes an *average benchmark* rather than a literal prediction for every team. Higher-capacity teams can start from less severe subgrouping even before explicit cross-subgroup bridges are sustained. We write the number of pre-bridge subgroup gaps as

$$\Gamma(n, z_g) = \big(\mathrm{Comp}(n) - 1\big)\psi(z_g), \qquad \psi'(z) < 0. \tag{12}$$

In the quantitative implementation we use an exponential form, $\psi(z) = \exp(-\phi z)$ up to a mean-preserving normalization, so that higher z_g compresses subgrouping while preserving the common-parameter benchmark on average. The implied pre-bridge effective neighborhood is

$$s^{pre}(n, z_g) = \frac{n}{1 + \Gamma(n, z_g)}. \tag{13}$$

Thus, heterogeneity in organizational capacity affects internal organization even before the team decides how many bridges to maintain: high-z_g teams begin from a less fragmented communication structure.

Convex bridge technology. Starting from $\Gamma(n, z_g)$ subgroup gaps, the team chooses a bridging intensity $\mu_g \in [0, 1]$, interpreted as the share of those gaps that are actively bridged. Let the per-team bridge-cost shifter be $\omega(z_g)$ with $\omega'(z) < 0$. With convex bridge-maintenance costs, per-member payoff becomes

$$u(n, z_g, \mu_g) = y_0 + \alpha \ln\big(s^{pre}(n, z_g)\big) - c\big(s^{pre}(n, z_g) - 1\big) - \delta\Gamma(n, z_g) + \Gamma(n, z_g)\left[\delta\mu_g - \frac{\omega(z_g)}{2n}\mu_g^2\right]. \tag{14}$$

The first line is the fragmented baseline evaluated at the pre-bridge neighborhood size. The bracketed term captures what bridges do. Each additional bridge reduces unresolved fragmentation losses, but maintaining many bridges is itself organizationally costly, and that cost is lower in teams with higher z_g.

Lemma 2 (Optimal bridging intensity under heterogeneous capacity)**.** *Fix* (n, z_g) *with* $\Gamma(n, z_g) > 0$. *Under* (14), *the objective is strictly concave in* $\mu_g \in [0, 1]$ *and the unique optimizer is*

$$\mu_g^*(n, z_g) = \min\left\{1,\ \frac{\delta n}{\omega(z_g)}\right\}. \tag{15}$$

Proof. The derivative of (14) with respect to μ_g is

$$\Gamma(n, z_g)\left[\delta - \frac{\omega(z_g)}{n}\mu_g\right],$$

and the second derivative is $-\Gamma(n, z_g)\omega(z_g)/n < 0$. Concavity gives a unique interior maximizer $\mu_g = \delta n/\omega(z_g)$, which is then clipped to the feasible interval $[0, 1]$. □

From fragmentation to hierarchy-like organization. Given (15), realized fragmentation is

$$\mathrm{Comp}^B(n, z_g) - 1 = (1 - \mu_g^*(n, z_g))\Gamma(n, z_g). \tag{16}$$

Higher organizational capacity therefore lowers fragmentation through two linked channels: it reduces the number of pre-bridge subgroup gaps and it raises the share of those gaps that can be bridged. In the empirical implementation, hierarchy is not modeled as a separate manager chosen in a distinct stage. Instead, it is a measured organizational outcome that becomes more likely when the team is less fragmented and more bridged. This is the economic transition we want the model to capture. Development does not need to create a formal manager instantaneously; it can instead make the team more integrated, more monitorable, and more legibly organized.

Observed productivity with organizational complementarity. The baseline model implies a fragmented payoff $u^*(n) = u(n, s^*(n))$. In the heterogeneous quantification, observed productivity is allowed to depend on both fragmentation relief and hierarchy-like organization. Let F_g denote the team's empirical fragmentation index and H_g its hierarchy index, with $\bar{F}_n$ and $\bar{H}_n$ denoting the corresponding size-specific baseline means. The empirical measurement equation used in the quantification is

$$y_g^{obs} = \nu_0 + a_g + u^*(n_g) + \beta_{f,n_g}(\bar{F}_{n_g} - F_g) + \beta_{fh}(\bar{F}_{n_g} - F_g)(H_g - \bar{H}_{n_g}) + \varepsilon_g. \tag{17}$$

This specification imposes a connected mechanism. Lower fragmentation has a direct productivity effect, and hierarchy matters only by amplifying the gains from lower fragmentation. Hierarchy is therefore not treated as a substitute for integration. The quantitative interpretation is: development first changes organization, and output rises when that organizational transformation makes a larger team easier to coordinate.

Proposition 6 (Development shifts teams from fragmentation toward hierarchy)**.** *Suppose* $\psi'(z) < 0$

and $\omega'(z) < 0$. Then higher organizational capacity z_g weakly raises $\mu_g^*(n, z_g)$ and weakly lowers realized fragmentation $\mathrm{Comp}^B(n, z_g)$. *Moreover, whenever H_g is increasing in bridge intensity and decreasing in residual fragmentation, development-induced increases in z_g move teams from more fragmented to more hierarchy-like organization.*

Proof. Equation (15) implies that $\mu_g^*(n, z_g)$ is weakly increasing in z_g because $\omega'(z) < 0$. Equation (12) implies that $\Gamma(n, z_g)$ is weakly decreasing in z_g because $\psi'(z) < 0$. Combining both results in (16) shows that realized fragmentation is weakly decreasing in z_g. If hierarchy-like organization is increasing in bridge intensity and decreasing in residual fragmentation, it follows immediately that higher z_g shifts the organization toward a more hierarchy-like state. □

Connection to the empirical analysis. The model speaks directly to the reduced-form sections of the paper. Proposition 4 rationalizes why Table 1 can show lower per-capita productivity in larger teams even when total production rises with size. Propositions 1–3 imply that larger teams should display lower normalized connectivity and more subgrouping in the interaction data, which is the first set of predictions we test in Section 5. The task-allocation evidence then evaluates whether these fragmented networks translate into weaker consensus and less effective specialization. Proposition 5 motivates the heterogeneous effects of the information-seeding treatment: if seeds are trapped inside narrow interaction neighborhoods, technology gains should be weaker or more uneven in larger teams. Finally, Proposition 6 gives a role to leadership and monitoring evidence as reduced-form proxies for whether some teams are able to create bridging roles or organizational layers that reconnect otherwise fragmented subgroups. The empirical sections therefore serve as reduced-form tests of the model's comparative statics, using specification (18) and its seed-interaction extension.

4 Data and Reduced-Form Evidence

We collected two sets of data: at the team level, at the individual level and at the network level between individuals. All our analysis is based on these information.

Production Team Data We envisioned the data collection for team-level outcomes to consist of in-person visits and weekly diaries to be completed by the teams, with the first in-person visit to take place approximately one month, and the second one four months after the training, respectively. Unfortunately, seven weeks after the start of the RCT, all field activities had to be halted because of the Omicron wave of Covid in India, forcing us to switch from in-person to phone-based data collection. Table A.1 provides an overview of how the production teams were tracked across time: after the start of the intervention, we managed to visit 49 teams in person to elicit production-related outcomes before Covid hit. For the remainder of the teams that we could not meet, the in-person visits were replaced by calls. We managed to reach by phone 153 more teams.[2] The Covid-induced lockdown lasted approximately 6 weeks; when restrictions were lifted, at the end of February 2022, we decided to meet in person all the groups we had interviewed by phone – we successfully met 144 out of 153 groups. Four months after the start of the intervention, we met all the teams again in person and carried out a team-level survey with at least 50% of team members being present. In all the visits, we asked teams whether they met and produced in the previous month. Column 2 of Table A.1 shows how many groups reported having produced within that time frame. Over the entire time period, 116 groups reported having produced soap within the whole duration, and 165 reported meeting, even if this does not necessarily mean that they produced soap.

Individual-level Data Two rounds of socio-economic surveys were carried out throughout the study: one at the start of the business training (baseline); the other one after four months (endline).[3] These surveys assessed women's socioeconomic profile, business skills, entrepreneurship, as well as decision-making power and wellbeing. Since the business training entailed a strong digital component, the surveys were also used to capture women's use of smartphones and digital technologies. We measured women's social connectedness by eliciting the extent of their social interactions. The endline survey contained questions to identify social interactions among training participants, both in person and through smartphones: how many times they met in person since the training was administered; whether and how many times they interacted via phone; what type of information they exchanged during in-person and digital interactions (e.g., purely business-related information, or also personal information). This allows us to track interactions within teams over time.

4.1 Summary Statistics and Balance Checks

Table A.2 presents summary statistics and balancing checks for the production teams included in the study. At the onset of the study, groups consisted of women of 33 years on average; more than 90% of group members were married; women had on average two children, all these characteristics being balanced across treatment arms. Overall, women in our sample appear fairly educated, with more than 60% of them having completed at least secondary school. In terms of economic activities, half of the women in our sample ran a business at the beginning of the study – this is not surprising as all the women involved in our study are part of women's collectives, through which they carry out small entrepreneurial activities –, while around 40% of them worked for a wage (either in agri or non-agri sector). Again these characteristics are all balanced at baseline. We instead find a slight imbalance in terms of per-capita household income and average business sales in the past 30 days. To account for these imbalances, we include these variables as controls in our main analysis.

[2] To cross-validate outcomes, two team members were called from each team to answer the same survey.
[3] The team-level survey was carried out at the end of the individual-level endline.

4.2 Empirical Analysis

We begin our analysis by estimating the following regression at the group level:

$$y_{pv} = \beta_0 + \beta_1 T_{1v} + \beta_2 T_{2v} + X_{pv} + \epsilon_{pv} \tag{18}$$

where y_{pv} is our team-level outcome of interest (e.g., number of soap bars produced; per capita productivity) for production team p in village v. T_1 and T_2 are dummies indicating whether the group is of small or medium size – the omitted dummy is for large-size teams.[4] We also include a vector of team-level characteristics X_{pv}. ϵ_{pv} is the error term. Standard errors are clustered at the village level.

Table 1 presents our main set of results in terms of duration of meetings and the production process, as well as the number of soap bars produced in total and per capita (PC) productivity. Panel A shows the main results for the full set of teams, irrespective of whether they met at least once to start production. Panel B instead presents results restricting the analysis to the groups who met at least once (but not necessarily ended up producing). Two main results stand out from this table: first, in terms of overall number of soap pieces produced, we find a positive relationship between group size and production (column 3). This is not surprising, as it simply reflects the larger dimension of production in larger groups. On the contrary, when we look at per-capita productivity in Column 4, we observe higher per-capita productivity in smaller teams than in larger teams: in the latter, women produce on average 1 soap bar per hour, whereas in small and medium-size teams they produce 62% and 33% more – the difference between small teams and large teams being also statistically significant. Results are similar, both in magnitude and significance, when we restrict our analysis to teams that met at least once (Panel B). In appendix D, we also include the results restricting the sample to teams that reported a positive production amount.

Overall, results from Table 1 suggest that the net effect between net gains from size are negative in this context. This could be completely mechanical if the production did not entail any advantage of large size. Yet, large teams could have specialized more or ripped the benefits of the information spread to one person being shared with more and going faster in the production. As a matter of fact, the most successful prodcution team was indeed a large group, named Muskan. We present their case study in Appendix section C.

The key question is therefore not simply whether larger teams produce less per worker, but *why* scale fails to translate into higher per-capita performance in this setting. Section 3 develops a conceptual framework in which larger teams endogenously break into smaller effective interaction neighborhoods, while a hierarchy extension clarifies when bridging roles or organizational layers can reconnect those subgroups. That framework is designed to map directly into the reduced-form evidence that follows. It predicts lower normalized connectivity and more fragmentation in larger

[4]Note that we do not have a *pure* control group where no training was administered. Given our main research questions – the impact of team size on productivity –all teams received the business training and we look at treatment effects across different group sizes.

teams, weaker consensus and task allocation when fragmentation is not offset by hierarchy, and heterogeneous technology diffusion depending on whether seeded information crosses subgroup boundaries. Section 5 then takes each of these predictions to the data.

Table 1: Treatment Effects on Production Process

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Panel A: Full Sample				
Small	-67.274**	-26.083*	-11.836**	0.512*
	(28.829)	(13.589)	(4.951)	(0.288)
Medium	-47.386	-17.268	-7.199	0.263
	(29.761)	(14.272)	(5.372)	(0.276)
Small = Medium	0.214	0.211	0.045	0.444
Mean (Large)	130.372	52.979	17.691	0.963
N	206	206	206	206
Panel B: Conditional on Meeting				
Small	-83.956**	-32.717*	-14.974**	0.660**
	(34.711)	(16.692)	(6.101)	(0.328)
Medium	-56.338	-20.554	-8.920	0.395
	(34.724)	(16.993)	(6.483)	(0.314)
Small = Medium	0.120	0.120	0.021	0.461
Mean (Large)	165.608	67.297	22.473	1.224
N	167	167	167	167

Note: Table 1 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. Controls are selected through LASSO. Robust standard errors (in parentheses) are clustered at the village level. Outcome variable definitions are in Table D.6. Meeting duration, production duration, and quantity produced are aggregated over all group meetings for each group. PC Productivity per hour is total quantity produced divided by meeting duration (in hours), divided by original group size. Sample size in Panel B corresponds to the number of groups that met but did not necessarily produce soap.

5 Endogenous Network Formation Inside Production Teams

Our main empirical results show that larger teams have lower per-capita productivity than medium and small teams. The central empirical question is therefore whether nominal scale actually expands the team's effective interaction neighborhood, or whether larger teams simply fragment into smaller communication units. This is the key distinction in the model. If growth fails to raise effective interaction size, then normalized connectivity should fall, subgrouping should increase, consensus-based task allocation should become harder, and the information-seeding treatment should become less effective unless some bridge nodes or organizational layers reconnect the team. Hsieh and Klenow [2009] show that, in the context of India and China, large firms face organizational constraints that can limit specialization. Our contribution is to test those organizational constraints in reduced form using direct measures of within-team interactions, labor allocation, monitoring, and technology diffusion.

The rest of this section follows that logic. We begin with the first-stage network evidence: do larger teams actually expand members' active interaction neighborhoods? We then move from sparsity to fragmentation by asking whether the missing links are disproportionately cross-subgroup links, so that larger teams break into smaller local clusters. Next, we study whether those fragmented networks are mirrored in weaker task allocation and consensus, and whether larger teams compensate by creating bridge nodes or simple hierarchical layers. Finally, we test whether the information seed is less effective when teams remain organizationally fragmented.

To test which of these margins dominates, we exploit the rich set of data we collected through our surveys, which contain specific targeted questions on team organization and labor division.

