Peer Effects in Social Networks: Evidence from an Entrepreneurship Experiment [†]

Vatsal Khandelwal [§]

Juni Singh [¶]

Abstract

We implement a randomized entrepreneurship program that induces peer interactions to study the mechanisms underlying peer effects – how they vary with network position, evolve over time, and can be engineered. We find that socially close and more connected peers generate short-term gains, while socially close but less connected peers generate long-term gains. This is driven by motivation from well-connected peers in the short term and ease of collaboration with less-connected peers in the longer term. We estimate peer effects and find that a 1σ increase in peers' entrepreneurial outcome increases an individual's future outcome by 0.4σ, but the effect declines by 0.06–0.08σ for each additional connection of the peer. Simulations show that pairing based on network position can be feasibly leveraged to increase treatment effects. Finally, we randomize a "connections" treatment in which peers share their network contacts. Despite network sparsity, sharing contacts does not generate additional gains and is concentrated within existing social groups, showing that peer effects are difficult to engineer in settings with high homophily.

JEL Codes: D85, O12, Z13, J16, L14, L26.

Keywords: Social networks; peer effects; social distance; entrepreneurship.

[†]We thank the Weiss Fund for funding the fieldwork for this project. We also thank our partner organizations, JPAL and Governance Lab, as well as our enumerators for implementing the project. We are very grateful to Marina Agranov, Abhijit Banerjee, Michael Callen, Margherita Comola, Horacio Larreguy, Sylvie Lambert, Karen Macours, Rohini Pande, Simon Quinn, Simone Schaner, Charles Sprenger, Alexander Teytelboym, Verena Wiedemann, and Christopher Woodruff for their helpful feedback. The field experiment and endline survey are registered as AEARCTR-0010081 on the AEA RCT Registry and has received IRB approval both at Caltech and the University of Oxford.

[§]University of Exeter. v.khandelwal@exeter.ac.uk

[¶]Caltech, World Bank. jsingh2@worldbank.org

1 Introduction

Peer effects are documented in a wide range of contexts (Calvó-Armengol et al. 2009, Duflo et al. 2011, Banerjee et al. 2013, Cai et al. 2015, Field et al. 2016, Beaman et al. 2021), but the mechanisms are less well understood (Bramoullé et al. 2020). First, it is not clear how peer identity can affect outcomes and whether these effects vary over time. For example, popular peers may be effective role models in the short term, while socially close peers may be better collaborators in the long term. Second, peers not only have direct effects but can also help indirectly by connecting individuals to their other contacts. The relative importance of this indirect, "bridging" channel, and the extent to which it can be facilitated to generate additional effects, remains largely unexplored.[1] Understanding these aspects is also crucial for assessing whether policy programs can be made more effective by recruiting specific peer types for specific roles and by leveraging peers as bridges to wider social networks.

In this paper, we induce peer interactions in social networks through an entrepreneurship training program to answer these questions. Our central contribution is to move beyond average peer effects and show how peer effects systematically vary by network position, change over time, and are not easy to engineer in settings with high homophily (i.e., preference for interacting with similar people). To show this, we first conducted a baseline survey and mapped the social networks[2] of 2840 women across 31 villages in Nepal and then randomized a subset to attend an intensive entrepreneurship training program. While individuals in pure control villages did not attend the program, individuals in treated villages were allocated into one of three groups. They either completed the program alone, attended with another woman assigned by the experimenter, or did not attend at all. Among those attending in pairs, we additionally randomized the implementation of a 'connections' treatment in which trainees were encouraged to share their social contacts with each other and think of ways these can help them set up a business.

We leverage the design to provide both reduced-form evidence and estimate peer effects. Conceptually, in terms of a linear-in-means framework (Manski 1993), our design tests both heterogeneity in how respondents weight different individuals in their network and whether peers can be leveraged to expand the set of individuals influencing a respondent's outcome. First, we exploit the variation in whether women attend the program alone versus with another woman in their village to assess whether social support can improve

[1] The "bonding" and "bridging" social capital were first introduced by Putnam (2000)

[2] Our main network measure captures all women in the village with whom an individual reports interacting with in any capacity.

outcomes.[3] This is especially promising as social networks in this setting are sparse[4] and exhibit a strong community structure with high homophily, suggesting that networking opportunities may be limited.[5] Second, we leverage whether individuals attended the connections treatment to compare the direct value of being trained with peers to the indirect value of accessing each other's social contacts. This allows us to assess whether increasing the perceived benefit of linking to new contacts and reducing the cost of reaching them can improve outcomes. Finally, we leverage the variation in the identity of the peer to study whether the treatment effect differs depending on their network position. In particular, we exploit variation in peer closeness (i.e., the number of network steps between two individuals) and centrality (i.e., the number of direct connections an individual has within the village network) across two survey waves to estimate heterogeneous peer effects[6]

We first measure short-term outcomes immediately after the training. We find that the training program significantly improves business knowledge and the intention to open a business. Being paired does not lead to additional gains relative to training alone on average. However, pairing is beneficial when the matched peer is socially close and more central than the respondent, i.e., has more connections in the village social network.[7] These pairs report higher business aspirations and greater take-up of additional resources than those trained with other pair types. These effects are similar across different measures of centrality and persist even after controlling for additional peer characteristics. Consistent with the effects of role models, we find that popular peers (i.e., in the highest quartile of centrality in their village) act as motivators. Those matched with popular peers have significantly higher aspirations after the training compared to those trained with less popular peers.

Sharing contacts via the connections treatment does not have a significant impact on short-term outcomes, even though women share 6.5 village-level contacts with each other on average, which is about two-thirds of the combined network size of both peers. Contact sharing is also significantly higher among those who are more steps away in the social network. At the same time, individuals are significantly less likely to share contacts with those of a different caste. Even among same-caste peers, pairs with higher homophily (in-group linking bias) in their caste share fewer contacts. Combined, this suggests that

[3]Peers can provide support through skill complementarities, motivation, and risk-sharing, especially since education, aspirations, and risk aversion are correlated with opening a business in our baseline data.

[4]These networks have an average density of 5.5% i.e., only 5.5% of total possible network links are present. This contrasts with the average density of 11.9% in villages in Karnataka in Banerjee et al. (2024).

[5]Links with individuals of the same caste are also 21.5 percentage points higher, on average, than the proportion of same-caste members in the village population, indicating high levels of homophily.

[6]Heterogeneity along these dimensions was pre-registered in our analysis plan. Details on how we depart from our pre-analysis plan are provided in Section 3.

[7]We define a person as socially close when their network distance is less than or equal to 2, i.e., strictly lower than the median social distance equal to 3.

the indirect benefits of networking may not be easy to leverage in homophilous networks.

Following this, we implemented a follow-up survey with a sub-sample one year after the intervention. We find that those trained alone are significantly more likely to take steps to open a business, but those who were matched with a pair have no average effect. Decomposing this by peer position as before, we find that those matched with a socially close peer have a significantly higher treatment effect on average outcomes compared to those matched with a socially distant peer. However, unlike the short-term results, those matched with more central peers are significantly less likely to have taken steps to open a business compared to those matched with other pair types. This suggests that peer centrality has different short-term and long-term effects, while social closeness has similar effects over time. We complement these findings with a random forest analysis (Breiman 2001), which shows that peer network characteristics are consistently ranked among the top three predictors of outcomes, even when compared to peer demographics.

The benefit from interacting with specific types of peers can arise due to multiple mechanisms, such as learning, encouragement, risk-sharing, or the probability of future interactions. We disentangle these mechanisms and find that the effects are primarily driven by motivation from a more central, socially close peer in the short term and the ability to interact with a socially close peer of a similar social status in the long term. This is due to the following reasons. First, the most commonly reported reason trainees thought pairing was beneficial was encouragement from their peer, followed by learning better together. Those matched with a more central peer were significantly more likely to report feeling encouraged. Second, socially close pairs performed better on in-training exercises and were significantly more likely to report wanting to meet in the future and start a business together than socially distant peers, suggesting stronger collaboration during and immediately after training. Third, the reversal of the effects of network position in the long term could be linked to long-term interactions. Trainees paired with a peer were more likely to have discussed business ideas within their village, and 31% reported contacting their matched peer from the training for advice, borrowing or lending money, or discussing business ideas. However, those matched with less central peers spoke less with others about businesses, and pairs across caste boundaries were also significantly less likely to have remained in contact, highlighting the importance of social status. This suggests that short-term motivational gains from central peers did not translate into sustained interactions, which instead occurred mainly among individuals of similar social status.

Next, we use the field data to estimate peer effects. We first present a simple model in which each agent chooses entrepreneurial effort as a function of her peer's effort. Peer effort enters through two channels: motivational benefits, where working with a central peer incentivizes effort, and collaboration frictions, where working with a central or socially distant peer makes effort more costly. The model generates heterogeneous peer effects that vary with social closeness and peer centrality over time. Taking the model to the data, we first find correlational evidence that peer centrality generates positive peer effects in the short term but not in the longer term. We identify peer effects by exploiting random pairing to eliminate endogenous sorting into pairs and by using data from two survey waves to mitigate the "reflection problem" (Manski 1993). We find that a 1 standard deviation increase in a peer's outcome raises the individual's outcome by 0.4 standard deviations, but the effect declines by 0.06–0.08 standard deviations for each additional increase in the peer's number of network connections. This heterogeneity persists even after controlling for individual and peer characteristics. Linking back to the model, this suggests that the impact of collaboration costs is higher in the longer term, as sustained interactions with central peers become more difficult without corresponding improvements in motivation.

Importantly, while pairings based on social distance can be easy to implement, the feasibility of pairing based on centrality remains uncertain, since network data can be hard to collect and there may be a limit to how many centrality-based matches we can construct within a fixed network. We find that the majority of our sample guesses their peers' degree centrality correctly within one standard deviation, implying that additional data collection may not be necessary. Moreover, as a policy implementation exercise, we simulate 10,000 counterfactual reassignments of peer pairings within each village in our dataset and plot the distributions of the average degree centrality gap by village. The resulting distributions show wide dispersion, indicating that strategic pairings based on centrality are indeed feasible and can generate high returns.

Our paper contributes to the existing literature in three ways. *First*, we contribute to the large literature on peer effects in economics by showing that their magnitude and sign vary systematically with network position and change over time. We show that this heterogeneity is both policy relevant and useful for understanding the mechanisms through which peers influence outcomes. While peer effects have been measured in various contexts, from the adoption of new technology (e.g. Beaman et al. (2021)), financial products (e.g. Banerjee et al. (2013), Cai et al. (2015)), education (e.g. Calvó-Armengol et al. (2009), Duflo et al. (2011)), and entrepreneurship (e.g. Lerner & Malmendier (2013), Field et al. (2016)), the majority of literature that estimates peer effects employs a linear-in-means model (Manski 1993), which largely treats peers as homogeneous and only captures average effects.[8] Recent literature leverages random variation in the peer's network position to show how it can heterogeneously affect different types of behaviour,

[8] The limitation of the linear-in-means assumption has also been noted in the context of peer effects in education (Sacerdote 2011) and more generally in Boucher et al. (2024), who show that this assumption can be misleading.

ranging from contract enforcement (Chandrasekhar et al. 2018), to savings (Breza & Chandrasekhar 2019), and the development of social skills (Zárate 2023). Contributing to this growing literature, we show that heterogeneity in peer effects indeed depends on network position and varies over time depending on the mechanism underlying the peer effect. If the main aim is to provide short-term motivation, then networking with socially close and central peers might be helpful. However, if the main aim is to encourage long-term collaborations, then networking with less central peers might be sensible. We also show that leveraging this heterogeneity to develop an adaptable networking strategy is feasible and can generate high returns.

Second, our experiment design allows us to distinguish between two contrasting mechanisms through which social capital can be beneficial—mechanisms that, to the best of our knowledge, have not been disentangled in the existing literature. Peers can directly influence outcomes, or they can serve as an indirect gateway to a broader social network, aiming to generate additional peer effects. We disentangle these direct and indirect channels using a novel 'connections' treatment and highlight the extent to which peer effects can be engineered. By testing this, our design contributes to the large literature on social capital (eg: Granovetter (1973), Glaeser et al. (2002), Durlauf & Fafchamps (2005)). In particular, we contribute to the literature on "bonding" and "bridging" social capital introduced by Putnam (2000) and to understand how responsive they are to social interventions. We find that encouraging network formation by facilitating 'bridging' may not be feasible in contexts with high homophily, as individuals primarily share contacts with those who have a similar social background, and greater homophily reduces the gains from sharing contacts within one's own group. Consequently, the indirect channel does not produce meaningful effects in our setting.

Finally, we contribute to the literature on entrepreneurship and the role of peers in improving business outcomes (Field et al. 2010, De Mel et al. 2014, Field et al. 2016, Cai & Szeidl 2018, Vasilaky & Leonard 2018, Fafchamps & Quinn 2018, Brooks et al. 2018, Carranza et al. 2018, Campos et al. 2019, Asiedu et al. 2023, Vega-Redondo et al. 2023). McKenzie et al. (2021) provide a comprehensive review of the literature on training programs. Unlike most of the existing literature, we focus on a sample of non-business owners and examine both the immediate and one-year impacts of business training on this group. We show that peer effects in entrepreneurship mask significant network-based heterogeneity in that there are optimal peer matches that can be leveraged to improve outcomes. By doing so, we contribute to existing literature that documents the average treatment effect of peers in improving business or agricultural outcomes (Field et al. 2016, Cai & Szeidl 2018, Vasilaky & Leonard 2018). Moreover, we show that peer effects can operate via different mechanisms over time and that peer matches must be designed accordingly to achieve the desired objective. Related to our paper, Vega-Redondo et al. (2023) also documents the role of peer heterogeneity by showing the effects of diverse and non-diverse peer groups on entrepreneurship and how this varies based on interaction format (virtual or in-person). In contrast, we examine the impact of a peer's network identity within the same interaction format, focusing on how individuals in the same social network and cultural context can support each other over time.