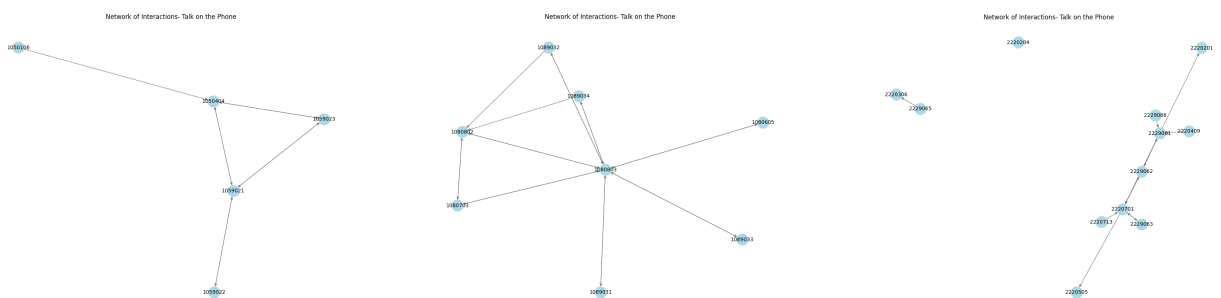
5.1 Measuring the Internal Organization of Production: Bilateral Networks

A central limitation in the empirical literature on team production is that the internal interaction architecture of teams is typically unobserved. Our data overcome this limitation by directly measuring bilateral interactions within each production team, allowing us to study how the internal network responds to scale and to evaluate the mechanism at the core of the model: whether larger teams remain integrated or endogenously fragment into subgroups.

This subsection asks whether nominal team size actually expands the effective interaction neighborhood inside the team. The model's key endogenous object is $s^*(n)$, the effective size of the set of coworkers with whom a member remains actively connected. A direct empirical proxy for that object is the number of same-team contacts implied by observed density, $\hat{s}(n) = 1 + \hat{d}(n)(n-1)$. Using the pooled network, this proxy is approximately 2.04 in small teams, 2.15 in medium teams, and 1.97 in large teams. In other words, as nominal team size more than doubles, the effective interaction neighborhood remains stuck at roughly two active contacts. That is the basic empirical fact behind the model's mechanism.

Our survey data let us see how that breakdown occurs. We ask team members to list with which other members they *i)* communicate over the phone, *ii)* communicate in person, *iii)* discuss production, and *iv)* ask for support during production. These bilateral reports let us observe not only whether larger teams are sparser, but whether communication becomes localized into subgroups. Although not all group members responded to the questionnaire, we find no differential attrition in responses by group size (see Table D.22 in the appendix).

On the Size and Shape of Networks The network of communication that was formed from each group differs by treatment. We use the variable "Talk Phone" indicating for each group member which people each one of them was in contact with during the survey period. We restrict our sample to groups where at least two respondents answered who they are connected to. Figure 4b shows an example of a graph created using the data on phone interactions.



(a) Graph of a Small Group (b) Graph of a Medium Group (c) Graph of a Large Group

Note: The graphs show an example of a different-sized group communication networks. Nodes represent members and links represent reported interactions, illustrating how within-group connectivity and network shape are measured in the analysis.

Table 2 reports the first-stage network evidence. Panel A shows that the raw number of reported contacts rises mechanically with group size. Panel B normalizes those contacts by team size and shows that members' active networks do not scale with headcount: normalized connectivity declines as teams get larger. This is the sense in which larger teams become effectively smaller than their nominal size would suggest. But lower density alone is not yet the mechanism. A sparse team can still function well if a few links bridge otherwise separate parts of the organization. The key question is therefore whether the missing links are disproportionately cross-subgroup links, so that larger teams fragment into smaller communication neighborhoods. That distinction matters because performance in complex tasks depends not just on how many links exist, but on whether they keep the team integrated [Sparrowe et al., 2001, Shaw, 1964].

The next question is whether larger teams compensate for weaker overall connectivity by generating bridge nodes or a broker core. The descriptive statistics below are useful because they speak directly to that possibility: they tell us whether some members emerge as especially central, especially connective, or especially important for linking otherwise distant parts of the network.

Definition (Degree centrality)**.** Degree centrality measures how connected a group member is, relative to group size. It gives us intuition on how active or visible a group member is, but does not lend us information on the importance of that node in the network.

$$DC(i) = \frac{d_i(g)}{n-1} \tag{19}$$

Definition (Freeman degree centralization)**.** Freeman degree centralization measures at the group level the extent to which the network structure is dominated by a single member. In effect, it compares the actual distribution of connections to a star graph, where one member is connected to everyone and all others are connected to only that leader.

$$H_D = \frac{\sum_{i=1}^{n}[C_{D,max} - C_D(i)]}{(n-1)(n-2)} \tag{20}$$

where $C_{D,max}$ is the highest observed degree centrality in the group, and $C_D(i)$ is the degree centrality of member i. A value of 1 represents a perfect star configuration, while a value of 0 indicates a perfectly regular graph where every member has the same number of connections.

Definition (Betweenness centrality)**.** Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes. A ratio close to 1 indicates that a member lies on most of the shortest paths connecting any other 2 members:

$$CB(i) = \frac{2}{(n-1)(n-2)} \sum_{s \neq i \neq t \in G} \frac{\sigma(s,t|i)}{\sigma(s,t)} \tag{21}$$

where G is the set of group members, $\sigma(s,t)$ is the number of shortest (s,t)-paths, and $\sigma(s,t|v)$ is the number of those paths passing through some member v other than s,t. If $s = t$, $\sigma(s,t) = 1$, and if $v \in \{s,t\}$, $\sigma(s,t|v) = 0$.

Definition (Closeness centrality)**.** Closeness centrality is the inverse of the average length of the shortest paths from a node to all other nodes in the network.

$$C(i) = C(i) = \frac{n-1}{\sum_{j \neq i} d(i,j)} \tag{22}$$

where $l(p,u)$ indicates the shortest path distance between u,p nodes.

Table 3 speaks directly to whether larger teams generate a compensating bridge architecture. If that were happening, we would expect larger teams to exhibit stronger brokerage or a more centralized core. We do not find that pattern. Relative to large teams, small and medium teams display higher degree centrality, higher betweenness centrality, and higher closeness centrality. Put differently, larger teams are not only less connected; they also fail to generate a stronger broker core that could reconnect otherwise separate subgroups. This is the network counterpart of the paper's hierarchy argument: bridging capacity appears limited precisely where nominal size is largest. Appendix Table D.24 reinforces the same point using complementary bridge-reach measures: large teams have lower top-k centrality share and lower largest-component coverage, consistent with weaker cross-subgroup reach.

Definition (Number of cliques)**.** A clique is a subset of nodes where every node is directly connected to every other node in that subset. [5]

Definition (Maximum clique size)**.** The number of members in the largest maximal clique found in the network. It gives us the largest possible fully-connected sub-unit.

Definition (Average clustering coefficient)**.** For each node i, the local clustering coefficient is the ratio of existing links between i's neighbors to the total possible links between them. It gives us a measure of how tightly knit a group is.

The average clustering is:

$$C = \frac{1}{n} \sum_{i \in G} \frac{2e_i}{k_i(k_i - 1)} \tag{23}$$

where k_i is the degree of node i and e_i is the number of edges between those neighbors.

Table 4 moves from sparsity to actual subgrouping. Larger groups display many more cliques, which is exactly what we would expect if communication localizes into smaller subsets of coworkers rather than expanding across the whole team. At the same time, the maximum clique size does not grow with group size beyond the smallest teams, suggesting that the size of the local "inner circle" remains bounded even as nominal headcount rises. The clustering results reinforce the same point: ties become concentrated within tightly knit local circles rather than across the team as a whole. Combined with the network figures above, this makes the subgroup interpretation visible in the data. Larger teams are not merely communicating less; they are operating through smaller, more weakly connected local clusters. Appendix Table D.24 shows the same pattern through additional subgroup statistics: large teams contain more communities, higher isolate shares, longer within-component path lengths, and higher modularity than small teams.

A fragmentation-based mechanism test. The network evidence is informative only if fragmentation also predicts what teams do and how much they produce. In the overlap sample used in the quantification, a one-unit increase in the empirical fragmentation index is associated with a 0.48 decline in per-capita productivity. A simple IV that instruments fragmentation with the randomized size dummies yields a coefficient of -0.94, with a first-stage F-statistic of 54.03. We interpret these estimates cautiously as the effect of size-induced fragmentation rather than as a fully

[5] We use the Maximal Clique definition, meaning adding another node that "breaks" a closed circle is null. Individual nodes are not considered a clique.

structural causal parameter. But they show that the size treatment moves performance precisely through the network object highlighted by the model. The rest of this section traces that same link through organization, hierarchy, and diffusion outcomes.

Baseline capacity and heterogeneous fragmentation. The network mechanism also implies that the cost of scale need not be the same for all teams. Using the baseline survey collected before random assignment, we build a standardized team-level capacity index from 21 member characteristics spanning education, business exposure, phone access and WhatsApp use, SHG participation, self-reported leadership, and autonomy. We first standardize these inputs at the individual level, average them into an individual organizational-capacity score, and then average across members; Appendix D.5 explains the construction in detail and shows that this team-mean measure predicts bridge-like network outcomes more strongly than top-member alternatives. Because team size is randomized, the interaction between assigned size and baseline capacity can be interpreted as variation in teams' ex ante ability to sustain bridges rather than endogenous sorting into larger groups. Appendix Table D.25 shows that higher baseline capacity predicts substantially less fragmentation and greater bridge reach. The large-team slopes are economically the steepest throughout, and the pairwise equality tests show that this amplification is statistically clearest for connected components: the capacity slope in 12-member teams is more negative than in 5-member teams ($p = 0.008$) and than in 8-member teams ($p = 0.044$). In block fixed-effect specifications with village-clustered standard errors, a one-standard-deviation increase in baseline capacity raises the bridge index by 0.126, 0.113, and 0.237 in teams of size 5, 8, and 12, respectively, while reducing the number of connected components by 0.254, 0.471, and 1.430. The implied high-minus-low capacity gap is especially large among large teams: the predicted bridge index is 0.475 higher and the predicted number of connected components is 2.860 lower. This is important conceptually because it shows that the penalty from nominal scale is not mechanical. Larger teams can remain substantially more integrated when they begin with higher organizational capacity. It is also the empirical reason the quantification in Section 7 treats organizational capacity as a team-level state variable rather than as noise around the average fragmentation model.

5.2 Fragmentation and Task Division

We study how teams organize themselves for production by asking how they decide who-does-what: whether assignments follow members' ability, individual preferences, or group consensus, and whether members in practice do all tasks or instead work in subgroups. These outcomes are the most direct organizational counterpart of the model's fragmentation cost. If larger teams split into smaller communication neighborhoods, task assignment should become less consensus-based and more duplicative. That is exactly the pattern in Table 5. We do not observe a statistically meaningful difference in whether tasks are assigned by ability (Column 1). But small and medium teams are more likely to assign tasks following preferences and, especially, group consensus. Small groups are also substantially less likely to report that "everyone does all tasks," which is precisely what one would expect when communication remains integrated enough for members to specialize. Together with the subgroup-work outcome, the results suggest that fragmentation is not merely a network description: it translates into weaker coordination and incomplete specialization in larger teams. This is the organizational channel through which the model's fragmentation cost shows up in the data.

5.3 Bridge Nodes, Monitoring, and Hierarchy

The centrality evidence already suggests that large teams do not spontaneously generate a stronger broker core. The hierarchy extension in Section 3 implies that size need not reduce performance if teams can create low-cost bridges across subgroups. We do not observe formal hierarchies directly, but monitoring and leadership variables provide reduced-form evidence on whether some organizational layer emerges as teams grow. Appendix Table D.23 adds supportive evidence on this margin: smaller and medium teams are more likely than large teams to converge on who the organizer is, and members in those teams place more concentrated organizer and "talked-to-most" nominations on the same individuals. We therefore look at the extent of monitoring in group production in Table 6. We find weak support for the view that larger teams compensate for fragmentation by building effective organizational layers. Panel A of column (1) shows that small groups are less likely than larger groups to use monitoring. Column (2) adds to findings in column (1) by showing that numbers of monitors per capita are higher in medium groups than in large groups. Taken together with the betweenness and centralization results above, this leads us to rule out the hypothesis that smaller groups perform better simply because their members are easier to monitor; the deeper issue is that larger groups do not appear to create enough bridging structure to remain integrated.

Panel A. Raw Number of Connections

Table 2: Number of Team Connections

	Talk on Phone	Meet in Person	Talk Production	Support in Production
	(1)	(2)	(3)	(4)
Small	-0.192**	-1.030***	-0.856**	-0.253
	(0.096)	(0.167)	(0.396)	(0.157)
Medium	0.156	-0.380*	-0.420	0.043
	(0.110)	(0.199)	(0.419)	(0.190)
Small = Medium	0.002	0.000	0.081	0.061
Mean (Large)	1.179	2.85	1.44	.589
N	1044	1063	821	812

Panel B. Connections Normalized by Group Size

	Talk on Phone	Meet in Person	Talk Production	Support in Production
	(1)	(2)	(3)	(4)
Small	0.101***	0.126***	-0.005	0.020
	(0.015)	(0.017)	(0.038)	(0.019)
Medium	0.070***	0.072***	0.004	0.029
	(0.013)	(0.020)	(0.040)	(0.020)
Small = Medium	0.084	0.007	0.790	0.672
Mean (Large)	.098	.237	.12	.049
N	1044	1063	821	812

Note: Table 2 reports estimates of specification (18). Panel A reports the raw number of group members each group member has interacted with (talked on the phone, met in person, talked during production, or received support from during production). Panel B reports the same outcomes normalized by group size. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in D.6. Sample size differs by individuals that received the question in the survey.

Table 3: Centrality Measures

	Degree Centrality	Betweenness Centrality	Closeness Centrality
	(1)	(2)	(3)
Small	0.343***	0.041***	0.186***
	(0.033)	(0.007)	(0.024)
Medium	0.149***	0.018***	0.090***
	(0.027)	(0.005)	(0.023)
Small = Medium	0.000	0.003	0.001
Mean (Large)	0.177	0.017	0.128
N	1126	1126	1126

Note: Table 3 is descriptive and does not estimate a regression specification. It reports team-level centrality measures constructed from members' reports on who was most active, most creative, most hardworking, most organized, most frequently contacted, and most frequently learned from.

Table 4: Cliquishness and Clustering Measures

	Nb. Cliques	Max Clique Size	Average Clustering
	(1)	(2)	(3)
Small	-3.974***	-0.314***	0.044
	(0.444)	(0.113)	(0.041)
Medium	-1.956***	0.010	0.071*
	(0.492)	(0.124)	(0.038)
Small = Medium	0.000	0.003	0.553
Mean (Large)	6.463	2.701	0.171
N	1126	1126	1126

Note: Table 4 reports the estimates of specification (18). We construc thte indepenent variables starting from the survey questionnaire. The survey asked group members to list the members who had the best organization skills, were most active in the group, were most creative, hardworking, the ones whom they talked to most and learned from the most.

Table 5: Treatment Effects on Production Organization

	Ability	Individual Preferences	Consensus	Everyone Does all Tasks	Work in Subgroups
	(1)	(2)	(3)	(4)	(5)
Panel A: Full Sample					
Small	0.055	0.076**	0.076**	-0.330***	-0.060
	(0.038)	(0.035)	(0.030)	(0.092)	(0.050)
Medium	0.033	0.010	-0.005	-0.039	0.071
	(0.042)	(0.021)	(0.026)	(0.103)	(0.065)
Small = Medium	0.606	0.133	0.048	0.001	0.038
Mean (Large)	0.021	0	0	0.66	0.067
N	206	206	206	206	192
Panel B: Conditional on Meeting					
Small	0.067	0.085**	0.092**	-0.433***	-0.093
	(0.052)	(0.042)	(0.039)	(0.100)	(0.062)
Medium	0.044	-0.004	-0.007	-0.075	0.067
	(0.057)	(0.030)	(0.031)	(0.100)	(0.076)
Small = Medium	0.665	0.115	0.049	0.000	0.035
Mean (Large)	0.027	0	0	0.838	0.081
N	167	167	167	167	163

Note: Table 5 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Variables in columns (1)-(4) take on the value 1 if a group divides tasks according to ability, individual preferences, consensus, or by assigning to all tasks to everyone, respectively, and 0 otherwise. Column (5)'s "Work in Subgroups" takes the value 1 if the group divided in order to produce. Outcome variable definitions are available in D.6. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

Table 6: Treatment Effects on Monitoring Production

	Use Monitoring	Monitors by Group Size	Has Group Leader
	(1)	(2)	(3)
Panel A: Full Sample			
Small	-0.040	0.024	0.015
	(0.080)	(0.017)	(0.049)
Medium	0.091	0.036*	-0.001
	(0.082)	(0.020)	(0.057)
Small = Medium	0.088	0.643	0.723
Mean (Large)	.191	.019	.933
N	206	201	193
Panel B: Conditional on Meeting			
Small	-0.012	0.026	0.026
	(0.095)	(0.021)	(0.047)
Medium	0.123	0.043*	-0.011
	(0.094)	(0.023)	(0.059)
Small = Medium	0.128	0.563	0.342
Mean (Large)	0.189	0.023	0.946
N	167	166	164

Note: Table 6 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in D.6. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

Additional evidence on leadership consensus and subgroup structure. Appendix D.4 strengthens this interpretation using mechanism evidence that is not in the main tables. First, smaller and medium teams exhibit significantly higher convergence around who the organizer is than large teams, suggesting that fragmentation makes leadership less legible precisely where scale is greatest. Second, additional network summaries not emphasized in the main text show that large teams contain more communities, higher isolate shares, lower top-k centrality share, lower largest-component coverage, and longer within-component path lengths. These appendix results reinforce the same point as the main network tables: larger teams are not only less connected on average, but also less able to organize clear bridge roles across internally separated subgroups.

5.4 Fragmentation and Technology Diffusion

The evidence so far is consistent with the hypothesis that coordination and communication costs prevented further labor specialization and, to some extent, functional organizational layers in larger groups. The diffusion treatment adds a sharper test. In the model, the relevant object for diffusion is the effective network share $s^*(n)/n$: if that share falls with size, seeded information will travel through a smaller fraction of the team even when the technology itself is valuable. The theory section therefore predicts that the impact of the seed should depend not only on the technology itself but also on whether the team remains connected enough for that information to reach beyond the initially treated members. According to classical theories like Becker and Murphy [1992b], access to new technology helps to increase the gains from specialization and, thus, the costs of coordination and communication. We create an entire randomization leg to test this hypothesis and further validate this mechanism. As described above, via phone, we explained a new technology for wrapping soap bars to two team members. To test our hypothesis, we run the following specification:

$$\begin{aligned} y_{pv} = \beta_0 + \beta_1 T_{1v} + \beta_2 T_{2v} + \beta_3 InfoSeeds_{pv} + \\ \beta_4 T_{1v} \times InfoSeeds_{pv} + \beta_5 T_{2v} \times InfoSeeds_{pv} + \beta_6 T_{3v} \times InfoSeeds_{pv} + \epsilon_{pv} \end{aligned} \quad (24)$$

where $InfoSeeds_{pv}$ denotes the random assignment to receive additional information related to cloth wrapping, and $T_{1v} \times InfoSeeds_{pv}$, $T_{2v} \times InfoSeeds_{pv}$ and $T_{3v} \times InfoSeeds_{pv}$ represent the interaction between receiving the information seed and different group sizes, respectively. Table 7 reports the results of specification 24. Column (4) shows that among teams that did not receive the additional information, smaller teams produced more soap bars per capita and time. Among those that received the information, medium-size teams did significantly better than medium-size teams that did not receive it. We do not finds similar results for large groups.

Table 7: Treatment Effects on Production Process- Technology Diffusion

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Small	-66.330	-27.712	-12.440	0.754*
	(43.754)	(21.501)	(7.803)	(0.382)
Medium	-99.840**	-40.681*	-15.269*	-0.201
	(45.447)	(22.950)	(8.580)	(0.310)
Small × Info Seeds	-15.277	-9.420	-2.730	-0.165
	(17.471)	(7.663)	(2.077)	(0.451)
Medium × Info Seeds	76.391***	27.900***	10.144***	1.093***
	(23.161)	(10.428)	(3.296)	(0.338)
Large × Info Seeds	-35.386	-21.214	-6.421	0.192
	(51.106)	(23.936)	(9.034)	(0.359)
Small = Medium	0.098	0.170	0.327	0.018
Small × Info Seeds = Medium × Info Seeds	0.002	0.006	0.002	0.023
Small × Info Seeds = Large × Info Seeds	0.711	0.635	0.683	0.531
Medium × Info Seeds = Large × Info Seeds	0.050	0.068	0.094	0.065
N	205	205	205	205

Note: Table 7 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in D.6. Sample size $N = 205$ corresponds to the full sample.

We can test our mechanism even more directly to observe whether among groups that received the new technology, medium groups tend to organize the task more efficiently. The results are striking as shown in table 8. When the info seed is given, the relationships with the division of labor stay the same for the groups that did not receive it. But it reverts for small and medium groups that did receive, and it improves for the large groups that did receive it. This confirms our current hypothesis that when technology improves, coordination costs get relatively smaller than the gains of specialization, expanding the size of the groups as a result.

Table 8: New Technology and Task Division among Different Size Teams

	Ability	Individual Preferences	Consensus	Everyone Does all Tasks	Work in Subgroups
	(1)	(2)	(3)	(4)	(5)
Small	0.059	0.071	0.166*	-0.319***	-0.076
	(0.059)	(0.066)	(0.087)	(0.108)	(0.079)
Medium	-0.025	-0.042	0.026	-0.141	0.222
	(0.053)	(0.032)	(0.045)	(0.143)	(0.137)
Small × Info Seeds	0.036	-0.095	-0.016	-0.152	0.098*
	(0.094)	(0.063)	(0.087)	(0.136)	(0.057)
Medium × Info Seeds	0.052	0.045	-0.000	0.095	-0.137
	(0.079)	(0.051)	(0.038)	(0.157)	(0.138)
Large × Info Seeds	0.087	-0.010	-0.028	-0.117	0.153
	(0.083)	(0.026)	(0.037)	(0.110)	(0.129)
Small = Medium	0.241	0.089	0.104	0.288	0.039
Small × Info Seeds = Medium × Info Seeds	0.896	0.112	0.870	0.254	0.142
Small × Info Seeds = Large × Info Seeds	0.687	0.244	0.898	0.842	0.694
Medium × Info Seeds = Large × Info Seeds	0.778	0.357	0.584	0.264	0.137
N	116	116	116	116	115

Note: Table 8 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in D.6. Sample size $N = 116$ corresponds to groups that recorded producing soap.

6 Robustness and Alternative Explanations

This section addresses alternative explanations and measurement concerns that could confound the interpretation of our main findings. The experiment was designed to hold constant key inputs and the production technology across teams, so that differences in performance can be attributed to the organizational consequences of scale and the endogenous interaction network. Nonetheless, team size could plausibly correlate with (i) exposure to external constraints (e.g., input procurement, marketing opportunities, or quality problems), (ii) differences in participation and reporting that mechanically affect productivity measures, or (iii) survey non-response that could bias measured network density. We therefore complement the main analysis with a set of robustness exercises that probe each of these channels.

External constraints: inputs, marketing, and quality A first concern is that the productivity differences we document could be driven by external constraints that interact with group size rather than by internal organization. We attempted to neutralize some of these issues by design: groups received the same production training and were provided with standardized inputs in quantities proportional to assigned group size. Still, teams could have faced different frictions in sourcing additional inputs, marketing their soap, or maintaining product quality. To assess this possibility, we collected direct measures of these external barriers. Appendix Table D.13 reports treatment effects on (i) marketing difficulties, (ii) input procurement issues, and (iii) soap quality problems. We do not find systematic evidence that smaller teams face fewer external barriers; if anything, medium teams report slightly fewer procurement issues than large teams. Overall, the pattern of coefficients suggests that external constraints are unlikely to be the main driver of the per-capita productivity ranking across team sizes.

Communication technology and phone access Because part of our interaction data is based on phone communication, differential access to phones could mechanically affect measured connectivity. Appendix Table D.14 shows that assignment to different group sizes does not differentially affect a broad index of decision-making autonomy. At the same time, members assigned to small and medium groups report somewhat greater phone access at endline relative to members assigned to large groups. Two considerations mitigate the concern that this drives the network and diffusion results. First, our interaction measures include both phone and in-person links, and the qualitative decline in normalized connectivity with group size appears across interaction modalities. Second, as we show below, survey response intensity is not systematically different across treatments, reducing the scope for differential measurement error.

Outcome construction, participation, and alternative productivity measures Our primary productivity measure is per-capita output per hour, computed by aggregating meeting-level production logs and dividing total output by total meeting duration and assigned group size (Table 1).

A mechanical concern is that assigned size may differ from effective labor input if attendance varies systematically with team size. Two pieces of evidence speak to this issue. First, Panel B of Table 1 shows that the productivity ranking is similar when restricting the sample to teams that met at least once, ruling out that the main result is driven by extensive-margin differences in whether teams activate at all. Second, Appendix Table D.8 documents meeting frequency and attendance patterns; while larger teams mechanically have more members present, attendance rates relative to assigned size do not display sharp differences across treatments.

We further check that our conclusions are robust to alternative definitions of productivity and to sample restrictions that reduce reporting noise. Appendix Table D.9 reports alternative measures including productivity per hour (not per-capita), per-capita production using realized attendance instead of assigned size, and high-quality output measures. Appendix Table D.7 restricts attention to teams with positive production. Across these definitions and restrictions, the qualitative pattern remains: smaller teams tend to outperform larger teams on a per-capita basis. Finally, Appendix Table D.10 shows consistent patterns when focusing on production outcomes measured during monitored visits, alleviating concerns that self-reported production logs drive the main results.

Covid interruption and survey modality A key feature of the fieldwork is that the Omicron wave forced a temporary shift from in-person monitoring to phone-based data collection. While we later returned to in-person surveys, differential exposure to this disruption could affect outcomes and measurement. In Appendix D, we therefore replicate the main specifications controlling not only for whether a group reports being affected by Covid, but also for whether the group was reached through a replacement phone call during the lockdown period. Tables D.15–D.20 show that controlling for this survey modality does not materially change the estimated size effects on production, organization, monitoring, diffusion, or external barriers.

Network measurement and non-response Our network outcomes are constructed from roster-style questions in which respondents list within-team contacts (phone communication, in-person communication, production discussions, and requests for support). A concern is that larger teams may have lower survey response intensity, mechanically lowering measured connectivity and density. We address this in two ways. First, in the main network analysis we focus on normalized measures (e.g., degree divided by team size) and we restrict attention to teams with sufficient respondent coverage. Second, Appendix Table D.22 tests directly whether the *number of nominated contacts*, normalized by team size, differs across treatments. We do not detect systematic differences in response intensity by assigned group size, suggesting that the documented decline in normalized connectivity with team size reflects genuine changes in interaction patterns rather than differential survey effort.

Unintended effects and wellbeing Finally, to assess whether the intervention affected well-being in ways that could indirectly influence performance (e.g., through stress or conflict), Appendix

Table D.12 reports treatment effects on mental health, financial stress, expectations, and physical illness. We find little evidence of adverse effects over the study horizon. This supports the interpretation that the performance differences across team sizes primarily reflect organizational and network mechanisms rather than changes in wellbeing.

Overall, the robustness exercises reinforce the central message of the paper: the negative relationship between nominal team size and per-capita productivity is not easily explained by external constraints, reporting artifacts, or differential survey response. Instead, it aligns with the mechanism emphasized throughout the paper—endogenous changes in internal interaction networks and coordination as teams grow.

7 Quantifying the Network Mechanism: Heterogeneity, Hierarchy, and Development

This section does two distinct jobs. First, it calibrates the paper's baseline fragmentation model to recover the *average* network mechanism in the data. Second, it extends that benchmark to allow for heterogeneity in organizational capacity and productive ability, because the reduced-form evidence shows clearly that teams of the same randomized size are not organizationally identical. The heterogeneous model is the paper's preferred quantitative object. It is the one that lets us ask the development question in a disciplined way: not whether a common scalar parameter improves for everyone at once, but whether development shifts the *distribution* of organizational capacity and therefore changes which teams remain fragmented, which teams become bridged, and which teams move toward hierarchy-like organization.

7.1 Baseline calibration of the average network mechanism

We begin with the common-parameter model because it provides the average benchmark from which the heterogeneous exercise departs. Concretely, we calibrate the common technology parameters (α, c, δ, y_0) to match three moments on the overlap sample of teams for which both network measures and productivity are observed: pooled network density, the number of connected components, and per-capita live productivity. These are the moments that map most directly into equations (9), (10), and Proposition 4. Table 9 separates the moments directly targeted in the objective from non-targeted validation moments.

The resulting estimates are $\alpha = 4.47$, $c = 2.48$, $\delta = 0.30$, and $y_0 = 1.95$. The implied effective interaction neighborhoods are small and strikingly flat: $s^*(n)$ rises only from 2.09 to 2.23 to 2.40 as nominal team size rises from 5 to 8 to 12. Put differently, the effective network share falls from 0.42 to 0.28 to 0.20 as teams get larger. A direct empirical proxy built from observed pooled density, $\hat{s}(n) = 1 + \hat{d}(n)(n-1)$, is similarly flat at 2.04, 2.15, and 1.97. This is the core quantitative fact behind the paper: nominal scale expands much faster than effective interaction scale.