The paper is organized as follows. Section 2 describes the context and experiment design. We discuss the reduced form and peer effects estimation strategy in Section 3. We present the short-term and long-term results of the intervention in Section 4 and discuss the mechanisms in Section 5. This is followed by Section 6 where we provide a theoretical framework, estimate peer effects, and discuss the feasibility of strategic pairing. Section 7 concludes.

2 Data and Experiment

We first conducted a detailed baseline survey with 2840 women across 31 villages in rural Nepal in September-October 2021. We collected data on demographic characteristics (age, marital status, caste, education), willingness to start a business, reasons for not doing so, risk aversion, and aspirations related to agricultural investments, non-agricultural investments, and income.[9] In addition, we administered a network survey to elicit information about social networks, as in Banerjee et al. (2013). These questions measure with whom individuals report interacting in any capacity, including visiting homes, giving or taking advice, going to a temple, or contacting during a health emergency. Individuals were asked to list as many names as they preferred. For the majority of the analysis, we treat links as undirected, i.e., individual i is assumed to be a friend of j if either of them mentions each other's names for any of the above interactions. We successfully elicited social networks for about 80% of village populations on average.[10]

2.1 Baseline

Baseline summary statistics are presented in Table A.1. The average age of women in our sample is 38, and 92% of them are married. Around 45% of our sample have no formal education. Roughly 22% of women report having opened businesses already, and 42% report a willingness to open a non-agricultural business. We find that 89% aspire to earn an income higher than their current income, while 23% aspire to spend more on non-agricultural business expenditures than their current investment. We also elicited

[9] Data on aspirations were collected according to the procedure outlined in Bernard & Seyoum Taffesse (2014).

[10] Our main results are robust to correcting the centrality measure to account for our sampling strategy. These results are discussed in Section F.

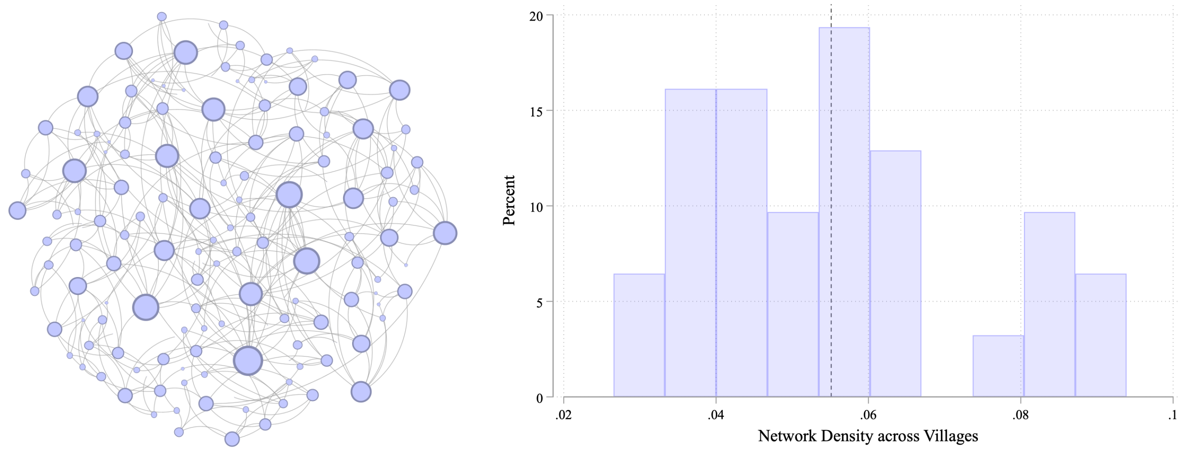


Figure 1: The figure on the right shows the social network in a village from our data. Larger nodes represent women with more number of connections. The figure on the left plots the network density for each village in our sample. Density is computed as the number of realized edges in the social network as a proportion of total possible edges. The number of total possible edges is given by $\frac{n(n-1)}{2}$ where n is the number of individuals in the network.

risk preferences using a standard choice experiment involving a series of lotteries and a fixed payment. On average, women are risk-averse with a risk-aversion level of around 4.6, where 6 stands for very risk-averse and 1 stands for risk-loving.

2.1.1 Network Statistics

Degree centrality is defined as the total number of undirected links an individual has in the village social network, i.e., the count of people who either named or were named by the individual as a friend. The average degree centrality in our sample is about 4.6. The left panel of Figure 1 plots the social network of one of the villages in the sample, showing the heterogeneity of the number of connections reported. Table A.2 shows that the number of connections is correlated with other demographic variables of interest, including education, caste, and marital status, with significantly lower links among unmarried women, women with no education, or women of a lower caste.

Networks in these settings have two key features. First, networks are sparse i.e. each person is connected to a small share of others in their village and most potential links are absent. We plot the density of networks, i.e., the number of observed links divided by the total possible links, computed for each village (Figure 1). The average network density across all villages is 5.5%, i.e., only 5.5% of links exist out of the total possible number of links given the number of individuals observed in each network. This suggests that the networks are very sparse.[11] Second, we find that these networks exhibit a community structure and a high level of homophily: intra-caste links are more common than links across castes, even after accounting for the number of same-caste members present in the village. On average, the proportion of same-caste links exceeds the proportion of same-caste village members by 21.5 percentage points. These findings suggest that pairing individuals and having them interact may involve social interactions that may not naturally occur in our setting.

2.1.2 What are the barriers that prevent women from opening businesses?

We find that women who already own a business have a 50% higher median income than those with no business ownership. Correlating whether or not they have opened businesses with their baseline characteristics in Table A.3, we find that those who own a business are younger, more educated, less risk-averse, and have higher aspirations for investment in non-agricultural business. This is also shown in Figure A.1 that employs a Lasso regression (at various penalty levels) to show which variables correlate with whether or not a business has been opened already out of a list of demographic and network characteristics. We find that being educated, and especially having obtained a university education, matters the most in explaining the decision to have already opened a business.

When those who haven't opened up businesses were asked to give reasons why they haven't done so already. 77% say that they do not have the skills or do not feel capable, while 24% say that they lack the financial ability. We correlate their willingness to open businesses with baseline demographics, number of network connections, aspirations, and other variables such as risk aversion. These results are shown in Table A.4. We find that those who are older and more risk-averse are correlated with being less willing to open businesses. This is also reinforced by Figure A.2 where we implement a Lasso regression, which selects age, risk-aversion, and education as the relevant correlates at high values of the penalty parameter. This suggests that risk-sharing and skill complementarities with peers could be helpful in setting up a business. Finally, those who have higher aspirations are more willing to open businesses, suggesting that peers can potentially be used to motivate and boost aspirations.

2.2 Experiment Design

We conducted the experiment in September 2022. The experiment consisted of a three-day entrepreneurship training program motivated by the SIYB module developed by the International Labour Organisation. The training typically lasted 3 hours per day, and

[11] The sparsity persists even after we correct each individual's degree centrality (i.e., number of links) to account for our sampling strategy. Average density is 7% in that case.

individuals were given NPR 100/day (i.e. 0.7$) for participation. The training focused on various topics, including defining entrepreneurship, listing the characteristics of a successful entrepreneur, learning how to build a business plan, learning about market scoping and market access, setting savings goals, and boosting aspirations using a video highlighting a successful female entrepreneur. Figure A.4 represents our two-step randomisation design.

First, we randomly allocated villages to Pure Control and Treatment. Following this, women in treatment villages were randomly allocated into one of the four groups as shown in Figure A.4 and discussed below.

Control: Women in the control group did not attend the training. They are treated as the spillover group in the one-year follow-up.

Treatment 1: Trained Alone: Women in this group attended the training alone, i.e., the enumerator seated the participant alone without pairing them with another individual. Other women from the same village were present in the classroom, but there were no coordinated activities for these women throughout the training. These individuals completed all the interaction-based activities of the training alone, e.g., developing a business plan or playing an investment and savings game.

Treatment 2: Trained with a Peer: Women in this group attended the training paired with another person from their social network. Women were allocated a peer by the enumerator at the training centre, ensuring that this peer is not self-selected. These pairs completed all the interaction-based activities of the training with each other, e.g., developing a business plan or playing an investment and savings game.

Treatment 3: Trained with a Peer + Connections Module: As in Treatment 2, women in this group also attended the training with another person from their social network. However, in addition, the pair received a "connections module" that encouraged them to share their network contacts with each other and think about how network ties can be helpful in opening a business. To guide participants, we first highlighted three main ways in which networks can be helpful: i) Information-Sharing ii) Complementarity in skills, and iii) Risk Sharing, and then asked them to think about how they can help each other. Following this, the pair were asked to write down and share their network contacts that could potentially help them in opening up a business together, and think of ways each of these contacts could be helpful.

Conceptual Motivation for the Connections Module:

Conceptually, the connections module aims to test whether increasing the perceived benefits of linking and reducing the cost of reaching new contacts can induce useful interactions.

Suppose individual i decides whether to interact with individual k with expected utility $E[U_{ik}] = \mu_{ik} - c_{ik}$, where μ_{ik} denotes the perceived benefit from interacting and c_{ik} is the cost of approaching k, including learning about them in the first place. In sparse, homophilous networks, μ_{ik} may be low and c_{ik} may be high. The module is designed to reveal who can be reached through the matched peer j and how these contacts might be useful, increasing μ_{ik} and lowering c_{ik} at the same time.

The connections module also allows us to compare between the direct and indirect benefits of peers, comparing the direct role played by peer motivation and collaboration, for example, with the indirect role played by peers in facilitating additional useful interactions with others in the network. Figure A.5 highlights the difference between these two channels.

2.3 Sampling and Implementation

Out of 2,840 women surveyed at baseline, we excluded those already operating businesses and four villages where training could not be implemented ex-ante. We then randomized the remaining 1,970 women across 27 villages into four treatment arms. Nearly half of this sample was included in the experiment and successfully surveyed at endline, and roughly two-thirds of those were followed up one year later. Figure B.1 summarizes participant numbers at the various stages of data collection.[12] Our main comparisons of interest, between those trained alone versus with peers and across peer types within the paired arms, are unaffected by attrition and compliance. However, we discuss these below to provide context for interpreting the average treatment effects of the experiment.

Dropouts and Attrition: The endline survey was conducted just before the training for the control group and immediately after the third day of training for those who attended training. We define dropout as a binary variable indicating whether the individual was present in the endline survey. Dropout rates were higher in treated villages than in pure control villages. This difference is driven by survey logistics rather than participant selection as women in treated villages completed the survey at the training centre, whereas those in pure control villages were surveyed at home. Consistent with this, within treated villages, dropout rates do not differ between those assigned to treatment and those in the within-village control group (Table B.3), suggesting that being offered training did not selectively affect participation. Dropout rates are also similar for those trained alone and those trained with a peer, which is our main comparison of interest. Baseline characteristics of dropouts and non-dropouts are balanced along age, income, caste-status,

[12] An additional 260 non-randomized women, including those from two excluded villages, also completed the endline. They are excluded from the reduced form analysis as we implement intent-to-treat analysis but their partners (who were present in the randomized list) are retained. Appendix B.3 provides additional details.

and, importantly, with baseline willingness to open a business. Differences are limited to education and the number of connections in the social network.(Table B.4)

We also conducted a phone-based follow-up survey one year later with a subset of 580 women from the endline sample. Table B.5 shows that attrition in this survey is not correlated with treatment status. Non-response was primarily due to changes in phone numbers.

Compliance: Next, we define compliance as a binary indicator equal to one if the individual's realized treatment status matched their assignment. Overall compliance in the endline sample is 68%. As shown in Table **??**, baseline characteristics of compliers and non-compliers are similar across almost all baseline characteristics. Table B.1 shows that compliance is uncorrelated with treatment status in the endline sample. However, if we exclude pure control villages, compliance is positively correlated with treatment status. As a result, we report intention-to-treat (ITT) estimates throughout the analysis. We also present robustness checks using within-village comparisons and IV specifications that instrument actual treatment status with treatment assignment. Further details on attrition and compliance are in Appendix Section B.3.

Balance Checks: To further show that dropouts and compliance issues do not affect internal validity, we check for balance in the baseline characteristics of individuals across the pure control and treatment groups who appear in our endline and follow-up samples, respectively. We check for balance on several baseline variables, including demographic, network, and business characteristics such as income, sources of income, age, education, caste, network connections (i.e., degree centrality), and aspirations, for all individuals across all treatment arms in the endline and follow-up sample. Table B.6 and B.7 show these results. We find that the sample is balanced on most characteristics across all pairwise comparisons between treatment arms. Out of 114 pairwise comparisons (6 between-arm comparisons across 19 variables in the endline sample), only 1 is significant at the 5% level, and only 8 in total are significant at the 10% level or below. We find similar results for the follow-up sample, where the majority of comparisons are not significant either. This suggests that the pure control group and treated groups are largely similar along baseline characteristics.