Table 9: Targeted and Non-Targeted Moments in the Baseline Quantification

Moment	Small ($n = 5$)	Medium ($n = 8$)	Large ($n = 12$)
Panel A. Targeted moments			
Network density: data	0.261	0.164	0.088
Network density: model	0.273	0.176	0.127
Connected components: data	2.079	2.942	4.795
Connected components: model	2.392	3.583	4.996
Per-capita productivity: data	2.162	1.727	1.157
Per-capita productivity: model	2.127	1.715	1.202
Panel B. Non-targeted validation moments			
Fragmentation index: data	-0.433	0.045	0.639
Fragmentation index: model-implied	-0.437	0.053	0.635
Modularity: data	0.074	0.162	0.226
Modularity: model-implied	0.081	0.150	0.232
Average path length in LCC: data	1.348	1.609	1.837
Average path length in LCC: model-implied	1.361	1.584	1.848
Largest-component share: data	0.757	0.731	0.655
Largest-component share: model-implied	0.764	0.717	0.662

Note: Table 9 separates the moments directly targeted in the baseline calibration from non-targeted validation moments. Panel A reports the three moment families used in the method-of-moments objective: pooled network density, the number of connected components, and per-capita live productivity. Panel B reports mechanism-relevant moments that are not directly targeted: the fragmentation index, modularity, average path length within the largest connected component, and the largest-component share. The data moments are computed on the overlap sample of teams for which both network measures and per-capita live productivity are observed. The model-implied values in Panel B are obtained by mapping the calibrated model's fragmentation object into these empirical moments.

The identifying variation in this common-parameter exercise is straightforward. Network density and the number of connected components jointly discipline the effective interaction neighborhood implied by the model, while per-capita productivity identifies the payoff parameters conditional on that network object. In particular, y_0 is pinned down by the overall level of productivity, whereas α, c, and δ are identified jointly through the tension between observed sparsity and fragmentation on the one hand and the decline in per-capita productivity with team size on the other. We view this benchmark as the correct *average* model of the network cost of scale. Its limitation is that it is silent about why some teams of the same randomized size remain much more integrated than others.

7.2 Why the baseline quantification is not enough

That omission matters. The mechanism section showed that within each size arm there is substantial heterogeneity in fragmentation, bridge reach, leadership concentration, monitoring, and productivity. More importantly, that heterogeneity is systematic rather than idiosyncratic. Teams that start with higher baseline organizational capacity fragment less, and this heterogeneity matters most in large teams. A quantification that treats every 12-member team as organizationally identical would therefore miss the central development margin in the data.

The heterogeneous quantification addresses that problem by introducing two distinct pre-treatment objects. The first is an *organizational-capacity* index, meant to capture ex ante ability to keep a team integrated as scale rises. The second is a *productive-ability* index, meant to capture output potential that is conceptually distinct from coordination. This separation is economically important. If the model used a single latent type to explain both network organization and output, it would mechanically attribute every productivity difference to "management" or every network difference to "skill." The baseline survey lets us do better than that.

7.3 Constructing the pre-treatment heterogeneity indices

The baseline survey covers 1,547 women who are later assigned to the 206 teams in the experiment. No post-treatment network or productivity outcome enters the construction of the indices. For each baseline input x_{ij}, we winsorize at the 1st and 99th percentiles and standardize in the full baseline sample, obtaining $\tilde{x}_{ij}$. We then build individual-level scores by averaging standardized inputs within economically motivated groups of variables.

Let $\mathcal{J}_Z$ denote the set of organizational-capacity inputs and $\mathcal{J}_A$ the set of productive-ability inputs. The individual scores are

$$Z_i = \frac{1}{|\mathcal{J}_Z|} \sum_{j \in \mathcal{J}_Z} \tilde{x}_{ij}, \qquad A_i = \frac{1}{|\mathcal{J}_A|} \sum_{j \in \mathcal{J}_A} \tilde{x}_{ij}. \tag{25}$$

The organizational-capacity inputs are: phone access, phone ownership, phone use, WhatsApp-group access, WhatsApp use, active SHG participation, inclusiveness of SHG decisions, comfort speaking in SHG meetings, phone and in-person communication inside the SHG, proactive behavior, multitasking, leadership, free choice, and work, food, and financial autonomy. The productive-ability inputs are: grade completed, business ownership, vocational training, business knowledge, creativity, and hardworking.

We then aggregate those individual scores to the team level. For organizational capacity, we compare the team mean, the top-two average, and the top member. The team mean predicts bridge-like network outcomes best: its average incremental R^2 relative to size-only controls is 0.0487, compared with 0.0418 for the top-two average and 0.0353 for the top member. We therefore define the team organizational-capacity index as

$$Z_g = \frac{1}{n_g} \sum_{i \in g} Z_i. \tag{26}$$

For productive ability, the top member performs best once organizational capacity is controlled for: its average incremental R^2 is 0.0318, compared with 0.0199 for the top-two average and 0.0089 for the team mean. We therefore define the team productive-ability index as

$$A_g = \max_{i \in g} A_i. \tag{27}$$

This difference in aggregators is informative. It says that organizational capacity looks like a team-wide environment, whereas productive ability looks more like a scarce upper-tail trait. In the estimation sample, the interquartile range is roughly $[-0.60, 0.68]$ for organizational capacity and $[-0.76, 0.82]$ for productive ability.

7.4 Stagewise estimation of the heterogeneous model

The heterogeneous model is estimated in two stages. The separation is deliberate. Stage 1 uses the network moments to identify how organizational capacity changes internal structure. Stage 2 then maps that simulated organization into observed productivity. This ordering prevents noisy productivity moments from determining the network block on their own and keeps the interpretation of organizational capacity tied to the paper's distinctive empirical contribution: direct observation of within-team networks.

Stage 1: organizational capacity, subgrouping, and bridging. Let z_g denote the team's organizational-capacity state. In the empirical implementation, z_g equals the standardized organizational-capacity index plus residual latent heterogeneity. Given z_g, the model allows organizational capacity to affect internal structure in two distinct ways. First, it lowers the number of subgroup gaps before bridges are created. Second, it lowers the cost of sustaining bridges across those gaps. Concretely, the estimated network block uses

$$\Gamma_g(n) = \big(\mathrm{Comp}(n) - 1\big) \exp(-\phi z_g), \tag{28}$$

$$\omega_g = \omega_0 \exp(-\omega_1 z_g), \tag{29}$$

$$\mu_g^* = \min\left\{1, \frac{\delta n_g}{\omega_g}\right\}, \tag{30}$$

$$\mathrm{Comp}_g^B - 1 = (1 - \mu_g^*)\Gamma_g(n_g). \tag{31}$$

This structure is the quantitative counterpart of the theory extension in Section 3. Development-like increases in organizational capacity can reduce fragmentation even before explicit bridging, and then reduce it further by making bridges easier to sustain.

The stage-1 moment vector targets the mean and standard deviation of connected components by randomized size and by organizational-capacity half. We use halves rather than terciles because the large-team upper tail is thin and otherwise makes the $n = 12$ cells too noisy. Largest-component share, isolate share, and the bridge index are treated as validation moments. This choice keeps the main identifying content transparent: the network block is being asked to explain who fragments and by how much, conditional on randomized size and ex ante organizational capacity.

The resulting estimates are $\omega_0 = 299.52$, $\omega_1 = 1.92$, $\sigma_z = 3.00$, and $\phi = 0.146$. These values imply that organizational capacity materially changes both the pre-bridge subgroup structure and the ease of reconnecting it. The validation equations track the main network outcomes reasonably well, with R^2 values of 0.903 for largest-component share, 0.862 for isolate share, and 0.841 for the bridge index. We do not interpret those fits mechanically as proof that every detail of the network is captured. We interpret them as evidence that the heterogeneous bridging model is recovering the right organizational variation: teams with higher z_g start from less severe subgrouping and are better able to reconnect the remaining gaps.

Stage 2: productive ability and the fragmentation–hierarchy complementarity. The second stage takes the simulated organization from Stage 1 as given and maps it into observed productivity. The key restriction is economic, not merely statistical: hierarchy is not allowed to substitute mechanically for fragmentation. Instead, lower fragmentation has a direct productivity effect, and hierarchy matters only by amplifying that effect. Let F_g denote the empirical fragmentation index used throughout the paper—the standardized average of connected components minus one, disconnectedness of the largest component, modularity, and average path length within the largest connected component. Let H_g denote a hierarchy index built from four observed organizational margins: concentration of skill on a focal leader, concentration of organizer nominations on a focal leader, organizer-nomination convergence, and the number of monitors. The estimated measurement equation is

$$y_g^{obs} = \nu_0 + a_g + u^*(n_g) + \beta_{f,n_g}(\bar{F}_{n_g} - F_g) + \beta_{fh}(\bar{F}_{n_g} - F_g)(H_g - \bar{H}_{n_g}) + \varepsilon_g. \tag{32}$$

This equation is the paper's preferred economic specification because it encodes the mechanism directly: a less fragmented team is more productive, and hierarchy raises productivity only when it helps the team exploit lower fragmentation.

The stage-2 moment vector uses four families of moments: mean productivity by randomized size and productive-ability half, the standard deviation of productivity by size, the within-size slope of productivity on fragmentation, and the within-size slope of productivity on hierarchy. The estimated intercept is $\nu_0 = 1.915$, the productive-ability coefficient is $\beta_a \approx 0$, the direct fragmentation effects are $\beta_{f,5} = 0.050$, $\beta_{f,8} = 0.025$, and $\beta_{f,12} = 0.020$, and the complementarity term is $\beta_{fh} = 2.288$. The direct productivity effect of lower fragmentation is therefore modest on its own, but the interaction with hierarchy-like organization is economically meaningful. That is precisely the interpretation we want: development changes organization first, and hierarchy determines how much of those organizational gains are converted into output.

This constrained complementarity specification fits less tightly than a looser fragmentation-only benchmark. We retain it anyway because the looser benchmark fits by shutting down the hierarchy margin altogether. The paper's economic question is not whether one can match productivity moments most flexibly with a reduced-form fragmentation slope. It is whether development moves teams from fragmentation toward hierarchy-like organization and whether that organizational transition matters for output. On that question, the constrained specification is the right benchmark, even if it sacrifices some in-sample fit.

7.5 Development as a shift in organizational capacity

With the heterogeneous model in hand, development is no longer represented as a common scalar reduction in fragmentation. Instead, it is represented as a shift in the distribution of organizational capacity. We implement two counterfactuals. The first is an internal benchmark that shifts teams toward the within-India organizational frontier. The second is an externally anchored benchmark that shifts teams toward a US-like distribution of management quality using public data from the World Management Survey (WMS). In both exercises we hold productive ability fixed, keep (α, c, δ) and the structural objects from the estimated $v13$ model fixed, and change only the organizational-capacity input. The interpretation is therefore deliberately narrow: this is a pure organizational-development counterfactual, not a full US productivity counterfactual.

For the within-India frontier, if Z_g is the estimated team-level organizational-capacity index, we replace it with

$$Z_g^{cf} = F_q^{-1}(F_{India}(Z_g)),$$

where $q \in \{0.75, 0.90\}$ indexes the top quartile or top decile of the observed Indian team distribution. This preserves ranks and asks what changes if all teams are exposed to the organizational conditions achieved by the upper tail of teams in the same environment.

For the external benchmark, we use the public WMS manufacturing sample for 2004–2015. Restricting to establishments with nonmissing management scores yields 711 Indian plants and 953 US plants. The average WMS management score is 2.527 in India with standard deviation 0.679, and 3.285 in the United States with standard deviation 0.612. Let m denote the WMS management score. We map WMS into the units of the estimated team-level organizational-capacity index using the India WMS moments:

$$z^{WMS} = \frac{m - \mu_{India}^{WMS}}{\sigma_{India}^{WMS}}.$$

This normalization is the right one for the present exercise because the model input Z_g is already standardized on the Indian SHG sample. After this transformation, the mapped India WMS distribution has mean zero and standard deviation one by construction, while the mapped US distribution has mean 1.116 and standard deviation 0.901. We then apply rank-preserving quantile mapping,

$$Z_g^{US} = F_{WMS,US}^{-1}(F_{India}(Z_g)) \,,$$

so that each Indian team keeps its percentile rank while facing a US-like organizational-capacity environment. No US productivity data are used at this stage. The productivity changes reported below are entirely model-implied, conditional on the Indian production environment and on the estimated link from organization to output in equation (32).

Table 10: Development Counterfactuals in the Heterogeneous Model

Scenario and outcome	Small ($n = 5$)	Medium ($n = 8$)	Large ($n = 12$)
Panel A. Within-India top quartile frontier			
Δ disconnected subgroups	-0.291	-0.651	-0.954
Δ bridge index	+0.096	+0.183	+0.227
Δ hierarchy index	+0.005	+0.011	+0.016
Δ per-capita productivity (%)	+0.18%	+0.28%	+0.69%
Panel B. Within-India top decile frontier			
Δ disconnected subgroups	-0.402	-0.865	-1.321
Δ bridge index	+0.134	+0.247	+0.321
Δ hierarchy index	+0.007	+0.015	+0.022
Δ per-capita productivity (%)	+0.27%	+0.45%	+1.10%
Panel C. WMS-US benchmark			
Δ disconnected subgroups	-0.281	-0.552	-0.867
Δ bridge index	+0.093	+0.156	+0.206
Δ hierarchy index	+0.005	+0.010	+0.015
Δ per-capita productivity (%)	+0.18%	+0.26%	+0.63%

Note: Panels A and B replace the observed organizational-capacity distribution with the within-India top quartile or top decile distribution, preserving team ranks and holding productive ability fixed. Panel C replaces the observed organizational-capacity distribution with a US-like benchmark constructed from the public WMS manufacturing sample, after mapping WMS scores into the units of the estimated organizational-capacity index using India WMS moments. "Disconnected subgroups" is the mean change in connected components minus one. The hierarchy index is a standardized average of leader-skill concentration, organizer-nomination concentration, organizer consensus, and the number of monitors. Productivity changes are model-implied percent changes under the preferred heterogeneous specification; they are not direct moments from US output data.

Table 10 now separates the internal and external development benchmarks. The within-India top-decile frontier remains the paper's upper-tail benchmark: it asks what happens when teams are shifted to the best organizational conditions already observed in the same environment. The WMS-US benchmark is less aggressive, but it is externally anchored and therefore more portable. Quantitatively, the mapped US benchmark is broader and less compressed than the within-India top decile. In the SHG sample, the within-India top-decile mapping has mean 1.719 and standard deviation 0.357, whereas the mapped WMS-US benchmark has mean 1.129 and standard deviation 0.894. Put differently, the US benchmark raises organizational capacity substantially relative to the Indian baseline, but it does not push almost every team into the extreme upper tail of the Indian distribution.

The two counterfactuals deliver the same qualitative result. Development changes internal organization much more than it changes productivity levels on impact, and those organizational gains are disproportionately large in bigger teams. Under the within-India top-decile frontier, the average 12-member team sheds 1.32 disconnected subgroups, raises the bridge index by 0.32, becomes more hierarchy-like, and raises per-capita productivity by 1.10%. Under the WMS-US benchmark, the average 12-member team still sheds 0.87 disconnected subgroups, raises the bridge index by 0.21, raises the hierarchy index by 0.015, and raises per-capita productivity by 0.63%. The external benchmark therefore delivers roughly two-thirds of the organizational gains and about 58% of the productivity gain generated by the within-India top-decile frontier in large teams.

This pattern is the main quantitative implication of the paper. Development first changes organization. It makes large teams less segmented, more bridged, and more legibly coordinated. Output then rises through that organizational transformation. The quantified gains therefore should not be read as saying that development instantly turns large teams into dramatically more productive units. They should be read as saying that development relaxes the organizational constraint on scale. The biggest immediate adjustment occurs in the network and hierarchy margins themselves. In that sense, the quantification sharpens the paper's broader development claim: when organizational capacity is scarce, growth generates fragmentation; when organizational capacity improves, larger teams move away from segmentation and toward the hierarchy-like coordination needed to make scale productive.

8 Conclusion

This paper studies a new mechanism through which scale can backfire in team production. The standard intuition is that larger teams should perform better because they allow more specialization, more knowledge sharing, and greater scope for adopting improved practices. Our results show why this need not happen. In a field experiment with newly created women's production teams in rural India, we randomly assign team size and seed a subset of members with information about an improved production technique. Larger teams produce more in total, but they are less productive per worker. The central explanation is that, as nominal team size increases, members do not expand their active interactions one-for-one. Instead, the internal network becomes more localized: communication is concentrated within smaller neighborhoods, cross-member links become scarcer, and the team starts to operate as a collection of subgroups rather than as a single integrated production unit. Making a production team bigger can therefore make it worse when effective interaction networks fail to scale with headcount.

This mechanism organizes the rest of the evidence in the paper. Larger teams are less connected relative to their size and display patterns consistent with endogenous subgroup formation. These changes in internal network structure are mirrored in organizational outcomes: larger teams are less able to reach consensus, less likely to allocate tasks in ways consistent with effective specialization, and less able to translate the information-seeding intervention into comparable productivity gains. The paper's claim is therefore sharper than a generic statement that coordination is harder in bigger groups. The point is that scale changes who interacts with whom inside the team, and that this endogenous reorganization of communication is what undermines the gains from size.

The broader implication is that diseconomies of scale may emerge before formal hierarchy or monitoring becomes the primary constraint. In low-capacity environments, organizations may remain small not only because supervision is limited, but because they struggle to sustain the cross-subgroup bridges required to keep communication integrated as they grow. When those bridges are missing, specialization, knowledge diffusion, and technology adoption all become harder to realize. This also helps explain why some large teams succeed while the average effect of size is negative: productive scale requires not just more workers, but an internal organization that preserves connectivity across them.

More generally, the results suggest that policies or managerial interventions should focus on the communication architecture of teams, not only on their size or formal structure. If the relevant barrier is endogenous subgroup formation, then creating bridging roles, strengthening ties across interaction neighborhoods, or ensuring that new information reaches disconnected parts of the team may matter more than simply adding members or intensifying monitoring. Quantifying these margins remains an important next step. But even without a fully quantified model, the reduced-form and network evidence point to a clear conclusion: the costs of scale can arise through the endogenous internal network of interaction, and this mechanism can be a first-order limit to productive growth.

References

Leila Agha, Keith Marzilli Ericson, Kimberley H Geissler, and James B Rebitzer. Team formation and performance: Evidence from healthcare referral networks. Technical report, National Bureau of Economic Research, 2018.

Venkatesh Bala and Sanjeev Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229, 2000a.

Venkatesh Bala and Sanjeev Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229, 2000b. doi: 10.1111/1468-0262.00155.

Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social incentives in the workplace. *The review of economic studies*, 77(2):417–458, 2010.

Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013a.

Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013b.

Vittorio Bassi, Jung Hyuk Lee, Alessandra Peter, Tommaso Porzio, Ritwika Sen, and Esau Tugume. Self-employment within the firm. Technical report, National Bureau of Economic Research, 2023.

Gary S Becker and Kevin M Murphy. The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, 107(4):1137–1160, 1992a.

Gary S Becker and Kevin M Murphy. The division of labor, coordination costs, and knowledge. *The Quarterly journal of economics*, 107(4):1137–1160, 1992b.

Nicholas Bloom and John Van Reenen. Why do management practices differ across firms and countries? *Journal of Economic Perspectives*, 24(1):203–224, 2010.

Nicholas Bloom, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. Does management matter? evidence from india. *The Quarterly Journal of Economics*, 128(1):1–51, 2013. doi: 10.1093/qje/qjs044.

Nicholas Bloom, Luis Garicano, Raffaella Sadun, and John Van Reenen. The distinct effects of information technology and communication technology on firm organization. *Management Science*, 60(12):2859–2885, 2014. doi: 10.1287/mnsc.2014.2013.

Patrick Bolton and Mathias Dewatripont. The firm as a communication network. *The Quarterly Journal of Economics*, 109(4):809–839, 1994. doi: 10.2307/2118349.

Stéphane Bonhomme. Teams: Heterogeneity, sorting, and complementarity. *arXiv preprint arXiv:2102.01802*, 2021.

Wyatt Brooks, Kevin Donovan, and Terence R Johnson. Mentors or teachers? microenterprise training in kenya. *American Economic Journal: Applied Economics*, 10(4):196–221, 2018.

Wyatt Brooks, Kevin Donovan, and Terence R Johnson. From micro to macro in an equilibrium diffusion model. *Globalization*, 2020.

Miriam Bruhn, Dean Karlan, and Antoinette Schoar. The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in mexico. *Journal of Political Economy*, 126(2):635–687, 2018. doi: 10.1086/696154.

Jing Cai and Adam Szeidl. Interfirm relationships and business performance. *The Quarterly Journal of Economics*, 133(3):1229–1282, 2018.

Alfred Dupont Chandler, Takashi Hikino, and Alfred D Chandler. *Scale and scope: The dynamics of industrial capitalism*. Harvard University Press, 2009.

Marc Claveria-Mayol, Pau Milán, and Nicolás Oviedo-Dávila. Incentive contracts and peer effects in the workplace, 2024. arXiv:2406.11712; revised 2025.

Timothy G Conley and Christopher R Udry. Learning about a new technology: Pineapple in ghana. *American Economic Review*, 100(1):35–69, 2010a.

Timothy G Conley and Christopher R Udry. Learning about a new technology: Pineapple in ghana. *American Economic Review*, 100(1):35–69, 2010b.

B Deloitte. Global human capital trends 2016, the new organization: different by design, 2016.

Wouter Dessein and Tano Santos. Adaptive organizations. *Journal of Political Economy*, 114(5): 956–995, 2006. doi: 10.1086/508031.

Kevin Devereux. Identifying the value of teamwork: Application to professional tennis. Technical report, Working Paper Series, 2018.

Niklas Engbom, Hannes Malmberg, Tommaso Porzio, Federico Rossi, and Todd Schoellman. Economic development according to chandler. NBER Working Paper 34483, National Bureau of Economic Research, 2025.

Luis Garicano. Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, 108(5):874–904, 2000a.

Luis Garicano. Hierarchies and the organization of knowledge in production. *Journal of political economy*, 108(5):874–904, 2000b.

Luis Garicano and Esteban Rossi-Hansberg. Organization and inequality in a knowledge economy. *The Quarterly Journal of Economics*, 121(4):1383–1435, 2006a.

Luis Garicano and Esteban Rossi-Hansberg. Organization and inequality in a knowledge economy. *The Quarterly journal of economics*, 121(4):1383–1435, 2006b.

Thomas J. Holmes and John J. Stevens. An alternative theory of the plant size distribution, with geography and intra- and international trade. *Journal of Political Economy*, 122(2):369–421, 2014. doi: 10.1086/674633.

Bengt Holmstrom. Moral hazard in teams. *The Bell journal of economics*, pages 324–340, 1982.

Chang-Tai Hsieh and Peter J Klenow. Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448, 2009.

Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, Princeton, NJ, 2008.

Matthew O Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74, 1996a.

Matthew O. Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74, 1996b. doi: 10.1006/jeth.1996.0108.

Robert E. Lucas. On the size distribution of business firms. *The Bell Journal of Economics*, 9(2): 508–523, 1978. doi: 10.2307/3003596.

Alexandre Mas and Enrico Moretti. Peers at work. *American Economic Review*, 99(1):112–45, 2009.

Michael J. Piore and Charles F. Sabel. *The Second Industrial Divide: Possibilities for Prosperity*. Basic Books, New York, 1984.

Marvin E Shaw. Communication networks. In *Advances in experimental social psychology*, volume 1, pages 111–147. Elsevier, 1964.

Raymond T Sparrowe, Robert C Liden, Sandy J Wayne, and Maria L Kraimer. Social networks and the performance of individuals and groups. *Academy of management journal*, 44(2):316–325, 2001.

Xiaokai Yang and Jeff Borland. A microeconomic mechanism for economic growth. *Journal of Political Economy*, 99(3):460–482, 1991. doi: 10.1086/261762.

Appendix
Table of Content

A Data Description

This section describes the data sources used in the study, how information was collected, and how raw records were consolidated for analysis. It documents the monitoring rounds, the harmonization of group responses, and the main data-processing decisions taken before estimation. It also reports balance and attrition checks used to assess sample comparability across treatment arms.

A.1 Data Collection and Organization

There was a gap of 10 days (6th Jan - 16th Jan) for making necessary changes to the survey instrument to shift from In-person Visits to Replacement Calls. Replacement Calls was launched on 16th Jan. In these 10 days, no group received In-person Visits or Calls.

We called any two women from the group we had of baseline - similar to the protocol followed in the call involving the information seed treatment.

Given that there are two respondents from each group, information on each meeting had to be harmonized to be collapsed into one observation per group.

Steps:

1. Exclude single-response observations and variables that are identifiers, member lists and categorical variables, and reshape them separately
2. Reshape the data to wide at TG level to have responses from 2 respondents in wide format
3. **Correcting inconsistencies** in responses between respondents for numeric and dummy variables
   (a) **For numerical variables:**
      i. Impute missing values based on responses of respondent that reported more group meetings
      ii. If both respondents reported meeting equally, take the means of the non-missing response of either respondent based on mismatch response in numeric variables
   (b) **For categorical variables:**
4. Merge back the identifiers, categorical variables and member lists to the corrected dataset. Then merge back single-response observations

Once the replacement call data was collapsed to one observation per group, we reshaped all the meeting data to wide to compare values of respondents, pick out any inconsistencies with meeting dates across replacement calls, in-person visits, and the second in-person visit, and merge all 3 datasets together.

Table A.1: Team Visits

	N. teams	N. teams that met at least once
First Visit (one month after BL)	49	29
Replacement Calls	*153*	*60*
First Visit (two and half months after BL)	144	13
Second Visit (four months after BL)	201	59

Note: Table A.1 is descriptive and does not estimate a regression specification. Column 1 reports the number of teams surveyed in each visit; Column 2 reports the number of teams that reported meeting for production in the relevant preceding time period in each round, respectively. We managed to visit only 49 teams one month after baseline because of the Covid Omicron outbreak. During the Covid lockdown, we replaced in-person visits with phone-based interviews, managing to reach out remotely to 153 teams. Once the Covid restrictions were lifted, we met in person 144 out of the 153 teams previously reached out via phone. The second visit was performed in person to all teams (we successfully met 201 teams).

A.2 Balance Checks and Attrition Tables

This subsection reports baseline comparability and differential attrition across treatment arms. It first presents group-level balance checks and differences by meeting likelihood. It then reports attrition estimates to assess whether sample retention differs systematically by treatment status.

Table A.2: Balance Checks - Group Level

Variable	N	(1) Small Mean/SE	N	(2) Medium Mean/SE	N	(3) Large Mean/SE	T-test Difference (1)-(2)	T-test Difference (1)-(3)	T-test Difference (2)-(3)
Share married	96	90.729 (1.504)	62	91.729 (1.154)	47	92.561 (0.949)	-1.000	-1.832	-0.832
Average respondent's age	96	33.960 (0.367)	62	33.185 (0.391)	47	33.887 (0.473)	0.775	0.074	-0.701
Average HH size	96	5.875 (0.107)	62	5.738 (0.131)	47	6.054 (0.198)	0.136	-0.180	-0.316
Average nb children	96	1.904 (0.057)	62	1.981 (0.053)	47	2.004 (0.064)	-0.077	-0.100	-0.023
Share general caste	96	3.958 (0.977)	62	2.110 (0.609)	47	3.766 (0.941)	1.848	0.192	-1.656
Share scheduled tribe and caste	96	38.229 (3.248)	62	34.884 (2.635)	47	36.729 (3.644)	3.346	1.500	-1.846
Share other backward classes	96	57.604 (2.835)	62	62.827 (2.835)	47	58.263 (3.893)	-5.223	-0.659	4.564
Completed at least secondary edu	96	63.646 (2.390)	62	65.654 (2.437)	47	63.265 (2.421)	-2.008	0.381	2.389
Share working for wage	96	44.306 (3.063)	62	46.756 (2.664)	47	44.780 (2.861)	-2.451	-0.475	1.976
Share running business	96	53.785 (2.513)	62	59.243 (3.165)	47	53.468 (3.422)	-5.458	0.317	5.775
Mean PC HH monthly income	96	2046.593 (197.254)	62	1419.656 (85.755)	47	1622.172 (150.250)	626.937***	424.420	-202.516
Mean wage last month	96	698.944 (61.951)	62	618.538 (57.446)	47	637.975 (61.004)	80.406	60.970	-19.437
Mean farm income last season	96	21564.250 (3509.945)	62	20893.859 (2096.889)	47	25762.989 (2868.252)	670.391	-4198.739	-4869.130
Mean business sales last month	96	1381.253 (212.778)	62	2346.692 (450.458)	47	1125.683 (211.969)	-965.439**	255.570	1221.009**
Mean PC HH overall monthly expenditure	96	7.028 (0.065)	62	6.966 (0.038)	47	6.947 (0.054)	0.061	0.081	0.020
Mean amount saved last month	96	1076.624 (105.031)	62	1016.465 (135.703)	47	987.206 (136.909)	60.160	89.419	29.259
Mean borrowed amount last 12m	96	29740.123 (2981.706)	62	26950.303 (2803.970)	47	28449.263 (2922.331)	2789.820	1290.860	-1498.961
Mean owning land	96	95.243 (1.046)	62	93.844 (1.152)	47	96.384 (0.831)	1.399	-1.141	-2.540**
Mean asset index KLK	96	-0.001 (0.034)	62	-0.061 (0.034)	47	-0.034 (0.038)	0.060	0.033	-0.027

Note: Table A.2 is descriptive and does not estimate a regression specification. It reports baseline group-level balance checks across small, medium, and large teams, presenting means, standard errors, and pairwise differences for demographic and economic characteristics used to assess ex-ante comparability across treatment arms.