2.3.1 Peer Assignment and Randomization Inference

Finally, we assess whether peer assignment is as-good-as-random. Pre-assignment of pairs was not feasible due to issues around compliance and attendance. Instead, peers were assigned on the spot by the research implementation team, with the goal of avoiding self-selection by ensuring that each woman was 'randomly' paired with another attendee from the same village. Since this approach can only approximate random assignment, we assess whether the implemented assignments are as-good-as-random by conducting randomization inference on the entire sample of individuals who were actually paired for the training. Specifically, we simulate multiple alternative peer assignments within each village and compute, for each assignment, a statistic capturing the total difference in characteristics between assigned pairs. We find that the observed value of the statistic from the actual pairing lies well within the distribution generated from these simulations, suggesting that the implemented peer assignments are statistically indistinguishable from random. These results are discussed in Section B.5.

3 Estimation Strategy

3.1 Endline Outcomes

We measure the effect of the experiment on five main endline outcomes collected after the training. These include the Knowledge Index, Business Aspirations Index, Business Index, Take-up Index, and Additional Steps consistent with our pre-analysis plan.

The indices are constructed as follows: (1) **Knowledge Index**: We ask five short questions that measure knowledge gained during the course of the training, and take the proportion of correct responses to construct the knowledge index. We collect this measure for both treatment and control groups.[13] (2) **Business Aspirations Index**: We compute a measure of business-related aspirations comprising yearly non-agricultural investment, monthly income, and savings, elicited following the procedure in Bernard & Seyoum Taffesse (2014). We ask individuals the minimum and maximum of the relevant variables in their neighbourhood, the current value, and what they aspire to in the corresponding time frame, and construct the weighted average of the individual's aspirations across these dimensions to create the index. (3) **Business Index**: This is constructed as a weighted average of responses to whether the individual is ready to invest in a business, whether they would submit a plan for a business competition, and how likely they are to start a business, capturing forward-looking intention and self-assessed likelihood of taking entrepreneurial action. (4) **Take-up Index**: This is a weighted average of hypothetical willingness to seek additional support. Specifically, we ask if individuals would attend additional paid training sessions or mentoring workshops in the next year, how much they would be willing to pay for each, and whether they are open to receiving advice from community members. (5) **Additional Steps**: This index captures concrete steps toward setting up a business, including whether the individual intends to open a savings account

[13] The questions are: (a) What do you understand by a business? (b) What characteristics are required to be a successful entrepreneur? (c) What do you mean by fixed assets? (d) What sector does a beauty parlour come under? (e) Above what break-even percentage does the business become risky?

or take a loan for the business. In each case, we construct a weighted average using the weighting procedure proposed in Anderson (2008) and normalize indices using the pure control group. [14]

Because each index aggregates several underlying variables, movements in these indices capture broad changes in behavior and are substantively meaningful.

3.2 Follow-up Outcomes

Treatment effects on follow-up outcomes are measured one year later with a subset of the endline sample using phone surveys. These outcomes include whether or not the individual has opened a business, their monthly income, agricultural investments, agricultural profits, whether they have opened a new savings account, the amount of money they save, and whether they have taken a loan. In addition to these main economic outcomes, we also measure other outcomes, including income aspirations, whether individuals sign up for a potential commitment savings account from which funds cannot be withdrawn except for business purposes, record-keeping for agriculture, and other outcomes regarding community interactions around advice-taking and collaborations.

We construct four indices to capture follow-up outcomes.[15] First, the **Outcomes Index** includes monthly income and agricultural profits. The **Steps Index** captures concrete actions taken toward setting up a business, including whether the individual reported opening a business, opening a new savings account, or joining a cooperative. We also include in this the level of business investments, agricultural investments, savings, and loans. Third, the **Mindset Index** captures forward-looking motivation and includes income aspirations, willingness to start a business, perceived self-efficacy in running a business, and willingness to open a (hypothetical) commitment savings account (both participation decision and amount committed), where the amount can only be withdrawn for business-related expenses. Finally, the **Business Practices Index** includes whether individuals keep records for agriculture or for business, capturing improvements in organization. As before, all indices are normalized relative to the pure control group.[16]

[14] Variables entering the aspirations index and the variables measuring payments for additional training/mentoring entering the take-up index were winsorised at the 99th percentile.

[15] We rely on an intuitive categorization since the follow-up survey was not pre-registered.

[16] All monetary variables were winsorized at the 1st and 99th percentiles to exclude outliers. This includes monthly income, total savings, agricultural profit, agricultural investment, and income aspirations. In contrast, business investment and loan amount were winsorized at the 99th percentile only, since the majority of respondents reported zero values for these variables.

3.3 Empirical Specification

3.3.1 Intent-to-treat Effects

We measure the impact of the treatment using the main specification described below, where we compare individuals in pure control villages with those in treated villages:

$$\begin{aligned}Y_i = \alpha &+ \beta_1 \text{ Trained Alone}_i + \beta_2 \text{ Trained With Peer}_i \\ &+ \beta_3 \left(\text{Trained With Peer+Connections Module}\right)_i + \epsilon_{iv}\end{aligned}$$

Y_i is an outcome measure for individual i, *Trained Alone* is a binary variable that takes the value 1 if the individual was intended to be treated alone and 0 otherwise. Similarly, *Trained With Peer* is a binary variable that takes value 1 if the individual was intended to be treated with a pair and *Trained With Peer+Connections Module* is a binary variable that takes value 1 if the individual was intended to be treated with a pair and attend the additional connections module. For the follow-up, we additionally include and control for an indicator variable for the spillover group (control individuals in treated villages). Standard errors are clustered at the village level.

As robustness exercises, we also estimate specifications that restrict comparisons to within-village controls and also instrument actual treatment status with assignment to obtain local average treatment effects. These results are reported in the Appendix Section C.1.

3.3.2 Impact of Training with different types of peers

Following this, we consider differences in outcomes for different pairs in treatments 2 and 3 compared to treatment 1. Our intention to leverage the random variation in the network identity of the peer and use it to study heterogeneity by social distance and network centrality was pre-registered in our pre-analysis plan. Let d_{ij} be the network distance between i and j, i.e., the number of links needed by i to reach j, and let ϕ_i be the network centrality (e.g., number of connections) of agent i.

We implement specifications using both continuous and discrete measurements of network statistics, as well as absolute and relative measurements of peer centrality, i.e., comparisons with the matched peer. First, we combine the paired treatment arms and interact the indicator for each with continuous measures of peer social distance and peer centrality, and implement the following specification.

$$\begin{aligned}Y_i = \alpha &+ \beta_1 \text{ Trained Alone}_i + \beta_2 \text{ Trained With Peer}_i \\ &+ \beta_3 \left(\text{Trained With Peer}_i \times d_{ij}\right) + \beta_4 \left(\text{Trained With Peer}_i \times \phi_j\right) + \epsilon_i\end{aligned}$$

Following this, we split individuals into whether or not they are socially close. We define a pair ij as socially close if the network distance between i and j is strictly less than the median social distance in the sample $d_{ij} = 3$. Similarly, we classify peers into two additional categories: if they have the same or fewer network connections ("Less Central") or more network connections ("More Central") than their matched peer. We then implement the following specification, where we regress outcomes on the indicator for being trained alone, trained with a peer, trained with a close peer, and trained with a more central peer.

$$\begin{aligned} Y_i = \alpha &+ \beta_1 \, \text{Trained Alone}_i + \beta_2 \, \text{Trained with Peer}_i \\ &+ \beta_3 \, \text{Trained with Close Peer}_i + \beta_4 \, \text{Trained with a More Central Peer}_i + \epsilon_i \end{aligned}$$

Finally, we classify peer type into one of four interaction categories: Close × Central, Close × Noncentral, Far × Central, and Far × Noncentral. For any pair ij, i is assigned to the category close-central if their matched close peer j has strictly higher degree centrality compared to i, i.e., $\phi_i - \phi_j < 0$ and i and j have social distance $d_{ij} \leq 2$. The other categories are included in a binary variable called Other Pair Type.

$$\begin{aligned} Y_i = \alpha + \beta_1 \, \text{Trained Alone}_i + \beta_2 \, (\text{Close} \times \text{Central})_i \\ + \beta_3 \, \text{Other Pair Types}_i + \beta_4 \, \phi_i + \epsilon_i \end{aligned}$$

Additional Controls: Note that an individual who is central in the network is mechanically more likely to be paired with a less central peer, even when pairing is random. This implies that an individual's centrality can be correlated with the group to which they are allocated in the above specifications. To ensure that this does not affect our results, we control for the individual's degree centrality in all regressions that use a relative measure of network centrality (eg: regressions using an indicator for whether the peer is "More Central" than them). In addition, in all regressions that compare peer types or analyze heterogeneity within the paired arms, we control for individuals who were intended to be paired but were not actually matched due to non-compliance. Finally, we also present specifications including controls for observable peer characteristics that may be correlated with peer centrality and interpret network centrality as a sufficient statistic capturing innate, unobserved traits such as charisma or role-model effects.

Alternative Measures of Centrality: We also show robustness of our main results to alternative undirected measures of centrality, including eigenvector centrality (importance from being connected to other well-connected nodes) and betweenness centrality (importance from lying on the shortest paths between others), and directed measures of centrality such as indegree (importance from receiving many links). We also check for robustness to an alternative measure of degree centrality that accounts for our sampling strategy and any missing links.

3.3.3 Peer Effects

In addition to the reduced form analysis, we identify peer effects by exploiting random pairing and the panel structure of our data, estimating a linear-in-means specification (Manski 1993) with heterogeneity based on network position. We use the entire sample of individuals who were paired during the training. As a first step, we examine whether outcomes of paired individuals are correlated at the endline and follow-up.

Using data from Treatments 2 and 3, we first estimate:

$$Y_{it} = \theta + \theta_o \mathbf{X}_i + \theta_p \mathbf{X}_j + \theta_e Y_{jt} + \theta_d (d_{ij} \times Y_{jt}) + \theta_c (\phi_j \times Y_{jt}) + \epsilon_i$$

where Y_{it} and Y_{jt} are outcome indices for individual i and peer j in wave t, d_{ij} is the social distance between them, and ϕ_j is peer j's centrality. $\mathbf{X}_i$ and $\mathbf{X}_j$ denote controls measuring own and peer characteristics. These correlations test whether the magnitude of association between own and peer outcomes varies systematically with closeness or centrality. However, because outcomes are measured contemporaneously, this exercise does not resolve the reflection problem highlighted in Manski (1993) despite pairs being formed randomly.

To address this, we regress each individual's follow-up outcome on their peer's endline outcome. The panel structure is critical because it allows us to use lagged peer outcomes to predict later own outcomes, thereby addressing the simultaneity concern that drives the reflection problem. We additionally control for the individual's own endline outcome as well as both own and peer characteristics. Importantly, we interact peer endline outcomes with measures of social closeness and peer centrality to test whether peer effects vary with network position:

$$Y_{i1} = \alpha + \beta_1 Y_{j0} + \rho \left(Y_{j0} \times \text{Close}_{ij}\right) + \eta \left(Y_{j0} \times \phi_j\right) + \beta_2 Y_{i0} + \gamma \mathbf{X}_i + \delta \mathbf{X}_j \epsilon_i$$

where Y_{i1} is the follow-up outcome of individual i, Y_{j0} is the endline outcome of peer j, Y_{i0} is the own endline outcome, $\mathbf{X}_i$ and $\mathbf{X}_j$ are own and peer covariates, and ϕ_j is peer centrality. η and ρ are the coefficients of interest that capture heterogeneous peer effects. Results of this specification are presented in Section 6. We also present variations of the above specification with village fixed effects and additionally controlling for the peer's outcome in the follow up survey.

4 Results

Table 1 shows the impact of the intervention on endline outcomes. We find that all treatment arms significantly increase the knowledge index by 0.7–0.8 standard deviations, with no differences across arms. Business aspirations are unaffected. Training alone raises the business index by 0.25 standard deviations (significant at 5%). The paired treatment with the connections module also increases the business index relative to the control, but differences are not significant when compared to other treatment arms. None of the treatments affect willingness to take additional steps or the take-up index. However, training alone leads to significantly more additional steps than training with a peer.

As a robustness check, Appendix Tables C.5–C.7 show that while ITT estimates using within-village controls are attenuated, IV specifications capturing local average treatment effects among compliers yield large effects for both training alone and training with a peer. However, in both cases, pairing continues to have no additional impact beyond training alone.

Table 1: Impact of the training on immediate outcomes.

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.789***	-0.139	0.253**	0.101	0.105
	(0.247)	(0.127)	(0.116)	(0.0878)	(0.0884)
Treatment with Peer	0.704***	-0.145	0.0991	-0.120	-0.0277
	(0.253)	(0.105)	(0.132)	(0.110)	(0.0953)
Treatment with Peer + Connections Module	0.799***	-0.177	0.210*	0.0511	-0.000710
	(0.244)	(0.117)	(0.111)	(0.0982)	(0.0909)
Constant	0	0	0	0	0
	(0.234)	(0.0932)	(0.0793)	(0.0616)	(0.0620)
Observations	768	768	765	746	751
R-squared	0.123	0.007	0.009	0.007	0.004
p: Trained Alone vs with Peer	0.183	0.951	0.255	0.0596	0.161
p: Trained Alone vs with Peer + CM	0.906	0.617	0.702	0.575	0.205
p: Treatment with Peer vs with Peer + CM	0.331	0.646	0.305	0.147	0.748

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The above regression treats the pure control group as the base category. Standard errors are robust and clustered at the village level.