Table A.3: Differences across Group, by likelihood of meeting

	(1) Did not Meet		(2) Met		(1)-(2) Pairwise t-test	
Variable	N/Clusters	Mean/(SE)	N/Clusters	Mean/(SE)	N/Clusters	Mean difference
Share married	39 / 36	89.808 (1.746)	165 / 124	91.929 (1.025)	204 / 145	-2.122
Share never married	39 / 36	1.154 (0.672)	165 / 124	1.707 (0.582)	204 / 145	-0.553
Share divorced or separeted	39 / 36	1.175 (0.577)	165 / 124	0.566 (0.235)	204 / 145	0.610
Share widowed	39 / 36	7.863 (1.513)	165 / 124	5.798 (0.745)	204 / 145	2.065
Share general caste	39 / 36	3.226 (1.333)	165 / 124	3.460 (0.787)	204 / 145	-0.233
Share schedules caste	39 / 36	11.004 (2.448)	165 / 124	12.384 (1.739)	204 / 145	-1.380
Share schedules tribe	39 / 36	23.462 (3.927)	165 / 124	24.863 (2.046)	204 / 145	-1.401
Share OBC	39 / 36	62.094 (4.419)	165 / 124	58.794 (2.406)	204 / 145	3.300
Mean age	39 / 36	33.468 (0.622)	165 / 124	33.816 (0.304)	204 / 145	-0.348
Mean household size	39 / 36	5.767 (0.190)	165 / 124	5.895 (0.097)	204 / 145	-0.128
Mean Nb children	39 / 36	1.921 (0.098)	165 / 124	1.943 (0.044)	204 / 145	-0.022

Note: Table A.3 is descriptive and does not estimate a regression specification. It reports differences across groups by likelihood of meeting.

Table A.4: Balance Checks - Group Meeting

Variable	(1) Did not Meet N/Clusters	(1) Did not Meet Mean/(SE)	(2) Met N/Clusters	(2) Met Mean/(SE)	(1)-(2) Pairwise t-test N/Clusters	(1)-(2) Pairwise t-test Mean difference
Share no education	39	8.120	165	7.611	204	0.509
	36	(2.048)	124	(1.067)	145	
Share primary education	39	17.137	165	18.362	204	-1.226
	36	(2.582)	124	(1.369)	145	
Share secondary education	39	28.654	165	27.868	204	0.786
	36	(2.441)	124	(1.472)	145	
Share high School	39	41.368	165	39.124	204	2.243
	36	(3.978)	124	(1.729)	145	
Share graduate education	39	3.654	165	4.550	204	-0.896
	36	(1.112)	124	(0.635)	145	
Share post grad. education	39	1.068	165	2.409	204	-1.341
	36	(0.611)	124	(0.555)	145	
Share works for wage	39	42.714	165	45.494	204	-2.780
	36	(3.064)	124	(1.807)	145	
Share runs business	39	50.940	165	56.441	204	-5.501
	36	(3.594)	124	(2.002)	145	
Mean income top coded at p99	39	6894.324	165	7991.354	204	-1097.030
	36	(889.693)	124	(438.876)	145	
Mean farm income uncdt	39	21754.908	165	25273.529	204	-3518.620
	36	(4557.164)	124	(2875.550)	145	
Mean business sales uncdt	39	6190.304	165	2401.504	204	3788.799
	36	(4814.742)	124	(454.896)	145	

Note: Table A.4 is descriptive and does not estimate a regression specification. It reports additional balance checks by meeting status, focusing on education, labor participation, and baseline income-related variables. Significance: ***=.01, **=.05, *=.1. Errors are clustered at the village level.

Table A.5: Differential Attrition at Endline: Ever met

VARIABLES	(1) Model 1	(2) Model 2
Small	-0.085	-0.090
	(0.071)	(0.073)
Large	-0.006	-0.006
	(0.082)	(0.083)
Mean grade completed		0.006
		(0.018)
Mean age		0.003
		(0.010)
Mean HH income		0.000
		(0.000)
Observations	201	201
R-squared	0.009	0.010
F-test Model 1	0.309	
F-test Model 2		0.290

Note: Table A.5 reports estimates of specification (18). It reports attrition rates at endline comparing groups of different size and controls for average team-member demographics. Robust standard errors are in parentheses. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

B Soap Production Process

This section summarizes the five-step soap production protocol taught during the intervention and highlights the operational requirements of each stage. It explains why timing, monitoring, and sequencing matter for production quality and consistency in team-based production. It also clarifies why task allocation and coordination are central for comparing performance across group sizes.

Although the product itself is simple, the production process is not a single undifferentiated task. The training decomposed soap making into a sequence of linked stages that require both accuracy and timing: workers must first prepare the soap base and any required water, then melt the base to the right consistency, quickly mix in color and fragrance, pour the liquid into molds without creating bubbles or spillage, and finally finish the bars so they cure uniformly and can be packaged for sale. Errors at an early stage propagate to later ones. If the base is overheated, if additives are not mixed at the right moment, or if the liquid is poured too slowly, the final bars can become uneven, unattractive, or difficult to sell. Even in a standardized production environment, quality therefore depends on whether the team can execute a tightly sequenced routine.

This structure is useful for the paper because it makes clear why internal organization matters even when the technology is relatively low-tech. Some steps can be divided across members, but only if the group maintains agreement about sequence, timing, and responsibility. Once the base is melted, the available window for adding ingredients and pouring into molds is short, so miscommunication can generate waste or duplication very quickly. The process therefore creates a natural role for monitoring and task assignment: someone has to keep track of measurements, someone has to handle melting and mixing, and someone has to ensure that molds, additives, and packaging materials are ready at the right moment. This is exactly why the soap-setting technology is informative for our purposes. It is standardized enough that all teams face the same production frontier, but it still requires enough coordination that differences in communication, specialization, and monitoring show up clearly in observed productivity.

The figure below summarizes the production routine taught during training. It should be read not merely as a technical recipe, but as the operational backbone of the experiment. Because all teams were instructed to follow the same sequence, differences across teams are less likely to reflect product heterogeneity or different production methods, and more likely to reflect how well members coordinated around a common production protocol.

1. Preparation

Measure base. Add water if brand requires.

↓

2. Melting

Microwave in mug (60-75s)
until smooth and liquid.

↓

3. Additives

Quickly add color (1-2 drops) and scent (1-2%).

↓

4. Molding

Pour into mold using glass
rod; tilt mug close to surface.

↓

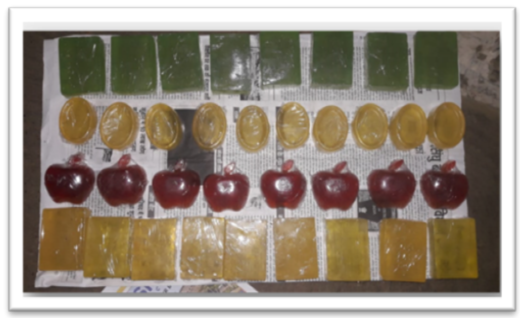
5. Finishing

Fill to the brim and apply IPA spray for bubbles.

C A Success Case Study in a Large Team: Muskan

This section presents a qualitative case study of Muskan, a large team that performed particularly well during the intervention. It describes the group's organization, production routines, and use of digital tools for coordination, communication, and sales. It is useful for the paper not as a generic success story, but as a large-team case that appears to have solved the two central problems emphasized in the main analysis: coordination across many members and the emergence of a hierarchy-like organizing layer. The discussion below draws on project case-study notes prepared after the intervention.

(a) Finished soaps prepared for sale.



(b) Finished soap bars in different shapes and colors.

Figure 5: Finished soaps from Muskan project follow-up documentation.

Muskan operated in a rural village in Chhattisgarh and is useful for the paper because it shows that the negative average size effect is not mechanical: a large team can do well when it sustains communication, solves coordination problems early, and makes leadership legible enough for the group to remain integrated.

The case-study notes report that Muskan began with 12 women randomly assembled from five SHGs for the soap-making training on 21 December 2021, and that 8 members remained active throughout the production phase. Since these women combined household responsibilities with other livelihoods, the first major challenge was coordinating meeting times and keeping the group active. What stands out in the notes is that the team solved this coordination problem through repeated communication rather than letting the larger group split into disconnected subgroups. Members created a WhatsApp group immediately after training, discussed schedules there, and selected meeting dates by consensus. When COVID restrictions made in-person meetings difficult, decisions continued through the same WhatsApp discussions and phone conversations instead of stalling production.

> *"The soap-making video shared during the training is watched by all members before coming to any production meeting."*
> — *Muskan member, case-study interview*

This communication structure appears to have gone together with a clearly identified focal organizer. The notes repeatedly describe decisions as collective, but they also indicate that one member consistently spoke on behalf of the group in the follow-up interview, which is exactly what one would expect when a large team has succeeded in making leadership clear rather than diffused. That focal role was complemented by stable organizational routines. All transactions, meeting details, and group activities were recorded in a diary. Absence from meetings triggered a fine of Rs. 100. Members who sold soap purchased it from the group at a common agreed price and were not allowed to undercut or raise that price. Taken together, these features suggest that Muskan combined consensus-based communication with a clearly legible leader, explicit rules, and written records.

> *"We impose a fine of Rs. 100 if any member is absent from any of the meetings. Thus, we ensure active participation from all."*
> — *Muskan member, case-study interview*

The production numbers are also substantial. Since training, the group held 20 meetings, including 15 production meetings, produced 852 soaps, and earned roughly Rs. 10,600, which they reinvested in raw materials. The notes report that they spent about Rs. 15,840 on raw materials from a vendor in Uttarakhand using pooled soap-sale revenue and personal contributions. Quality control also appears to have been unusually disciplined. Members repeatedly returned to the soap-making video shared on their phones during training whenever problems arose in a production stage, which helped them follow the same procedures and maintain quality standards across meetings.

Digital tools were central not only for internal coordination but also for outreach and marketing. Mobile phones were used to contact dealers and buyers; WhatsApp and Facebook were used to advertise the soaps; and a substantial share of orders came from people in other villages. Muskan also relied on cost-effective business practices such as self-packaging with plastic food wraps, cheaper stickers for labeling, and smaller or customized molds. When local raw-material supply was scarce, the group first relied on contacts provided by the Mahasamund team and then arranged direct orders from the Uttarakhand supplier. The case-study notes further report that the women later started selling soap-making raw materials at reasonable prices to other training groups, effectively turning one of their original coordination bottlenecks into an additional business line.

Muskan therefore matters less as a feel-good anecdote than as a mechanism case. The group appears to have preserved exactly the organizational margins that the rest of the paper shows are hard to sustain in larger teams: repeated communication, agreement on meeting timing, transparent record-keeping, enforceable rules, and a clearly identified leader who made coordination legible. In that sense, Muskan is a positive example of how a large team can perform well when it addresses the two main frictions in the paper: coordination across members and the creation of a hierarchy-like organizing structure. Muskan does not overturn the average result that larger teams underperform. It shows instead that scale can work when a large team succeeds in building the bridging and organizational capacity needed to remain integrated.

D Data Analysis Appendix

This section documents how core outcome variables were constructed from meeting-level records and how group-level aggregates were generated. It reports supplementary regression tables, alternative outcome definitions, and robustness checks including covid-related specifications. It also includes additional evidence on team interaction patterns used to interpret the mechanisms in the main text.

D.1 Data Construction

Initial data related to group production was collected at the meeting level, meaning that groups were asked to fill in a log of every meeting that they had, even if they did not end up producing. Production variables Meeting Duration, Production Duration, and Quantity Produced were summed across all the meetings a group had, whereas attendance measures were taken as the average number of people that attended all meetings. Productivity measures were calculated by taking the ratio of these totals, and in the case of attendance, dividing by average attendance instead of original group size.

Table D.6: Variable Definitions

Variable	**Definition**
A. Production Variables	
Meeting duration	Group meeting duration (in minutes), total of all group meetings
Production duration	Group production duration (in minutes), total of all group meetings
Quantity produced	Soap quantity produced by group during production meeting, total of all group meetings
PC Productivity (per hour)	Hourly rate of soap production, divided by original group size
B. Organization Variables	
Ability	(=1) if group divides tasks by ability
Individual preferences	(=1) if group divides tasks according to individual-level preferences
Consensus	(=1) if group divides tasks by consensus
Everyone does all tasks	(=1) if all tasks are done by everyone
Work in subgroups	(=1) if groups split into subgroups to produce
Use Monitoring	(=1) if group has women monitoring others during production to ensure productivity and good soap quality
Monitors by group size	Number of monitors divided by group size
Has group leader	(=1) if group has a leader
C. Productivity Variables	
Hourly productivity	Total soap bars produced divided by production duration, in hours
Production PC (Size)	Total soap bars produced divided by original group size
Production PC (Attendance)	Total soap bars produced divided by average number of attending members
Productivity PC (Attended)	Total soap bars produced divided by average number of attending members, per hour of production duration
Share of high quality soaps	Share of high quality soaps produced out of the total
PC high quality production	Share of high quality soaps produced per capita
D. External Barriers Variables	
Input procurement issues	(=1) if group faced any problems related to procuring more raw materials

Variable	Definition
Marketing issues	(=1) if group faced any problems related to marketing of the soap product
Soap quality issues	(=1) if group faced any problems related to the quality of soaps
E. Individual-level Variables	
Owns savings account	(=1) if owns a personal savings account under their name in any financial institution
Account usage (frequency)	(=1) if savings account is used more than once a week, (=5) if less than once a month
Savings amount	Money saved in the past 30 days
Borrowed recently	(=1) if HH borrowed money from someone since baseline date
Amount borrowed	Amount HH borrowed since baseline date
Phone access	(=1) if had access to any mobile phone at home in the past 30 days
Autonomy KLK	KLK index of the degree of autonomy in household financial, health, education, and food-consumption decisions
Illness	(=0) if never feels ill, (=1) if has felt ill
Worry about money	(=0) if during last week never worried about providing basic needs, (=1) if during last week has worried
Kessler Scale	Kessler scale of wellbeing, based on feeling nervous, depressed, worthless, restless, and worn out. A higher score indicates worse mental health
Future expectations	In one year expects life will be better (=0), or more or less the same / worse (=1)

Note: Table D.6 reports definitions of all main outcomes and controls used in the paper, covering production, organization, productivity, external barriers, and individual-level welfare indicators.