We also find that the connections module leads to no statistically significant gains relative to other arms. This is despite the fact that individuals assigned to the connections treatment shared 6.5 contacts on average with each other and discussed how these connections can help them set up a business. In fact, as shown in the dyadic regressions in Table E.6, more connections were pooled among those with a larger social distance (i.e., those who have less overlap in their social networks). Moreover, those who shared more connections are more likely to report wanting to start a business together, even after controlling for social closeness and peer centrality (Table E.7). However, significantly more connections were shared between individuals in the same caste, and there is no significant correlation between the quadratic difference in degrees and the number of contacts pooled (Table E.6), suggesting that high-degree individuals did not preferentially share contacts with low-degree individuals. These patterns suggest that it may be difficult to engineer new network links in settings with high homophily. A Lasso regression (Figure E.3) shows that caste similarity is the strongest predictor of number of contacts pooled, followed by own degree centrality and education.

However, as we will discuss in Section 5.2, higher levels of caste homophily (in group linking bias) reduce contact sharing even within same-caste pairs, suggesting decreasing returns to sameness.[17] This is because when both individuals draw links from the same social circle and their social circles are more homophilous to begin with, there are fewer distinct connections to share in the first place. This highlights the limits of using the contact-sharing approach to build links in homophilous networks.

Next, Table C.1 shows the types of businesses individuals report that they want to open after being trained, conditional on being ready to invest. We find that those who are paired together during the training are significantly more likely to want to open an agricultural business together compared to the control group. This result could be consistent with the existence of the *parma* or *bola* system in Nepalese villages where peers support each other with agriculture-related tasks through reciprocity-based labour provision programs (Messerschmidt 1981, Sherpa 2005, Bhattarai 2006).

4.1 Effects of Peer Identity

We exploit variation in the identity of the matched peer in order to understand the conditions under which pairing can be helpful. We measure treatment effects using social distance and both absolute and relative measures of network centrality. First, we interact peer social distance and peer centrality with an indicator for being in the paired treatment arms. We find in Table 2(a) that higher social distance among peers significantly reduces endline aspirations. A one unit increase in social distance reduces aspirations by 0.08 standard deviations (significant at 1%). We then implement a similar specification where we use a discrete classification of social closeness as defined earlier and a relative comparison of an individual's own degree with their peer. Consistent with the previous result, Table 2(b) shows that being trained with a close peer increases aspirations by 0.28 standard deviations (significant at 5%) relative to the pure control group. The effect of

[17] Homophily is defined as the extent to which an individual's network overrepresents their own caste relative to its share in the village population, i.e., $\frac{\text{share of same-caste links} - \text{share of own caste in the population}}{\text{share of own caste in the population}}$.

being paired with a more central peer is not significant, except for the take-up index, where those matched with more central peers perform significantly better (significant at 10%).

While these specifications highlight the importance of social distance, they do not account for complementarity between social distance and network centrality i.e. is it better if the socially close peer is well connected?. To test this, we implement a two-by-two categorization by social distance and centrality: socially close peers that are more central than the individual, socially close peers that are less central, distant peers that are more central, and distant peers that are less central.

As shown in Table 2(c), those who are paired with a more central and close peer have higher aspirations ($+0.26\sigma$), significantly above the other pairs (p=0.01) and significantly above those treated alone (p=0.04). Other pair types have significantly lower aspirations than the control group (0.24 σ, p<0.05). This implies that a central peer helps motivate when the peer is socially close. In fact, Table C.2 shows that among socially close peers, being paired with a more central peer significantly increases aspirations relative to a less central peer. Consistent with this, as we will shortly discuss in Section 5, those who were matched with central peers were also significantly more likely to list encouragement as a reason why pairing was useful for them.

Table 2(c) further shows that those matched with a more central and close peer have a higher value of the take-up index ($+0.18$ σ), significantly higher compared to the other pairs (p=0.07). As with aspirations, socially close peers who are more central also lead to higher take-up than less central close peers (Table C.2).

In terms of other outcomes, Table 2(c) shows that individuals paired with a central and close peer have a significantly higher value of the business index ($+0.39$ σ, p<0.1) relative to the control group as well. While this effect is larger in magnitude than the effect of other pairs, the difference is not significant.

Figure 2 visualises the results in Table 2(c). Those matched with close and more central peers have significantly higher aspirations and take-up, and the remaining outcomes show a similar pattern, with these pairs outperforming both those trained alone and other pair types. Combined, these results suggest that central and close peers can boost motivation and the willingness to demand additional resources immediately after the training.

4.1.1 Popular Peers raise Aspirations:

Consistent with the above result on aspirations, Table C.4 shows that popular peers can act as motivators. We define popular peers as those with centrality greater than or

Table 2: Effects on Endline Outcomes by Peer Type

(a) By Social Distance and Peer Centrality

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.789***	-0.139	0.253**	0.101	0.105
	(0.248)	(0.127)	(0.116)	(0.0879)	(0.0886)
Trained with Peer	1.173***	0.242	0.336	0.00834	-0.135
	(0.283)	(0.178)	(0.245)	(0.157)	(0.192)
Trained with Peer × Peer distance	-0.0294	-0.0830***	-0.0235	0.00483	0.0203
	(0.0266)	(0.0223)	(0.0448)	(0.0303)	(0.0215)
Trained with Peer × Peer degree	-0.0255	-0.0241	-0.000329	0.0108	0.0295
	(0.0196)	(0.0170)	(0.0258)	(0.0155)	(0.0199)
Constant	0	0	0	-0	-0
	(0.235)	(0.0933)	(0.0794)	(0.0617)	(0.0621)
Observations	753	753	750	731	737
R-squared	0.166	0.015	0.019	0.019	0.020

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) By Social Closeness and More/Less Central

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.774***	-0.150	0.229*	0.0754	0.0618
	(0.247)	(0.127)	(0.113)	(0.0854)	(0.0917)
Trained with Peer	0.953***	-0.294**	0.224	0.0438	0.00682
	(0.247)	(0.124)	(0.133)	(0.103)	(0.0966)
Trained with Peer (Close)	-0.00626	0.281**	0.0394	-0.0456	-0.00880
	(0.0982)	(0.114)	(0.170)	(0.130)	(0.0678)
Trained with Peer (More Central)	-0.0611	0.0909	-0.0195	0.0443	0.132*
	(0.105)	(0.115)	(0.138)	(0.150)	(0.0690)
Degree	0.0181	0.0141	0.0384*	0.0318*	0.0283*
	(0.0148)	(0.0212)	(0.0193)	(0.0178)	(0.0149)
Constant	-0.0861	-0.0729	-0.182	-0.144	-0.130
	(0.244)	(0.107)	(0.135)	(0.116)	(0.0851)
Observations	756	756	753	734	740
R-squared	0.167	0.016	0.026	0.023	0.023

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

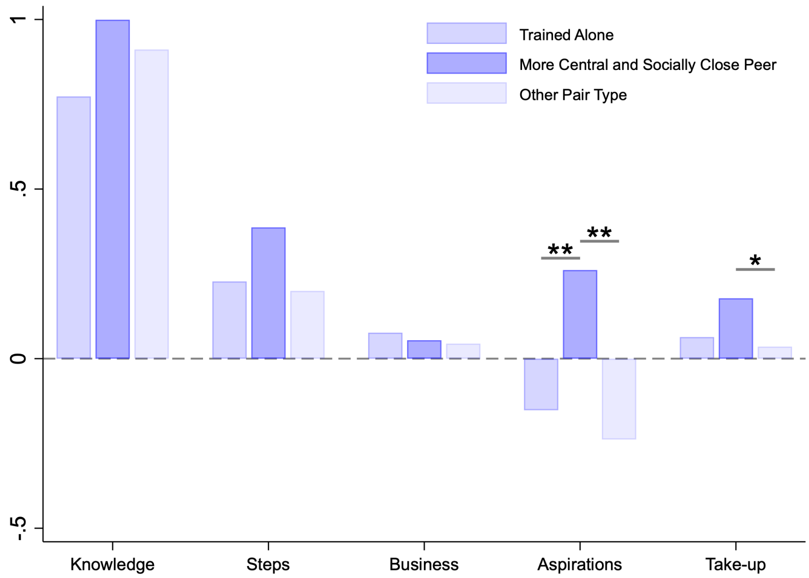
(c) By Social Closeness Interacted with More/Less Central

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.773***	-0.152	0.228*	0.0765	0.0636
	(0.248)	(0.127)	(0.114)	(0.0844)	(0.0910)
Close x More Central	0.999***	0.261	0.387*	0.0542	0.178
	(0.288)	(0.200)	(0.199)	(0.212)	(0.109)
Other Pair Types	0.911***	-0.238**	0.200*	0.0443	0.0355
	(0.244)	(0.112)	(0.113)	(0.0874)	(0.0941)
Degree	0.0207	0.0192	0.0410**	0.0297*	0.0247*
	(0.0134)	(0.0193)	(0.0177)	(0.0168)	(0.0136)
Constant	-0.0985	-0.0973	-0.195	-0.134	-0.113
	(0.242)	(0.103)	(0.129)	(0.111)	(0.0808)
Observations	756	756	753	734	740
R-squared	0.167	0.021	0.027	0.023	0.022
p: Central and Close vs Other Pairs	0.620	0.0106	0.417	0.967	0.0666
p: Central and Close vs T1	0.213	0.0391	0.426	0.910	0.266

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The above regressions treat individuals in the pure control villages as the base category and include an indicator for those who were intended to be paired but remained unmatched. We additionally control for the individual's own degree centrality in regressions that include relative comparisons between own and peer centrality. Standard errors are robust and clustered at the village level.

Figure 2: Effects on Short Term Outcomes by Centrality and Distance



The figure plots treatment effects on endline outcomes for those trained alone, those trained with a socially close and more central peer, and those trained with other pair types. Stars denote if pairwise differences are significant at the 1% (***), 5% (**), or 10% (*) significance levels respectively.

equal to the 75th quartile of that centrality measure in the respective village. Relative to other pair types, being paired with a popular and close peer consistently raises aspirations across various centrality measures: degree (p=0.027), indegree (p=0.052), and eigenvector (p=0.071), suggesting that central peers act as motivators. In contrast, there are no significant differences for knowledge, intentions, steps, or take-up.

4.1.2 Peer Characteristics and Homophily

Next, we consider other characteristics beyond centrality to understand whether the effect of network position of the peer persists even after controlling for other characteristics of the peer. We find that training with a central and close peer outperforms other pair types in terms of the effect on aspirations and take-up of additional resources, even after controlling for other peer characteristics and for the effects of homophily i.e. the effect of being paired with peers who are similar in baseline characteristics. Moreover, we also find that the effect of being matched with a more central and close peer also outperforms being trained alone (for the aspirations, take-up, and business index) once we account for homophily. These results are shown in Tables E.1, and E.2.

To show this, we first construct a similarity index to assess if the similarity between the peers along various characteristics leads to stronger treatment effects in the endline.

An individual in Treatments 2 or 3 (i.e. the paired treatment arms) is more similar to their peer if they are in the same income group, age group, caste, marital status, or education. We construct an index of these variables.[18] Table E.1 in the appendix shows these results, where we compare Treatment 1 with various pair types and additionally control for similarity, which is normalised to be between 0 and 1.[19] We find that the effects of being matched with a socially close peer who is central persist and are significantly higher than being trained alone for aspirations. The effect of close and more central peers on aspirations and take-up index is still significantly higher than the effect of being matched with other pair types. These results also persist when we instead control for peer characteristics, such as their caste, age, education, income, and marital status. Compared to being trained alone (T1), being trained with a central friend yields substantially larger improvements in aspirations even after peer characteristics are controlled for. These results are shown in Table E.2.

Importance of Network Variables using Random Forests

We complement the above findings by implementing a predictive exercise using random forests (Breiman 2001). For each pre-specified index, we train a random forest model for the peer treatment arms that includes only peer characteristics. These characteristics include both network measures (eg: peer degree, whether peer is more central, social distance, interaction between centrality and closeness) and demographics (peer education, income, age, marital status, and caste). To assess variable importance, we measure how much the model's prediction error increases when the values of a given predictor are randomly shuffled across observations. We then report the three predictors that reduce accuracy the most. This exercise is descriptive, but it highlights which peer traits systematically explain variation in outcomes. As seen in Panel A of Figure E.2, we find that the network characteristics of the peer are consistently selected as the most important predictors of the knowledge, aspirations, and business index.

4.1.3 Alternative Measures of Centrality

Finally, we show that the same pattern holds when we use alternative definitions of centrality. Results comparing treatment effects using various centrality measures are shown in Table C.3. Using indegree and eigenvector centrality, those matched with more central and socially close pairs report significantly higher aspirations than other pair types (p=0.083 and p=0.019, respectively), though there are no differences for take-up. As per betweenness centrality, aspirations do not differ, but those matched with more central and socially close pairs exhibit higher take-up (p=0.068). These patterns are consistent

[18] As before, we use the method proposed in Anderson (2008) to construct the index.

[19] This implies that similarity is equal to 1 for the most similar pairs in the sample.

with the results obtained using degree centrality in our analysis.

In Section F, we show that the heterogeneous effects of distance and centrality on aspirations and take up persist even after we consider an alternative measure of degree centrality accounting for partial sampling.

4.2 Long Term Effects

We now discuss the effects on outcomes measured one year after the program. About 3% of individuals in the follow-up sample had opened new businesses after one year of training. As shown in Table 3, training alone significantly improves the business steps index by 0.49 standard deviations. Training with a peer shows no effects apart from a marginal increase in the steps index relative to the pure control group (p value=0.109).[20] We also detect positive spillover effects among individuals in the spillover group within the same village. These effects are not significantly different from those trained alone, but they are higher for the steps and mindset indices compared to those trained with a peer. Importantly, these differences persist even after applying post-double selection lasso Belloni et al. (2014) (Table D.3) to account for any potential baseline differences.

The weak average effects in the peer arm likely reflect a combination of productive and unproductive matches, making it important to test for the heterogeneity.

Table 3: Effects on Long Term Outcomes

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Spillover	0.279	0.585***	0.280	0.138
	(0.190)	(0.195)	(0.203)	(0.147)
Trained Alone	0.194	0.490***	0.173	-0.0201
	(0.137)	(0.163)	(0.188)	(0.145)
Trained with Peer	0.103	0.215	0.0110	-0.0188
	(0.115)	(0.130)	(0.171)	(0.115)
Constant	-0	-0	-0	0
	(0.0700)	(0.0721)	(0.157)	(0.0858)
Observations	576	580	580	580
R-squared	0.007	0.028	0.010	0.003
p: Trained Alone v/s Spillover	0.642	0.639	0.482	0.318
p: Trained Alone v/s Trained with Peer	0.428	0.0799	0.123	0.991
p: Trained with Peer v/s Spillover	0.272	0.0533	0.0505	0.217

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: The above regression treats the pure control group as the base category. Standard errors are robust and clustered at the village level.

[20] As a robustness check, Appendix Table D.4 shows that IV estimates yield large effects of training alone on the steps index among compliers, while pairing again has no additional impact.

4.2.1 Effects of Peer Identity

To do this, as before, we decompose the effects by pair types using continuous and discrete measures of social distance and peer centrality as well as relative comparisons between own and peer centrality (see Tables 4(a)-(c)). In Table 4(a), we find that individuals trained with a peer with high social distance have significantly lower treatment effects on the outcomes index, mindset index, and business practices index. Table 4(b) shows that those matched with a more central peer are worse off, in that they have a significantly lower value of the steps index relative to those trained with a less central peer. Combined, these results suggest that the benefits of being matched with a close peer extend to the longer term, but the benefits of centrality are reversed. This is in contrast to the short-term results, where we detected strong motivational effects from training with a close and central peer. While central and close peers may initially boost motivation and engagement, less central but close peers might offer more durable support, potentially due to being able to interact more often. We will provide more evidence on these mechanisms in Section 5.

What happens when we interact distance and centrality? Unlike the endline results where we detected complementarity between social closeness and network centrality, Table 4(c) shows that the effects of being matched with a less central but close peer on all outcomes are in fact larger than other pair types, albeit not statistically significant. Table D.2 further decomposes these effects between the various pair types and shows that both the Steps and Mindset indices are substantially larger for those matched with socially close but less-central peers than with socially close and more-central peers. The difference between those matched with more central and less central peers is statistically significant for the Steps index ($p = 0.046$).

In fact, central and socially close peers have a significantly lower effect on the steps index compared to those treated alone (p value=0.04) as seen in Table D.1. This suggests that long-term benefits (in terms of taking steps to open a business) are more strongly associated with socially close peers, and the additional 'motivational' effect of being matched with a more central peer in the endline does not translate into improvements in the longer term. We will return to these follow-up results when we estimate peer effects in Section 6.

Importance of Network Variables using Random Forests: As before, we implement random forests for the paired treatment arms, and rank peer characteristics by how much the model prediction error increases when each variable is randomly permuted. Peer characteristics include both network measures (eg: peer degree, whether peer is more central, social distance, interaction between centrality and closeness) and demographics (peer education, income, age, marital status, and caste). As shown in Panel B of Figure

Table 4: Effects on Long-Term Outcomes

(a) By Social Distance and Peer Centrality

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Trained Alone	0.194	0.490***	0.173	-0.0201
	(0.137)	(0.164)	(0.188)	(0.146)
Trained with Peer	0.541*	0.355	0.489	0.305
	(0.316)	(0.385)	(0.343)	(0.219)
Trained with Peer × Peer distance	-0.0980**	0.0139	-0.101*	-0.0993**
	(0.0471)	(0.0707)	(0.0529)	(0.0456)
Trained with Peer × Peer degree	-0.0173	-0.0518	-0.0226	0.0107
	(0.0271)	(0.0326)	(0.0290)	(0.0222)
Constant	0	0	-0	-0
	(0.0702)	(0.0723)	(0.158)	(0.0860)
Observations	564	568	568	568
R-squared	0.012	0.034	0.018	0.012

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) By Social Closeness and More/Less Central

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Trained Alone	0.191	0.454***	0.164	-0.0234
	(0.135)	(0.150)	(0.181)	(0.146)
Trained with Peer	0.0167	0.324	0.00222	-0.0351
	(0.174)	(0.194)	(0.206)	(0.126)
Trained with Peer (Close)	0.413*	0.0211	0.177	0.219
	(0.217)	(0.183)	(0.248)	(0.174)
Trained with Peer (More Central)	-0.103	-0.460**	-0.0753	-0.00818
	(0.230)	(0.181)	(0.198)	(0.133)
Degree	0.00648	0.00657	0.00842	0.00346
	(0.0221)	(0.0257)	(0.0182)	(0.0251)
Constant	-0.0395	-0.0372	-0.0475	-0.0258
	(0.137)	(0.148)	(0.185)	(0.162)
Observations	566	570	570	570
R-squared	0.015	0.038	0.013	0.006

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(c) By Social Closeness Interacted with More/Less Central

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Spillover	0.289	0.593***	0.288	0.148
	(0.193)	(0.196)	(0.201)	(0.149)
Trained Alone	0.189	0.448***	0.164	-0.0243
	(0.135)	(0.150)	(0.181)	(0.145)
Close x Less Central	0.362	0.267	0.248	0.111
	(0.225)	(0.295)	(0.258)	(0.176)
Other Pair Types	0.0442	0.115	-0.0227	0.0104
	(0.133)	(0.140)	(0.168)	(0.128)
Degree	0.00955	0.0166	0.00817	0.00504
	(0.0221)	(0.0235)	(0.0195)	(0.0242)
Constant	-0.0546	-0.0864	-0.0463	-0.0335
	(0.136)	(0.138)	(0.188)	(0.158)
Observations	566	570	570	570
R-squared	0.011	0.031	0.014	0.004
p: Less Central and Socially Close vs Other Pairs	0.215	0.568	0.217	0.578
p: Less Central nd Socially Close vs T1	0.433	0.533	0.727	0.403

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats individuals in the pure control villages as the base category and includes an indicator for the spillover group and those who were intended to be paired but remained unmatched. We additionally control for the individual's own degree centrality in regressions that include relative comparisons between own and peer centrality. Standard errors are robust and clustered at the village level.

E.2, peer network characteristics consistently emerge as one of the top 3 most important predictors of follow-up outcomes, often surpassing peer demographic characteristics.

5 Mechanisms

In this section, we disentangle channels through which different types of peers affect outcomes in the short and long term, drawing on a variety of additional survey outcomes. We first examine short-run mechanisms—joint learning during training, encouragement, and pooling of contacts. We then turn to longer-run mechanisms, asking whether peers facilitate risk-sharing, sustain collaboration, and remain in contact one year later. Together, these results highlight the role of peer identity over time.

First, we asked the perceptions of women who attended the training in pairs as to how pairing was useful rather than training alone. Figure E.1 shows the responses to this question. 37% of women paired report having received encouragement from their partner. This is followed by 30% of the women who report that the training material was easier to grasp due to being paired in the training. About 18% say that the training helped because they shared network contacts. Importantly, the responses to this question do not differ by whether the individual was paired with a more central or socially close peer.

5.1 Do pairs learn better together?

5.1.1 Impact on knowledge during and after the training

We now proceed with disentangling mechanisms that can explain how pairing can help and why the effects are heterogeneous by centrality and distance. The training could be helpful for peers as they can help each other learn the material better. To evaluate this mechanism, we test how the intervention affects knowledge for different types of peers. Column 1 of Table E.3 shows that while the intervention improved the knowledge index, this does not differ by whether the individual was trained alone or with a peer.

Next, we also measure knowledge gained by the training by two additional variables measured in exercises conducted during the training: game profit and yearly profit. Game profit records the performance on the first day of the training, where women played an investment and savings game. Yearly Profit, on the other hand, was measured on the last day of the training, which is the amount of profit made in the exercise involving the creation of a business plan. Columns 2 and 3 of Table E.3 show that these endline outcomes do not differ by whether the individual was trained alone or with a peer.

It could also be the case that women who attended the training with a socially close peer may have been able to better discuss the material being taught. Table E.4 shows that the effect of being trained with a socially close peer on knowledge during training, measured in terms of performance in the profit-making exercises, is much higher than treatment 1 and much higher than being trained with a socially distant peer, and significantly so for profit in the business plan developed during the training. This suggests that socially close peers may be learning better together.

5.2 Do pairs provide access to network connections?

Gaining access to network connections by pooling network contacts is the second channel that we look at. Treatment 3, i.e., where women were paired during the training, and a connection module was designed precisely to check if this mechanism is at work. As previously discussed, this treatment fares better in terms of magnitude but the effects on most outcomes are not significantly different when compared to Treatment 1 or 2.

On average, women in Treatment 3 pool 6.5 contacts. Being from the same caste leads to significantly more contact sharing, with same-caste pairs listing about 1.3–1.5 additional shared contacts. As discussed earlier, pairs with a higher social distance also pool significantly more contacts (Table E.6). After controlling for social closeness and peer centrality, we find that the number of contacts listed increases the probability of reporting a willingness to start a business with the matched peer at endline.(Table E.7) However, we do not find heterogeneous effects by the number of contacts pooled in Treatment 3 (Table E.9). Taken together, this evidence suggests that pooling network contacts may not be a key mechanism and is likely constrained by caste boundaries. In fact, those who are in the same caste share a significantly higher number of contacts (1.3-1.5 more contacts on average) than other pairs (Table E.6). A Lasso regression (Figure E.3) identifies caste similarity as the main predictor of contact pooling in Treatment 3, followed by own centrality, degree gap, and similarity in education. The coefficient on caste remains stable even under higher values of the penalty parameters.

Impact of Homophily: We define caste homophily as the extent to which a woman's network over-represents her own caste relative to its share in the village population:

$$\text{Homophily} = \frac{\text{share of same-caste links} - \text{share of own caste in the population}}{\text{share of own caste in the population}}.$$

Higher values indicate stronger same-caste preference. Table E.8 shows that homophily is negatively and significantly associated with the number of contacts pooled, both among all participants who attended the connections module and among those randomly assigned to it. Same-caste pairs share more contacts on average, but this advantage declines sharply as caste homophily increases. This is consistent with decreasing returns to sameness as there is likely to be a lot of overlap in existing links. Inward-looking networks therefore limit the scope for contact sharing.

5.3 Do pairs help each other share risk?

Pairing can also improve outcomes if the peer is less risk-averse or if the peer provides financial assistance. We do not find this to be the case. First, we do not find any heterogeneous effects by baseline risk aversion of the matched peer. As seen in Table E.10, being paired with a peer who is less risk-averse does not impact the outcomes of the training. Further, one year following the training, we measured the number of savings groups that the respondent was a part of and whether they had joined a new cooperative in the village. We find that those treated alone, along with the spillover group, are significantly more likely to have joined a new cooperative, but the effect is not significant for those trained with a peer, suggesting that the paired treatments did not lead to an increased involvement with the wider network in terms of saving together and sharing risk. These results are shown in Table E.11.

5.4 Do pairs encourage each other?

Finally, a key channel driving peer effects in the endline survey could be encouragement. First, the most common benefit of pairing reported by respondents was encouragement, and it is significantly more likely to be reported by those whose peers have higher degree centrality or who are matched with a more central peer. This is shown in Table E.12, where we control for additional peer characteristics and find significant effects of peer centrality on encouragement being reported as a reason.

This is also corroborated by our main results discussed in the previous section: we detected a significant effect on aspirations for those matched with socially close peers, and among socially close peers, business aspirations are higher when paired with more central relative to less central peers (p value=0.05; Table C.2).

Finally, we define a popular individual as one whose centrality is greater than or equal to the village median. Results in Table C.4 show that individuals paired with 'popular' and close peers report significantly higher aspirations across all four centrality measures. Taken together, these findings suggest that central peers operate as motivators in the short term.

5.5 Are pairs likely to interact in the future?

Lastly, we find that individuals with lower social distance (i.e., socially close) to their matched peer are significantly more likely to want to open a business together and to meet in the future (Panel A of Table 5). In contrast, being assigned to a peer with high centrality does not differentially affect these outcomes.

One year later, we find that individuals who are trained with a peer are also significantly more likely than the control group to report discussing business-related concerns with anyone in the village. This is shown in Table E.13. When asked whether they spoke to their matched peer, we find that 31% report saying yes. When asked what they spoke about, 83% report speaking to the peer for advice (in general), and the remainder report interacting to borrow/lend money and about setting up businesses. However, as shown in Table E.13, these network interactions were significantly more likely than in the control group when the matched peer was less central (Column 2), and within the paired arm, when the matched peer was socially close (Column 3).