Table D.7: Treatment Effects on Production Conditional on Producing

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Small	-56.204**	-14.494	-9.489***	1.294***
	(26.785)	(9.880)	(3.345)	(0.346)
Medium	-38.696	-2.843	-2.871	0.518
	(26.721)	(10.753)	(3.905)	(0.315)
Small = Medium	0.386	0.203	0.025	0.050
Mean (Large)	188.646	72.083	21.896	1.701
N	114	114	114	114

Note: Table D.7 reports estimates of specification (18). It reports production outcomes conditional on positive production. Columns use alternative production and productivity metrics; controls are selected through LASSO and robust standard errors (in parentheses) are clustered at the village level.

Table D.8: Meeting Frequency and Attendance

	Times Met (1)	Attendance (2)	Attendance (Normalized) (3)
Panel A: Full Sample			
Small	-1.742	-2.451***	-0.036
	(1.076)	(0.556)	(0.056)
Medium	-1.813	-0.748	0.055
	(1.196)	(0.630)	(0.064)
Small = Medium	0.887	0.000	0.119
Mean (Large)	3.213	3.809	.317
N	206	201	201
Panel B: Conditional on Meeting			
Small	-2.274*	-2.851***	-0.024
	(1.337)	(0.605)	(0.062)
Medium	-2.297	-0.751	0.086
	(1.456)	(0.681)	(0.070)
Small = Medium	0.969	0.000	0.085
Mean (Large)	4.081	4.632	.386
N	167	166	166

Note: Table D.8 reports estimates of specification (18). It reports meeting frequency and attendance outcomes. The table complements production outcomes by showing whether group size affects participation intensity; controls are selected through LASSO and robust standard errors are clustered at the village level.

Table D.9: Productivity Measures

	Productivity (per hour)	PC Production (Size)	PC Production (Attendance)	PC Productivity (per hour)	PC Productivity (by attendance, per hour)	Share of High Quality Soaps	PC High Quality Production
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Small	-5.963**	0.558	0.681	1.294***	2.208**	-0.067	3.529***
	(2.869)	(0.346)	(0.544)	(0.346)	(0.955)	(0.043)	(1.292)
Medium	-2.644	0.479	0.405	0.518	0.913	-0.081*	1.545
	(2.949)	(0.402)	(0.597)	(0.315)	(1.016)	(0.045)	(1.600)
Small = Medium	0.159	0.851	0.661	0.050	0.275	0.784	0.257
Mean (Large)	20.407	1.825	3.565	1.701	3.617	.928	6.618
N	114	114	114	114	114	114	114

Note: Table D.9 reports estimates of specification (18). It reports estimates using alternative productivity definitions, including per-hour, per-capita (by original size and by attendance), and high-quality-output measures. Controls are selected through LASSO. Robust standard errors in parentheses are clustered at the village level. ***, **, * indicate significance at the 1%, 5%, and 10% levels.

Table D.10: Monitored Production

	Quantity Produced	Production Duration	Soap Quality Rating	Number of Soap Quality	Visit Productivity	PC Visit Productivity (Attendance-Time)	Visit PC Productivity (Size-Time)	Visit PC Productivity (Attendance)	PC Visit Productivity (Size)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Small	-2.084***	-1.610	0.107	-0.350	-3.394***	1.021***	0.979***	0.480***	0.475***
	(0.488)	(1.581)	(0.274)	(0.564)	(0.997)	(0.206)	(0.135)	(0.064)	(0.109)
Medium	0.234	2.510	0.266	0.328	-0.409	0.678***	0.565***	0.358***	0.421**
	(0.769)	(1.893)	(0.270)	(0.559)	(1.195)	(0.256)	(0.140)	(0.092)	(0.172)
Small = Medium	0.005	0.035	0.545	0.102	0.020	0.231	0.023	0.260	0.780
Mean (Large)	7.721	33.7	2.256	2.977	14.069	1.8	1.172	.643	1.007
N	190	180	190	190	180	180	180	190	190

Note: Table D.10 reports estimates of specification (18). It reports estimates on production outcomes measured during monitored visits. Outcomes include quantity, duration, quality, and multiple productivity indicators based on attendance and group size. Controls are selected through LASSO, and robust standard errors in parentheses are clustered at the village level. ***, **, * indicate significance at the 1%, 5%, and 10% levels.

Table D.11: Productivity Measures (Alternative)

	Productivity (per hour)	PC Production (Size)	PC Production (Attendance)	PC Productivity (per hour)	PC Productivity (by attendance, per hour)	Share of High Quality Soaps	PC High Quality Production
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Small	-5.963**	0.558	0.681	1.294***	2.208**	-0.067	3.529***
	(2.869)	(0.346)	(0.544)	(0.346)	(0.955)	(0.043)	(1.292)
Medium	-2.644	0.479	0.405	0.518	0.913	-0.081*	1.545
	(2.949)	(0.402)	(0.597)	(0.315)	(1.016)	(0.045)	(1.600)
Small = Medium	0.159	0.851	0.661	0.050	0.275	0.784	0.257
Mean (Large)	20.407	1.825	3.565	1.701	3.617	.928	6.618
N	114	114	114	114	114	114	114

Note: Table D.11 reports estimates of specification (18). It reports estimates using alternative productivity metrics shown in Table D.9. Controls are selected through LASSO, robust standard errors are clustered at the village level, and ***, **, * indicate significance at the 1%, 5%, and 10% levels.

Table D.12: Treatment Effects on Mental Health, Expectations and Physical Health

	Kessler Scale	Worry About Money	Future Expectations	Illness
	(1)	(2)	(3)	(4)
Small	0.234	0.016	0.027*	0.011
	(0.260)	(0.027)	(0.016)	(0.027)
Medium	0.033	-0.023	-0.006	0.022
	(0.259)	(0.031)	(0.016)	(0.027)
P-values				
Small = Medium	0.402	0.238	0.075	0.688
Mean (Large)	3.33	.468	.055	.2
N	1443	1432	1443	1432

Note: Table D.12 reports estimates of specification (18). It reports estimates on mental and physical well-being outcomes, including illness, financial stress, Kessler-scale mental health, and future expectations. Controls are selected through LASSO and robust standard errors are clustered at the village level. ***, **, * indicate significance at the 1%, 5%, and 10% levels.

Additional robustness tables The tables in this subsection collect supporting robustness checks that are informative but not central enough for the main text. They report reduced-form evidence on external barriers to production, including packaging, marketing, and input-procurement constraints, as well as phone communication outcomes that speak to whether larger teams were simply harder to reach outside production. Taken together, they help rule out the view that the main size effects are driven by external frictions or survey-contact differences rather than by internal organization. We therefore keep them in the appendix while using them to discipline alternative interpretations of the network mechanism.

D.2 Controlling for Covid

In the previous regressions, we had a dummy variable controlling for whether a group was impacted by covid, as reported by that group. Here, we also present the tables as in the main text, but this time controlling for whether the group had to receive a replacement call because of the covid lockdown.

Table D.13: Treatment Effects on External Barriers

	Input Procurement Issues (1)	Marketing Issues (2)	Soap Quality Issues (3)
Small	-0.093	-0.084	-0.140
	(0.124)	(0.122)	(0.102)
Medium	-0.239**	-0.037	-0.051
	(0.120)	(0.121)	(0.109)
Small = Medium	0.191	0.653	0.326
Mean (Large)	0.438	0.406	0.25
N	144	146	146

Note: Table D.13 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. A smaller sample size is being used due to the question being introduced in the final stages of the survey.

D.3 Team Interaction Data

This subsection reports supplementary descriptive evidence from the roster-based interaction module used to construct the network measures in the main text. Table D.21 summarizes which members are most frequently nominated as organized, active, creative, hardworking, frequently contacted, and frequently learned from, providing a raw picture of how salient roles are distributed within teams. Table D.22 then reports the number of within-team contacts each respondent names, normalized by team size. These appendix tables are useful for interpreting the network data and for showing that the main connectivity patterns are not driven by systematic differences in survey response intensity across treatment arms.

D.4 Leadership Consensus and Additional Subgroup Evidence

The main text interprets centralization, betweenness, and monitoring as reduced-form evidence on whether teams generate broker roles or hierarchy-like structures as they grow. Ayah's recent leadership and nomination analysis adds two useful pieces of evidence that point in the same direction.

Table D.14: Treatment Effects on Phone Communication

	Autonomy (KLK) (1)	Phone Access (2)
Small	0.022	0.026**
	(0.057)	(0.011)
Medium	0.083	0.033***
	(0.053)	(0.011)
P-values		
Small = Medium	0.296	0.413
Mean (Large)	0.076	0.964
N	1092	1092

Note: Table D.14 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. Sample includes all women whose groups met up, but did not necessarily produce.

First, smaller and medium teams are more likely than large teams to agree on who the organizer is. Second, the raw network measures contain additional subgroup statistics that make the same fragmentation pattern visible beyond density and connected components. We view these appendix results as supportive mechanism evidence rather than as headline causal estimates.

Table D.23 summarizes the leadership-consensus block. Panel A reports the treatment effects on nomination convergence from the group-level leader-mismatch exercise. Relative to large teams, both small and medium teams display substantially higher convergence in organizer nominations. Panel B reports two individual-level nomination margins from the same family of exercises: in smaller teams, members receive more organizer nominations and, especially, more "talked-to-most" nominations than in large teams. This is consistent with the idea that bridge roles are more legible and more concentrated in smaller teams. Relatedly, the same analysis shows that a skill-network alignment index is positively associated with per-capita productivity in OLS specifications (0.0175 without size controls and 0.0158 with size controls), but the corresponding IV is weak (KP $F = 1.29$). We therefore treat the alignment-productivity relationship as suggestive rather than as a separate causal mechanism result.

Table D.24 reports additional subgroup statistics computed from the pooled network-measures dataset and collapsed to the group level. These moments are not necessary to establish the main density and components results, but they are useful because they show directly what fragmentation looks like. As team size rises, the average number of communities more than doubles, isolate shares increase, the share of the group captured by the top central nodes falls, the largest connected component covers a smaller fraction of the team, within-component path lengths rise, and modularity increases. Taken together, these statistics show that larger teams do not simply stretch a single connected structure. They split into more internally localized neighborhoods with weaker reach across them.

D.5 Baseline Capacity and Heterogeneous Fragmentation

The main text argues that larger teams underperform because their internal networks fragment as nominal size increases. An immediate question is whether this pattern is inevitable, or whether some teams are better equipped ex ante to sustain cross-subgroup links. To study that margin, we construct a team-level baseline capacity index from the pre-treatment individual survey and interact it with the randomized size assignment.

Construction of the index. The baseline survey covers 1,547 women in the 206 experimental teams. We use 21 pre-treatment items that plausibly capture organizational or managerial capacity: grade completed, business ownership, vocational training, phone access and ownership, WhatsApp group access and use, SHG participation, inclusiveness of SHG decisions, comfort speaking in SHG meetings, phone and in-person SHG communication, creativity, hardworking, proactive behavior, multitasking, leadership, free choice, and work, food, and financial autonomy. We orient each measure so that higher values correspond to greater organizational capacity, winsorize the numeric inputs at the 1st and 99th percentiles, and standardize them in the baseline sample. We then average the available standardized components to form an individual capacity score, requiring at least eight observed components. The team-level index is the mean of these individual scores within the team. In the estimation sample, we re-standardize this team index to mean zero and unit variance, so a one-unit change corresponds to a one-standard-deviation difference in baseline capacity. Its interquartile range runs from roughly -0.62 to 0.67.

Choice of team aggregator. A priori, the relevant organizational object could be the average skill mix of the team or the presence of one or two standout coordinators. We therefore compare three aggregators of member capacity: the team mean (z_{mean}), the top member (z_{top1}), and the average of the top two members (z_{top2}). The team mean provides the strongest predictive fit for bridge-like network outcomes, with an average incremental R^2 of 0.071 relative to size-only controls, compared with 0.046 for z_{top2} and 0.029 for z_{top1}. We therefore use the standardized team mean as the benchmark baseline capacity index.

Table D.25 reports the size-specific slope of this standardized index in pooled regressions that include size dummies, size-by-capacity interactions, and block fixed effects, with standard errors clustered at the village level. Because each village contributes only one assigned size arm, village fixed effects are not feasible. Panel A shows that baseline capacity strongly predicts whether teams remain integrated as they scale. Higher-capacity teams have fewer connected components, lower isolate shares, and greater bridge reach, with the largest coefficients appearing in the 12-member teams. Panel B reports the corresponding pairwise equality tests across size-specific slopes. The clearest formal amplification appears for connected components: the large-team slope differs from the small-team slope with $p = 0.008$ and from the medium-team slope with $p = 0.044$. For bridge index, largest connected component share, and isolate share, the large-team slopes are also economically steeper, although the pairwise differences are estimated less precisely. Evaluated at $z = +1$ and $z = -1$, the implied large-team difference is economically large: the predicted bridge index is 0.475 higher, the largest connected component share is 0.247 higher, the isolate share is 0.236 lower, and the number of connected components is lower by about 2.86. The direct productivity gradients are positive but imprecisely estimated, which is consistent with organizational capacity operating primarily through internal network structure rather than through an immediate additive productivity shifter.