Importantly, as shown in Panel B of Table 5, those matched with a same-caste peer or with someone with low social distance are significantly more likely to have interacted with each other. Being matched with someone with high centrality does not affect these outcomes. This highlights the potential role played by the ease of collaboration among individuals of socially close individuals with a similar social standing and corroborates the previous empirical findings.

5.6 Summary of Mechanisms

Overall, our evidence points to two distinct mechanisms. Women paired with close peers are more likely to plan future meetings and state a willingness to open businesses, and maintain contact one year later when they are of the same caste. We find no evidence that close peers facilitated risk-sharing, suggesting that ease of collaboration is the key channel. This is also reinforced by the finding that close pairs perform better on in-training measures than those paired with distant peers.

Central peers, by contrast, matter in the short run because they act as motivators. Respondents paired with central peers were more likely to report encouragement as the reason why pairing was beneficial and displayed higher aspirations at endline. However, these effects disappear in the long run, with higher effects on the steps index among less central peers, consistent with the idea that central peers are less available for sustained collaboration. Taken together, the results indicate that motivation dominates in the short run while collaboration costs determine who is effective in the long run.

Table 5: Effect of Pair Type on Collaboration and Communication

(a) Willingness to Start a Business Together

VARIABLES	(1) Pairs will meet in Future	(2) Pairs will start Business Together	(3) Pairs will meet in Future	(4) Pairs will start Business Together
Peer Centrality	-0.00227	0.00349		
	(0.00720)	(0.0137)		
Social Distance	-0.0295**	-0.0278		
	(0.0132)	(0.0198)		
Paired (High Centrality)			-0.0402	0.00969
			(0.0475)	(0.0688)
Paired (Close)			0.0822**	0.146**
			(0.0415)	(0.0677)
Constant	0.977***	0.478***	0.836***	0.285***
	(0.0619)	(0.113)	(0.0465)	(0.0804)
Observations	325	324	331	330
R-squared	0.642	0.164	0.644	0.177

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) Talking to Matched Peer over 1 Year

VARIABLES	(1) Talk to matched peer	(2) Talk to matched peer	(3) Talk to matched peer	(4) Talk to matched peer
Peer Centrality	0.00472	0.00365		
	(0.0197)	(0.0194)		
Social Distance	-0.0465**	-0.0393*		
	(0.0235)	(0.0233)		
Same caste		0.247***		0.282***
		(0.0835)		(0.0801)
Paired (High Centrality)			-0.00122	-0.00736
			(0.0943)	(0.0909)
Paired (Close)			0.0853	0.0660
			(0.0961)	(0.0938)
Constant	0.472***	0.246	0.283**	0.0869
	(0.166)	(0.180)	(0.110)	(0.117)
Observations	177	177	181	181
R-squared	0.160	0.193	0.136	0.179

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Panel (a) restricts the endline sample to participants assigned to the paired treatment arms. Panel (b) restricts the follow-up sample to the paired treatment arms. Robust standard errors in parentheses.

6 Estimation of Peer Effects

We now present a simple model to understand the patterns in the data and estimate peer effects, exploiting the variation in the identity of the peer. Consider two matched women i and j who choose entrepreneurial effort $e_{it} \geq 0$ at time periods $t \in \{0, 1\}$, where $t = 0$ denotes the training/endline period and $t = 1$ denotes the one–year follow-up. Utility in period t can be written as follows:

$$U_{it} = \theta_t e_{it} - \frac{a}{2} e_{it}^2 \, + \, \beta_t \, M(c_{ij}, \phi_j) \, e_{it} e_{jt} \, - \, \kappa_t \, C(c_{ij}, \phi_j) \, e_{it} e_{jt}, \tag{1}$$

where $\theta > 0$, $a > 0$, c_{ij} is a measure of closeness between i and j and ϕ_j is peer j's centrality. The parameters β_t and κ_t measure the relative importance of motivation and collaboration frictions at horizon t, with $\kappa_1 > \kappa_0$ implying that maintaining collaborations post–training is more difficult than working together during the training.

Throughout, subscripts denote partial derivatives, so that $M_c \equiv \frac{\partial M(c,\phi)}{\partial c}$, $M_\phi \equiv \frac{\partial M(c,\phi)}{\partial \phi}$, $C_c \equiv \frac{\partial C(c,\phi)}{\partial c}$, and $C_\phi \equiv \frac{\partial C(c,\phi)}{\partial \phi}$. As per our empirical findings, $M_\phi(c,\phi) \geq 0$ (more central peers are more motivating), $M_c(c,\phi) \geq 0$ (closer peers are more motivating), $C_c(c,\phi) < 0$ (closeness reduces collaboration costs), and $C_\phi(c,\phi) > 0$ (collaborating with more central peers is costlier).

6.0.1 Best Responses and Equilibrium.

Differentiating (1) with respect to e_{it} yields

$$\frac{\partial U_{it}}{\partial e_{it}} = \theta_t - a e_{it} + \big[\beta_t M(c_{ij},\phi_j) - \kappa_t C(c_{ij},\phi_j)\big] e_{jt} = 0, \tag{2}$$

so the linear best response is

$$e_{it} = A_t + b_t(\phi_j, c_{ij})\, e_{jt}, \quad A_t = \frac{\theta_t}{a}, \quad b_t(\phi_j, c_{ij}) = \frac{\beta_t M(c_{ij},\phi_j) - \kappa_t C(c_{ij},\phi_j)}{a}. \tag{3}$$

Lemma: Equilibrium existence and uniqueness If $a > 0$ and $b_t(\phi_i, c_{ij})\, b_t(\phi_j, c_{ij}) < 1$, then a unique interior Nash equilibrium exists and is given by:

$$e_{it} = \frac{A_t + b_t(\phi_j, c_{ij}) A_t}{1 - b_t(\phi_i, c_{ij})\, b_t(\phi_j, c_{ij})}. \tag{4}$$

This can be seen by solving Equation 3 for $e_{jt}$and solving the system of equations.

Proposition 1: Closeness increases effort. Given $M_c(c_{ij},\phi) \geq 0$, $C_c(c_{ij},\phi) < 0$, and $b_t(\phi_i, c_{ij})\, b_t(\phi_j, c_{ij}) < 1$, if peer efforts are strategic complements, that is $b_t(\phi_m, c_{ij}) > 0$ for both $m \in \{i,j\}$ (equivalently $\beta_t M(c_{ij},\phi) > \kappa_t C(c_{ij},\phi)$), then holding (ϕ_i,ϕ_j) fixed,

$$\frac{\partial e_{it}}{\partial c_{ij}} \ > \ 0.$$

Intuitively, when the motivational effect of peer effort exceeds the cost of collaboration (i.e. $\beta_t M(c_{ij},\phi) > \kappa_t C(c_{ij},\phi)$), peer efforts are complements. In this case, greater closeness both raises motivation and lowers collaboration costs, and increases equilibrium effort.

Short vs. Long Run Differences in the Impact of Closeness Because $b_{t,c} = \frac{1}{a}\big(\beta_t M_c - \kappa_t C_c\big)$, closeness both strengthens complementarities and reduces collaboration frictions. A larger κ_t reduces $b_{t,c}$ and thus lowers the sensitivity of equilibrium effort to closeness. Hence, the positive effect of closeness is stronger at $t = 0$ (small κ_0) and weaker at $t = 1$ (larger κ_1). Intuitively, as it becomes more difficult to maintain collaborations after training, closeness matters less because collaboration frictions become more costly.

Proposition 2: Peer centrality increases effort if and only if motivation is higher than collaboration cost. Given $M_\phi(c_{ij}, \phi) \geq 0$, $C_\phi(c_{ij}, \phi) > 0$, and $b_t(\phi_i, c_{ij})\, b_t(\phi_j, c_{ij}) < 1$, if peer efforts are strategic complements, that is $b_t(\phi_m, c_{ij}) > 0$ for both $m \in \{i, j\}$, then the effect of peer centrality on effort is positive if and only if the motivational effect of centrality exceeds the associated collaboration cost, i.e.

$$\frac{\partial b_t(\phi_j, c_{ij})}{\partial \phi_j} = \frac{1}{a}\Big(\beta_t M_\phi(c_{ij}, \phi_j) - \kappa_t C_\phi(c_{ij}, \phi_j)\Big) \; > \; 0.$$

Intuitively, when effort choices are complements, higher peer centrality raises equilibrium effort only when it increases motivation more than it raises collaboration costs.

The next proposition highlights how the effect of peer centrality can vary over time due to its differential effects on motivation and collaboration.

Proposition 3: The effect of peer centrality can reverse over time If peer efforts are strategic complements, that is $b_t(\phi_m, c_{ij}) > 0$ for both $m \in \{i, j\}$, then higher peer centrality can increase effort in the short run but reduce it in the long run.

This is because at $t = 0$, when κ_0 is small enough, the motivational effect of centrality dominates and $\frac{\partial e_{i0}}{\partial \phi_j} > 0$. At $t = 1$, a larger κ_1 increases the relative cost of collaboration, and for sufficiently high κ_1 the effect can turn negative, so that $\frac{\partial e_{i1}}{\partial \phi_j} < 0$. Intuitively, while the motivational benefits of working with a central peer are immediate, the costs of coordination with a central (but likely busy) peer are larger in the longer term.

6.0.2 Correlational Evidence

We estimate the best response function above using data on all individuals who were paired during the training.[21]. This is to test if the magnitude of strategic complementarities might differ based on the network positions of the paired individuals. It is important to note that while random pairing mitigates concerns around correlated effects due to sorting, the lack of exogenous variation in effort does not allow us to solve the reflection

[21] Throughout this section, we do not exclude non-compliers or non-randomized participants, as the estimation of peer effects relies only on the random formation of pairs, established in Section 2.3.1

problem arising from the simultaneous effect of peer effort on own effort and vice versa (Manski 1993). However, we first conduct this exercise to check if the correlations in both time periods are in line with the predictions in the preceding empirical analysis. Then, we address this concern in Section 6.1 by testing how endline outcomes of peers affect the respondent's follow-up outcomes, thereby addressing the reflection problem.

We first estimate the equation for the first-order conditions for the endline and follow-up waves separately because the empirical analysis has previously shown that centrality plays a different role depending on the time frame when effects are observed. We use data from the paired treatments (i.e., Treatments 2 and 3) for this exercise. We construct an outcome measure in each wave to capture short-term effects by constructing a standardized Anderson Index (Anderson 2008) combining all the endline variables that enter the knowledge, business aspirations, business, steps, and take-up indices. Similarly, we combine all follow-up indices into a single outcome variable as well. We interpret these outcome indices as reduced-form proxies for entrepreneurial effort. They capture both their own actions and peer influence, as implied by the first-order condition. The indices summarize observable entrepreneurial activity, and the model is designed to capture how these measures vary with peer characteristics rather than to recover structural primitives.

Using these measures, we then estimate the following equation:

$$e_i^* = \theta + \theta_o \mathbf{X}_i + \theta_p \mathbf{X}_j + \theta_e e_j^* + \theta_d c_{ij} e_j^* + \theta_c \phi_j e_j^*$$

The vector $\mathbf{X}_i$ includes individual characteristics such as age, education, caste, income, and number of social connections, and $\mathbf{X}_j$ includes the analogous peer characteristics.

The results are shown in Table 6.[22] We find evidence of heterogeneity in the magnitude of strategic complementarity depending on how central the peers are. This effect persists in the endline even when we control for additional own and peer characteristics, including their age, caste, income, and education level. However, we find that this heterogeneity varies over time. As shown in the table, average outcomes are higher when agents are connected to more central peers in the endline, but these effects vanish in the follow-up. This is consistent with Proposition 3, where the collaboration cost term becomes more important in the follow-up, offsetting the motivational effect of central peers.

[22] The sample for the paired treatments in the follow-up is smaller, as we only include cases where we observe both the individual and their peer after 1 year.

Table 6: Effect of Closeness and Centrality in the Short and Long Run

VARIABLES	(1) Outcome (Endline)	(2) Outcome (Endline)	(3) Outcome (Follow-up)	(4) Outcome (Follow-up)
Peer Outcome	-0.0287	-0.0915	0.131	0.157
	(0.104)	(0.0951)	(0.172)	(0.170)
Close x Peer Outcome	-0.116	-0.0905	0.117	0.0751
	(0.0872)	(0.0760)	(0.153)	(0.158)
Peer Degree X Peer Outcome	0.0426*	0.0474**	-0.00431	-0.00515
	(0.0220)	(0.0187)	(0.0269)	(0.0264)
Constant	0.199***	1.047***	-0.000422	0.936
	(0.0424)	(0.272)	(0.0696)	(0.706)
Observations	454	452	186	186
R-squared	0.027	0.212	0.030	0.178

Notes: The above table shows the effect of the matched peer's outcome on the individual's own outcome. The first two columns use data from the endline wave, while the last two columns use data from the follow-up wave. Columns (2) and (4) additionally control for individual and peer characteristics.

6.1 Identification of Peer Effects

Next, as shown in Table 7, we identify peer effects by regressing the follow-up outcomes of the individual on the endline outcomes of their peers to mitigate the reflection problem Manski (1993), which is otherwise not addressed solely by random matching. As shown in the table, we find that peer centrality significantly reduces the effect of the endline peer outcome on own follow-up outcomes. A one–standard-deviation increase in the peer's endline outcome raises the respondent's follow-up outcome by 0.35–0.38 standard deviations. Each additional peer connection reduces this effect by about 0.06–0.08 standard deviations, implying that the peer effect is lower for more connected peers. This effect is significant even after controlling for the own and peers' demographic characteristics and additionally for village fixed effects. For sufficiently high levels of peer centrality, the net peer effect can even turn negative, consistent with higher collaboration costs outweighing motivational gains.

It could be argued that the peer's endline outcome is endogenous if it is correlated with their follow-up outcome, which in turn affects the individual's follow-up outcome. Table G.1 shows that the results remain robust when we control for the peer's own follow-up outcome.

Following the model, the reduced-form coefficient on peer outcomes can be interpreted as follows:

$$\rho_t = \frac{1}{a}[\beta_t M(c_{ij}, \phi_j) - \kappa_t C(c_{ij}, \phi_j)].$$

A positive ρ indicates that the motivational channel dominates collaboration frictions,

Table 7: Identification of Peer Effects

VARIABLES	(1) Outcome (Follow-up)	(2) Outcome (Follow-up)	(3) Outcome (Follow-up)
Peer Outcome (Endline)	0.381**	0.353**	0.376**
	(0.151)	(0.169)	(0.183)
Close X Peer Outcome	-0.0509	-0.129	-0.172
	(0.167)	(0.176)	(0.179)
Peer Degree X Peer Outcome	-0.0759**	-0.0672**	-0.0644**
	(0.0294)	(0.0315)	(0.0313)
Own Outcome (Endline)	0.257***	0.130**	0.124*
	(0.0628)	(0.0648)	(0.0707)
Constant	-0.0666	1.001*	0.483
	(0.0597)	(0.565)	(0.645)
Observations	288	288	288
R-squared	0.056	0.141	0.278
Own Characteristics	No	Yes	Yes
Peer Characteristics	No	Yes	Yes
Village FE	No	No	Yes

Notes: The table reports the effect of the matched peer's endline outcome on the individual's own follow-up outcome. Close × Peer Outcome interacts the peer outcome with an indicator for whether the peer was socially close. Peer Degree × Peer Outcome interacts the peer outcome with the peer's degree centrality. Columns (2) and (3) control for the individual's and peer's age, income, education, caste, and network degree. Column (3) additionally includes village fixed effects.

while a negative ρ implies that collaboration costs outweigh motivational gains. Our results suggest that κ_t (the impact of collaboration costs) is high in the longer term as maintaining collaboration with central peers becomes more difficult without corresponding improvements in motivation. This reverses the motivational effects observed at the endline.

6.1.1 Feasibility of Peer Pairings

While pairings based on social distance can be easy to implement, the feasibility of pairing based on network centralities and the potential returns from such pairings remain unclear. First, it is not clear if it is feasible to collect network data and engineer such pairs in the first place. We find that 63% of individuals in our setting guess the number of connections of their randomly matched peer within a range of ± 2 connections (≈ 1 standard deviation), and about 41% guess it correctly within a range of ± 1 connection. This aligns with findings in the literature showing that individuals can possess accurate knowledge about who is central in their network (Banerjee et al. 2019), and it may not require additional data collection.

Secondly, it is not clear how many such pairs we can form within a village in the first place, given the fixed social network. To show this, we simulate 10,000 counterfactual reassignments of peer pairings within each village and find that strategic pairings based on centralities are indeed feasible and could thereby generate high returns.

For each village, we first grouped individuals into randomly chosen dyads. Then, we computed the differences in degree centralities of each dyad, and computed the average degree centrality gap in the village. We then repeat this process 10,000 times for each village. The distribution of the average centrality gaps is plotted in Figure G.1.

The figure shows that while some villages have tightly clustered gaps, the majority display much wider distributions, implying that pairs with large gaps in centralities are indeed feasible to implement. These simulations confirm that strategic matching within villages, based on network position, is feasible and, in many cases, could be used to generate substantial treatment effects in the short-term.

7 Conclusion

In this paper, we show that the strength and direction of peer effects depend on peer identity, vary over time, and that peer effects are difficult to engineer in networks with high homophily. We show that the magnitude and mechanisms underlying peer effects depend on the network position of the peer: socially close peers are easier to collaborate with, while central peers provide short-term motivational benefits but do not sustain long-run collaboration. Using our experimental setup, we then estimate peer effects and document heterogeneity by network position that can even reverse the sign of the average effect. Combined, these results highlight that average peer effects, often estimated via linear-in-means models in the existing literature, can obscure heterogeneity that is both policy relevant and central to understanding underlying mechanisms.

We also use our experiment to compare the direct value of peers to their indirect value in terms of providing access to the wider social network. To do this, we study the effect of a treatment arm that encourages individuals to share their contacts and think of ways these contacts could help them set up a business. This treatment arm directly aims to facilitate new links through peers' existing networks. We find that contact sharing has limited benefits, as it is primarily concentrated among individuals in the same caste, likely preventing the formation of new ties. This highlights the difficulty of engineering new peer effects in settings with high homophily and points to future interventions that can test conditions under which more diverse and useful connections can be formed.

Our findings also highlight the benefits of networking strategically rather than at random, as peer identity affects motivation and collaboration in heterogeneous ways. While our results demonstrate that such strategic pairings are feasible and may not require additional network data, they do not capture the private costs of networking. These costs may be higher when interacting with central peers than with less central but socially close ones. In effect, our findings speak to the net benefit of interacting with different types of peers, but the private benefits may differ. Costs can range from informational (e.g., knowing who is central, which we show is not binding in our setting) to interactional (e.g., confidence, time). While we cannot measure such costs, reducing them through appropriate platforms may help leverage the benefits of strategic networking.

Appendix

A Baseline Results

Table A.1: Summary Statistics

	Mean	SD
Age	37.98	(10.85)
Marital Status		
Divorced	0.00141	(0.0375)
Married	0.918	(0.274)
Unmarried	0.0669	(0.250)
Widow	0.0134	(0.115)
Education		
No Education	0.326	(0.469)
Informal Education	0.126	(0.332)
Primary (Class 1-5)	0.155	(0.362)
Secondary (Class 6-10)	0.249	(0.432)
Higher Education (Class 11, 12)	0.104	(0.305)
University	0.0402	(0.196)
Belongs to Upper Caste	0.341	(0.474)
Degree Centrality	4.585	(2.140)
Eigen Vector Centrality	0.00978	(0.0133)
Own Non Agr. Business	0.220	(0.415)
Feel not Capable	0.277	(0.448)
Have no skills	0.496	(0.500)
Financial Reasons	0.239	(0.427)
No support from family	0.0129	(0.113)
Willing to Open a Business	0.419	(0.494)
Risk Aversion (1-6)	4.610	(1.406)
Aspirations (Agricultural Expenditure)	245802.6	(331604.3)
Aspirations (Non Agri. Expenditure)	443675.3	(861565.3)
Aspirations (Income)	141669.0	(658774.2)
Aspires to Higher Income	0.890	(0.313)
Aspires to Higher Non Agri. Exp	0.229	(0.420)
Observations	2840	

Notes: The above table reports summary statistics (i.e. mean and standard deviation) for baseline characteristics. The sample includes all baseline respondents with non-missing data.

Table A.2: Correlations between number of network connections and baseline characteristics.

	Degree Centrality
Age	-0.00957
Marital Status	
Divorced	0.0469*
Married	0.0800***
Unmarried	-0.0830***
Widow	-0.0257
Education	
No Education	-0.0744***
Informal Education	0.0404*
Primary (Class 1-5)	0.0216
Secondary (Class 6-10)	0.0847***
Higher Education (Class 11, 12)	-0.0402*
University	-0.0552**
Belongs to Upper Caste	0.0719***
Own Non Agr. Business	-0.0437*
Willing to Open a Business	0.0171
Risk Aversion (1-6)	-0.0681***
Aspirations (Agricultural Expenditure)	0.0236
Aspirations (Non Agri. Expenditure)	-0.0374
Aspirations (Income)	-0.0856***
Aspires to Higher Income	-0.0382
Aspires to Higher Non Agri. Exp	-0.0218

Notes: The above table reports variable-wise correlations between degree-centrality (i.e. number of connections in the social network) and other baseline characteristics.

Table A.3: Correlations between whether or not an individual has opened a business and their baseline characteristics.

	Own Non Agr. Business
Age	-0.0944***
Marital Status	
Divorced	0.00270
Married	-0.0354
Unmarried	0.0499**
Widow	-0.0250
Education	
No Education	-0.181***
Informal Education	-0.00791
Primary (Class 1-5)	0.00828
Secondary (Class 6-10)	0.00354
Higher Education (Class 11, 12)	0.129***
University	0.221***
Belongs to Upper Caste	-0.00378
Degree Centrality	-0.0437*
Eigen Vector Centrality	-0.0491**
Risk Aversion (1-6)	-0.0910***
Aspirations (Agricultural Expenditure)	-0.0364
Aspirations (Non Agri. Expenditure)	0.273***
Aspirations (Income)	0.155***
Aspires to Higher Income	-0.0429*
Aspires to Higher Non Agri. Exp	0.382***

Notes: The above table reports variable-wise correlations between whether an individual has already opened a non-agricultural business and other baseline characteristics.

Table A.4: Correlations between whether or not an individual is willing to open a business and their baseline characteristics.

	Willing to Open a Business
Age	-0.391***
Marital Status	
Divorced	0.0187
Married	-0.0508*
Unmarried	0.0965***
Widow	-0.0826***
Education	
No Education	-0.300***
Informal Education	-0.104***
Primary (Class 1-5)	0.0367
Secondary (Class 6-10)	0.251***
Higher Education (Class 11, 12)	0.167***
University	0.100***
Belongs to Upper Caste	-0.0541*
Degree Centrality	0.0171
Eigen Vector Centrality	0.00140
Own Non Agr. Business	0
Feel not Capable	-0.00718
Willing to Open a Business	1
Risk Aversion (1-6)	-0.103***
Aspirations (Agricultural Expenditure)	0.0455*
Aspirations (Non Agri. Expenditure)	0.167***
Aspirations (Income)	-0.0250
Aspires to Higher Income	0.0372
Aspires to Higher Non Agri. Exp	0.239***

Notes: The above table reports variable-wise correlations between baseline willingness to open a non-agricultural business and other baseline characteristics.

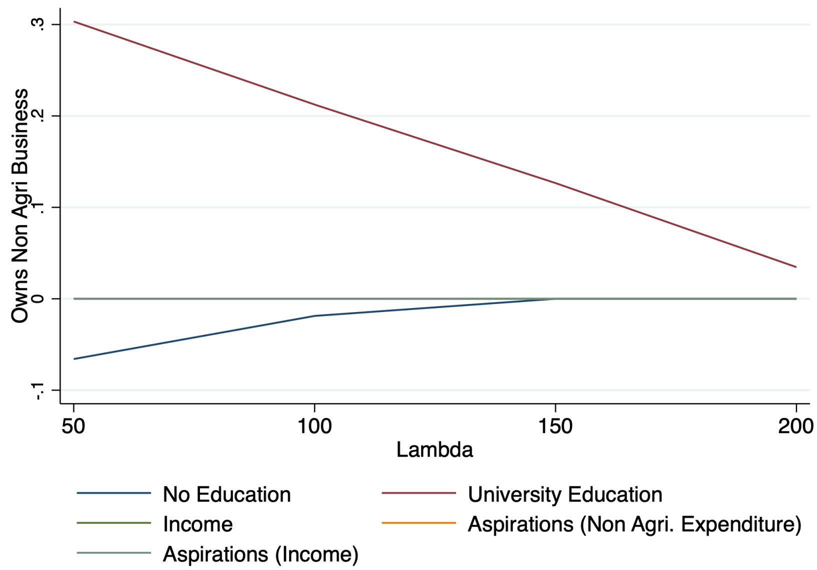


Figure A.1: The above graph plots correlations between selected variables and owning a business for different values of penalty parameters in a Lasso regression. Only variables selected out of a large list of demographic and network variables at $\lambda = 100$ are displayed. No Education and University Education are binary variables equal to 1 if the condition is true, and High Non. Agricultural Expenditure Aspirations is a binary variable equal to 1 if aspirations are higher than the baseline level.

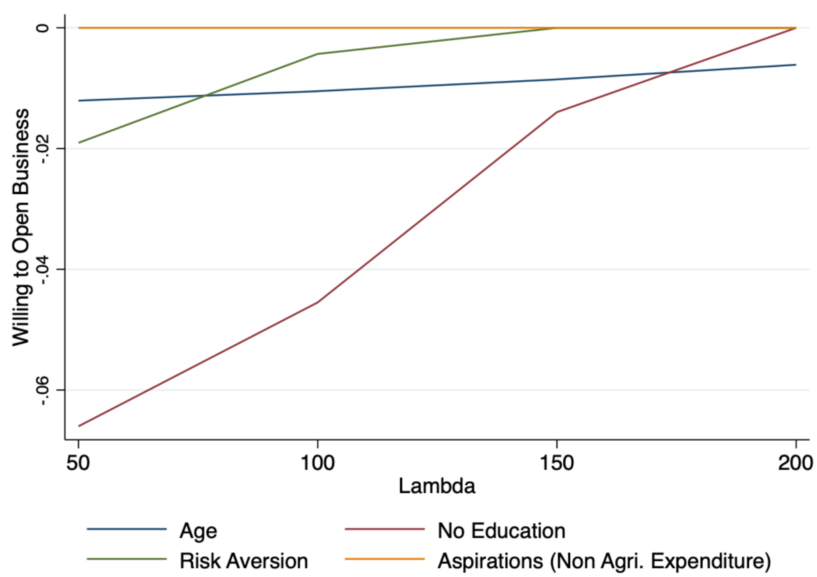


Figure A.2: The above graph plots correlations between selected variables and willingness to open businesses for different values of penalty parameters in a Lasso regression. Only variables selected out of a large list of demographic and network variables at $\lambda = 100$ are displayed. No Education is a binary variable equal to 1 if true. Risk Aversion is measured on a scale of 1-6 using a series of choices between lotteries and fixed outcomes presented to participants in the baseline survey.