Table D.15: Treatment Effects on Production Process

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Panel A: Full Sample				
Small	-34.056	-8.601	-5.286**	0.624**
	(20.635)	(7.650)	(2.542)	(0.290)
Medium	-14.727	2.120	-0.366	0.375
	(23.215)	(9.010)	(3.189)	(0.295)
Small = Medium	0.245	0.140	0.029	0.444
Mean (Large)	103.278	38.444	11.678	.907
N	204	204	204	204
Panel B: Conditional on Meeting				
Small	-44.448*	-11.552	-7.248**	0.673**
	(24.082)	(9.143)	(3.055)	(0.334)
Medium	-17.741	2.847	-0.943	0.395
	(25.803)	(9.914)	(3.720)	(0.368)
Small = Medium	0.141	0.071	0.014	0.439
Mean (Large)	132.786	49.429	15.014	1.166
N	165	165	165	165

Note: Table D.15 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in D.6. Meeting duration, production duration, and quantity produced are calculated by summing over all group meetings for each group. PC Productivity per hour is calculated by taking the ratios of total quantity produced to meeting duration (in hours), and dividing it by the original group size. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

Table D.16: Treatment Effects on Production Organization

	Ability	Individual Preferences	Consensus	Everyone Does all Tasks	Work in Subgroups
	(1)	(2)	(3)	(4)	(5)
Panel A: Full Sample					
Small	0.044	0.076**	0.088***	-0.287***	-0.061
	(0.037)	(0.035)	(0.033)	(0.093)	(0.049)
Medium	0.025	0.011	0.007	0.001	0.065
	(0.042)	(0.020)	(0.023)	(0.105)	(0.063)
Small = Medium	0.633	0.142	0.056	0.002	0.050
Mean (Large)	.022	0	0	.644	.07
N	204	204	204	204	190
Panel B: Conditional on Meeting					
Small	0.043	0.081*	0.113**	-0.386***	-0.089
	(0.049)	(0.043)	(0.045)	(0.101)	(0.059)
Medium	0.021	-0.008	0.013	-0.029	0.067
	(0.056)	(0.027)	(0.028)	(0.101)	(0.072)
Small = Medium	0.673	0.118	0.054	0.000	0.041
Mean (Large)	.029	0	0	.829	.086
N	165	165	165	165	161

Note: Table D.16 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Variables in columns (1)-(4) take on the value 1 if a group divides tasks according to ability, individual preferences, consensus, or by assigning to all tasks to everyone, respectively, and 0 otherwise. Column (5)'s "Work in Subgroups" takes the value 1 if the group divided in order to produce. Outcome variable definitions are available in D.6. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

Table D.17: Treatment Effects on Monitoring Production

	Use Monitoring	Monitors by Group Size	Has Group Leader
	(1)	(2)	(3)
Panel A: Full Sample			
Small	-0.047	0.025	0.020
	(0.082)	(0.018)	(0.052)
Medium	0.081	0.038*	0.006
	(0.084)	(0.019)	(0.058)
Small = Medium	0.096	0.629	0.756
Mean (Large)	.2	.019	.93
N	204	199	191
Panel B: Conditional on Meeting			
Small	-0.010	0.026	0.029
	(0.097)	(0.023)	(0.053)
Medium	0.121	0.044*	-0.007
	(0.096)	(0.023)	(0.065)
Small = Medium	0.135	0.561	0.348
Mean (Large)	.2	.024	.943
N	165	164	162

Note: Table D.17 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in D.6. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

Table D.18: Treatment Effects on Production Process- Technology Diffusion

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Small	-17.387	-0.825	-2.020	0.855**
	(25.256)	(10.251)	(2.874)	(0.392)
Medium	-48.492*	-7.921	-3.708	-0.131
	(27.060)	(12.341)	(3.137)	(0.303)
Small × Info Seeds	-13.543	-8.062	-2.039	-0.153
	(16.987)	(7.258)	(1.807)	(0.449)
Medium × Info Seeds	74.614***	23.289**	9.642***	1.136***
	(22.896)	(11.536)	(3.251)	(0.332)
Large × Info Seeds	19.148	7.188	4.767	0.307
	(36.809)	(13.568)	(4.752)	(0.363)
Small = Medium	0.099	0.477	0.474	0.013
Small × Info Seeds = Medium × Info Seeds	0.003	0.025	0.002	0.019
Small × Info Seeds = Large × Info Seeds	0.429	0.331	0.197	0.427
Medium × Info Seeds = Large × Info Seeds	0.203	0.371	0.391	0.089
N	203	203	203	203

Note: Table D.18 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in D.6. Sample size $N = 114$ corresponds to groups that recorded producing soap.

Table D.19: New Technology and Task Division among Different Size Teams

	Ability	Individual Preferences	Consensus	Everyone Does all Tasks	Work in Subgroups
	(1)	(2)	(3)	(4)	(5)
Small	0.093	0.075	0.131	-0.323***	-0.069
	(0.064)	(0.062)	(0.083)	(0.110)	(0.088)
Medium	0.028	-0.037	-0.026	-0.150	0.240
	(0.058)	(0.029)	(0.054)	(0.146)	(0.149)
Small × Info Seeds	0.053	-0.093	-0.031	-0.158	0.109*
	(0.091)	(0.066)	(0.084)	(0.140)	(0.055)
Medium × Info Seeds	0.040	0.046	0.010	0.098	-0.145
	(0.079)	(0.052)	(0.044)	(0.158)	(0.138)
Large × Info Seeds	0.125	-0.003	-0.069	-0.124	0.162
	(0.087)	(0.023)	(0.043)	(0.114)	(0.136)
Small = Medium	0.377	0.105	0.079	0.312	0.035
Small × Info Seeds = Medium × Info Seeds	0.913	0.135	0.661	0.250	0.113
Small × Info Seeds = Large × Info Seeds	0.561	0.209	0.689	0.850	0.713
Medium × Info Seeds = Large × Info Seeds	0.486	0.378	0.196	0.243	0.131
N	114	114	114	114	113

Note: Table D.19 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in D.6. Sample size $N = 114$ corresponds to groups that recorded producing soap.

Table D.20: Treatment Effects on External Barriers

	Input Procurement Issues (1)	Marketing Issues (2)	Soap Quality Issues (3)
Small	-0.119	-0.090	-0.143
	(0.123)	(0.121)	(0.104)
Medium	-0.259**	-0.043	-0.053
	(0.119)	(0.121)	(0.111)
Small = Medium	0.208	0.646	0.313
Mean (Large)	.438	.406	.25
N	144	146	146

Note: Table D.20 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. A smaller sample size is being used due to the question being introduced in the final stages of the survey.

Table D.21: Team Connections

Variable: Select the best at	Mean	Std	Min	Max	More than 3	All Team	N
Organizing Meetings	1.34199	1.14764	1	12	40	14	1155
Being active & moti-vated	2.08997	2.36618	1	12	156	102	1156
Creativity	1.63754	1.71147	1	12	88	53	1156
Being hard-working	2.41696	2.68151	1	12	231	123	1156
Communicating with (you)	1.82526	1.45902	1	12	79	19	1156
(You) Learning from	1.4282	1.54333	1	12	55	28	1156

Note: Table D.21 is descriptive and does not estimate a regression specification. It reports summary statistics from survey questions where group members identified who was most organized, most active, most creative, most hardworking, most frequently contacted, and most frequently learned from.

Table D.22: Number of (Normalized) Responses

	Talk on Phone	Meet in Person	Talk Production	Support in Production
	(1)	(2)	(3)	(4)
Small	-0.008	0.031	-0.023	-0.022
	(0.048)	(0.031)	(0.044)	(0.031)
Medium	0.055	0.043	-0.016	-0.001
	(0.044)	(0.034)	(0.046)	(0.034)
Small = Medium	0.165	0.678	0.852	0.448
Mean (Large)	.472	.736	.169	.11
N	1388	1388	1388	1388

Note: Table D.22 reports estimates of specification (18). ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Each column represents the number of group members each group member has interacted with (talked to on the phone, in person, talked during production, or has helped during production), divided by the group size. Outcome variable definitions are available in D.6. Sample size differs by individuals that received the question in the survey.

Table D.23: Leadership Consensus and Nominated Hubs by Team Size

	Small − Large	Medium − Large	Mean (Large)	N
Panel A. Group-level leadership consensus				
Nomination convergence (organizer)	0.171**	0.162**	0.125	125
Panel B. Individual nomination margins				
Organizer nominations received	0.060***	0.032**	0.054	1412
"Talked-to-most" nominations received	0.183***	0.071***	0.138	1412

Note: Panel A reports the coefficients for small and medium teams relative to large teams from Ayah's group-level nomination-convergence exercise. Panel B reports the corresponding small-minus-large and medium-minus-large gaps from the individual-level nomination-indegree exercise. Stars reproduce the significance markers in the underlying output files: ***, **, * denote significance at the 1%, 5%, and 10% levels. These appendix results are descriptive reduced-form evidence on how legible and concentrated leadership roles are across team sizes.

Table D.24: Additional Subgroup Metrics by Team Size

	Small	Medium	Large
Number of communities	2.381	3.865	6.077
Isolate share	0.199	0.229	0.299
Top-k centrality share	0.851	0.696	0.607
Largest connected component share	0.757	0.731	0.655
Average path length in LCC	1.348	1.609	1.837
Modularity	0.074	0.162	0.226
Number of groups	63	52	39

Note: This table reports group-level means from the pooled network-measures dataset, after collapsing the individual-level network outputs to the team level. The rows summarize subgroup structure rather than average density. Higher values of the number of communities, isolate share, average path length, and modularity all indicate greater internal separation; lower values of the top-k centrality share and the largest connected component share indicate weaker bridging reach across the team.

Table D.25: Baseline Capacity and Heterogeneous Fragmentation by Team Size

	Small ($n = 5$)	Medium ($n = 8$)	Large ($n = 12$)
Panel A. Size-specific slope of baseline capacity			
Bridge index	0.126*	0.113*	0.237**
	(0.065)	(0.059)	(0.108)
Largest connected component share	0.060**	0.060**	0.124***
	(0.025)	(0.026)	(0.038)
Isolate share	-0.044**	-0.061***	-0.118***
	(0.020)	(0.023)	(0.033)
Connected components	-0.254***	-0.471**	-1.430***
	(0.092)	(0.187)	(0.428)
Per-capita productivity	0.063	0.036	0.063
	(0.072)	(0.122)	(0.120)
Panel B. Equality of size-specific slopes (p-values)			
	Small = Medium	Small = Large	Medium = Large
Bridge index	0.880	0.377	0.340
Largest connected component share	0.996	0.165	0.181
Isolate share	0.584	0.058	0.154
Connected components	0.308	0.008	0.044
Per-capita productivity	0.847	1.000	0.873
Block fixed effects	Yes	Yes	Yes
Village-clustered standard errors	Yes	Yes	Yes

Note: Panel A reports the size-specific slope of a one-standard-deviation increase in the standardized team-level baseline capacity index from pooled regressions that include size indicators, the baseline capacity index, size-by-capacity interactions, and block fixed effects. Standard errors clustered at the village level are reported in parentheses. Panel B reports p-values for equality tests across those size-specific slopes. The baseline capacity index is constructed from 21 pre-treatment member characteristics, standardized at the individual level and then averaged across members within team. Network-outcome regressions use $N = 154$ teams; the productivity regression uses $N = 182$ teams. Because each village contributes only one assigned size arm, village fixed effects are not included. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

D.6 Why We Do Not Estimate a Separate Manager-Only Counterfactual

Earlier drafts considered a reduced-form coordination-manager technology in which appointing a manager absorbed some share of the unresolved coordination penalty directly. We do not use that exercise in the final quantification. Once the preferred heterogeneous model allows organizational capacity to change subgrouping, bridging, and hierarchy jointly, a separate manager-only counterfactual is no longer a primitive object but a special case of the broader organizational-capacity shift.

There is also an identification reason not to keep that older exercise as a headline quantitative result. In our data, leadership concentration, organizer consensus, monitoring, bridge reach, and fragmentation move together. A manager-only sensitivity grid can therefore fit by reallocating the same organizational gains across margins without being separately disciplined by the data. For that reason, the final quantification treats hierarchy as an organizational outcome generated by capacity and bridging rather than as an independent technology chosen in a separate stage.

E Wellbeing and Empowerment Impact on Team Members

In this section, we study whether team size had differential impacts on team members' wellbeing. The hypothesis is that if teams were more productive and organized their production process more efficiently, there might be also positive spillovers to individual-level outcomes like financial empowerment or mental health. To assess the extent of spillover effects from the group to individual members, we estimate the following regression equation at the individual level:

$$y_{ipv} = \beta_0 + \beta_1 T_{1ipv} + \beta_2 T_{2ipv} + X_{ipv} + \epsilon_{ipv} \tag{33}$$

where y_{ipv} represents individual-level outcomes, T_{1ipv} and T_{2ipv} are dummies for whether woman i in production team p in village v was assigned to either a large or medium group. We also include a vector of baseline individual-level characteristics X_{ipv}. While 1452 women were involved in the experiment, we restrict the individual-level analysis only to women who were part of active teams (1092 women) to make sure we capture the effect on team production.

Table D.14 shows that there is no differential impact on autonomy indicators (grouping autonomy in financial, health-related, education-related, and consumption decision-making) across members belonging to different groups. One potential reason for this lack of result is the relatively short time frame of the intervention. On the contrary, we observe that members of larger teams have less phone access than members in small and medium teams at endline.

We then turn to study treatment effects on physical and mental health. Results are reported in table E.26. Overall, we do not find any significant treatment effects on mental or physical health. Again, one potential explanation for the lack of findings is the relatively short time span of the intervention – improvements in outcomes like mental health or expectations about the future may take longer to realise.

Finally, we study the impact of belonging to different group sizes on borrowing and saving as reported in table E.27. We do not observe the differential impact on saving. However, we find at the endline that women belonging to large groups have borrowed 6,000 and 5,000 Indian Rupees less than women in small and medium groups, respectively. This finding is consistent with the hypothesis that members belonging to more productive groups wanted to finance the business and make it more sustainable.

Table E.26: Treatment Effects on Mental Health, Expectations and Physical Health

	Kessler Scale	Worry About Money	Future Expectations	Illness
	(1)	(2)	(3)	(4)
Small	0.283	-0.008	0.018	0.043
	(0.305)	(0.030)	(0.019)	(0.033)
Medium	0.199	-0.019	-0.013	0.040
	(0.297)	(0.032)	(0.017)	(0.031)
P-values				
Small = Medium	0.768	0.732	0.170	0.923
Mean (Large)	3.288	.48	.053	.193
N	1092	1087	1092	1087

Note: Table E.26 reports estimates of specification (18). It reports estimates on individual-level mental and physical health outcomes in the active-team sample. Controls are selected through LASSO, robust standard errors in parentheses are clustered at the village level, and ***, **, * indicate significance at the 1%, 5%, and 10% levels.

Table E.27: Treatment Effects on Saving and Borrowing

	Savings			Borrowing	
	Owns Savings Account	Account Usage (Frequency)	Savings Amount	Borrowed this Year	Amount Borrowed
	(1)	(2)	(3)	(4)	(5)
Small	0.001	-0.101	185.143	0.037	6000.368**
	(0.012)	(0.099)	(255.522)	(0.034)	(2527.242)
Medium	-0.008	-0.025	-165.286	0.038	5555.297**
	(0.015)	(0.107)	(152.752)	(0.037)	(2557.697)
P-values					
Small = Medium	0.524	0.390	0.188	0.973	0.878
Mean (Large)	.978	4.350	813.835	.325	8833.358
N	1092	1039	1092	1084	1056

Note: Table E.27 reports estimates of specification (18). It reports saving and borrowing indicators at the individual level, including in the sample only members that belong to active teams. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level.