Figure A.3: Exeriment Design

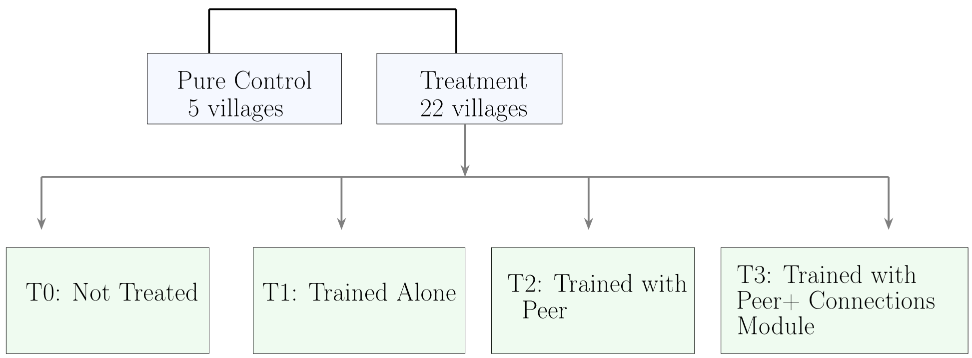


Figure A.4: The figure illustrates the random assignment of 27 villages into pure control and treatment groups. Within treatment villages, individuals were further randomized into one of three arms: trained alone (T1), trained with a peer (T2), or trained with a peer plus a connections module (T3).

Figure A.5: Motivation behind the Connections Module

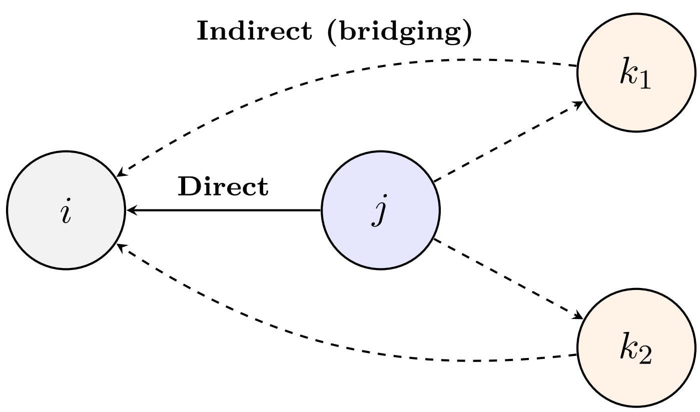


Figure: The solid arrow shows how peer j directly influences i (eg: via motivation, learning). The dashed arrows indicate how peer j indirectly benefits i by connecting her to contacts k_1 and k_2.

B Implementation

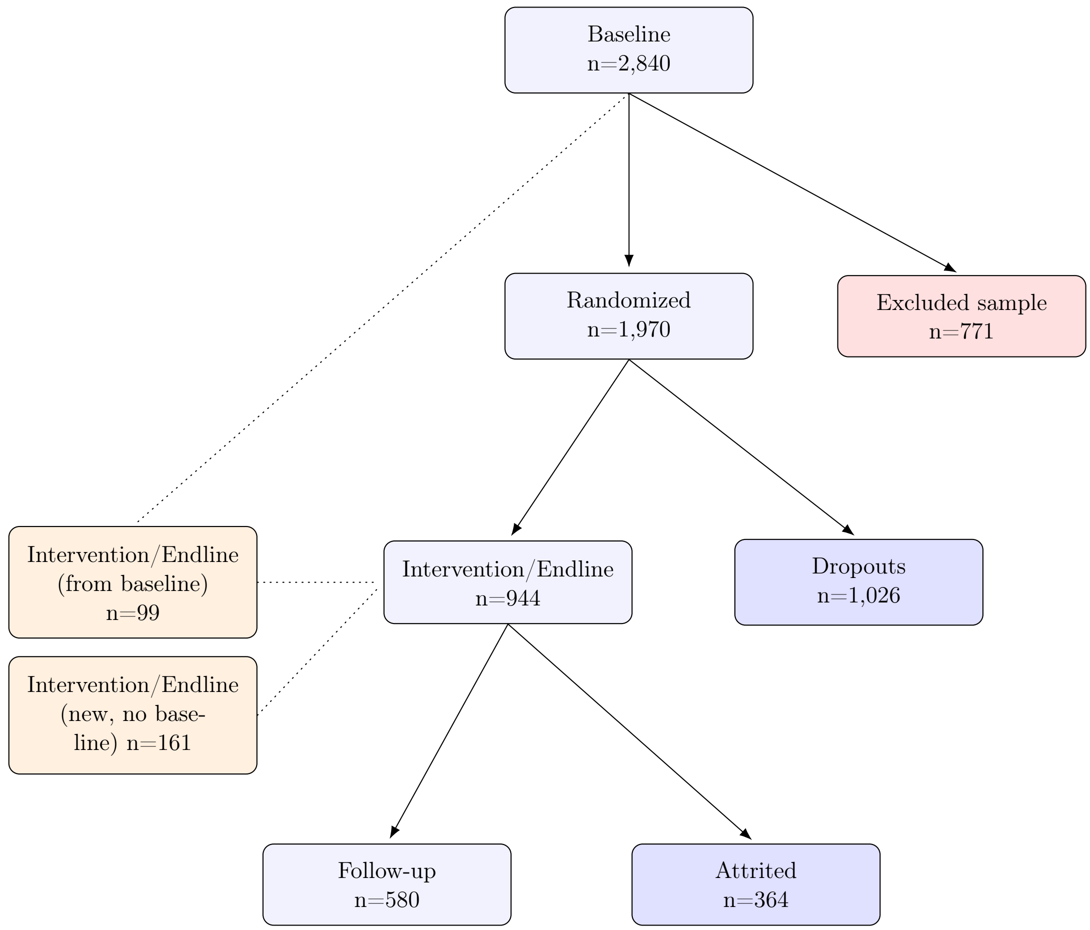


Figure B.1: The figure shows participant numbers at each stage of data collection. From the baseline sample (n = 2,840), 1,970 women were randomized into treatment arms, while 771 were excluded ex-ante. Of those randomized, 944 participated in the intervention and endline survey, and 580 were successfully followed up one year later. The "Intervention/Endline" box also includes 99 baseline respondents who attended despite not being randomized and 161 new participants without baseline data. These non-randomized individuals are excluded from all intent-to-treat analyses.

B.1 Compliance

Compliance is defined as a binary indicator equal to one if the individual's realized treatment status matched their assignment. Overall compliance in the endline sample is 68%, or 58% when excluding pure control villages. Overall compliance in the entire randomized sample is 51%. This includes those whom we could not find and who do not appear in the endline survey (i.e., dropouts), as well as those who did not adhere to their assigned treatment, e.g. did not get trained despite being assigned to training.

Compliance is not correlated with treatment status in the endline sample (Column 2 in Table B.1) but this pattern disappears if we exclude the pure control villages (Column 2). In that case, compliance is higher for those assigned to treatment, i.e., those invited for training were more likely to attend training compared to compliance among those asked not to attend. We do not find any key differences in the characteristics of compliers and non-compliers in this subsample, as shown in Table B.2, except that compliers are more likely to be married.

To address any issues arising due to compliance, we present intention-to-treat (ITT) estimates in the main results to avoid any bias from endogenous selection into training. Moreover, we also show how the endline and follow-up sample are balanced along a long list of baseline characteristics.

Table B.1: Correlation of Compliance with Treatment Status (Endline Sample)

VARIABLES	(1) Compliance	(2) Compliance
Trained Alone	-0.0562	0.301***
	(0.0929)	(0.0501)
Trained with Peer	-0.105	0.252***
	(0.0932)	(0.0511)
Trained with Peer + CM	-0.0982	0.259***
	(0.0955)	(0.0528)
Constant	0.732***	0.375***
	(0.0849)	(0.0366)
Observations	944	710
R-squared	0.010	0.058
Trained Alone vs Paired	0.276	0.327
Paired vs Paired+CM	0.868	0.901
Trained Alone vs Paired+CM	0.508	0.413

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Compliance is defined as a binary measure indicating adherence to the assigned treatment in the endline sample. Column 2 excludes the pure control villages.

Table B.2: Baseline Characteristics of Compliers in Treated Villages (Endline Sample)

	0	1	(1) vs. (2), p-value
Age	38.649	38.849	0.795
Income	24459.091	24311.509	0.940
No Education	0.352	0.366	0.712
Informal Education	0.154	0.156	0.950
Primary Education	0.141	0.176	0.216
Secondary Education	0.265	0.239	0.430
Higher Education	0.081	0.051	0.115
University Education	0.007	0.012	0.467
Upper Caste	0.331	0.324	0.834
Married	0.920	0.954	0.061
Divorced	0.000	0.000	
Widow	0.023	0.012	0.252
Degree Centrality	5.144	5.359	0.245
Willing to Open a Business	0.452	0.404	0.211
Risk Aversion (1-6)	4.525	4.781	0.015

Notes: This table compares baseline characteristics of individuals who complied with the assigned treatment with those who did not, in treated villages, excluding dropouts (i.e. who are not present in the endline survey).

B.2 Endline Attrition

Table B.3 examines whether dropouts from our randomized sample (N=1970), i.e., those who did not appear in the endline survey (N=944), differ by treatment status. In the full sample (Column 1), dropout rates are higher among those assigned to treatment compared to the pooled control group, which includes both controls in treated villages and households in pure control villages. Once we restrict attention to treated villages (Columns 2–3), this pattern disappears: dropout rates are not systematically different between those assigned to treatment and those in the control group within the same village. This suggests that the higher attrition observed in Column 1 reflects survey modality rather than selection related to treatment assignment i.e., surveys were largely conducted at training centers in treatment villages but at households in pure control villages. Importantly, survey attrition does not differ between those trained alone and those trained with a peer, which is the main comparison of interest. While dropout is somewhat higher in the Peer + Connections Module arm relative to the Alone and Peer arms, the differences are only marginally significant, and participants were not informed ex-ante about the module.

Table B.3: Initial Dropouts by Treatment Status

VARIABLES	(1) Initial Dropout	(2) Initial Dropout	(3) Initial Dropout
Trained Alone	0.0694*	-0.0276	-0.0276
	(0.0386)	(0.0351)	(0.0351)
Trained with Peer	0.0757*	-0.0213	
	(0.0411)	(0.0351)	
Trained with Peer + CM	0.133***	0.0364	
	(0.0398)	(0.0349)	
Paired			0.00748
			(0.0304)
Constant	0.464***	0.561***	0.561***
	(0.0359)	(0.0248)	(0.0248)
Observations	1,970	1,606	1,606
R-squared	0.010	0.003	0.001
p-value: Alone vs. Peer	0.830	0.858	
p-value: Peer vs. Peer+CM	0.0291	0.0991	
p-value: Alone vs. Peer+CM	0.0488	0.0673	
p-value: Alone vs. Paired			0.249

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table compares the proportion of dropouts by intended treatment status within the randomized sample where Column 2 excludes the pure control villages and Column 3 combines the paired treatment arms into one variable.

Table B.4 further compares baseline characteristics of individuals who dropped out before the endline to those who remained. Dropouts show significant differences in terms of education, are less likely to be married, and are significantly less central in the village network. Other characteristics, including age, income, caste, risk aversion, and baseline willingness to open a business, are balanced. Baseline balance checks between the treated and control groups (Appendix Section B.6) show that these initial dropouts do not result in systematic selection across key variables, including baseline willingness to open a business, which is in fact, higher in the control group.

Table B.4: Baseline Characteristics of Dropouts

	0	1	(1) vs. (2), p-value
Age	38.765	37.970	0.152
Income	24373.855	25470.248	0.427
No Education	0.360	0.401	0.098
Informal Education	0.155	0.094	0.000
Primary Education	0.161	0.142	0.284
Secondary Education	0.250	0.249	0.959
Higher Education	0.064	0.092	0.039
University Education	0.010	0.023	0.040
Upper Caste	0.327	0.324	0.907
Married	0.939	0.913	0.046
Divorced	0.000	0.002	0.208
Widow	0.017	0.015	0.701
Degree Centrality	5.268	4.245	0.000
Willing to Open a Business	0.424	0.422	0.931
Risk Aversion (1-6)	4.673	4.676	0.972

Notes: This table compares baseline characteristics of individuals who dropped out before the endline with those who remained in the sample.

B.3 Follow-up Attrition

Table B.3 tests whether attrition in the follow-up survey differs by treatment assignment. Attrition is 40% overall (constant) and does not vary significantly across Spillover, Trained Alone, or Trained with Peer. Pairwise tests confirm no differential attrition. This reduces concern that long-run results are driven by selection.

Table B.5: Correlation of Follow-up Survey Attrition with Treatment Status

VARIABLES	(1) Attrition
Spillover	0.00170
	(0.0511)
Trained alone	-0.0506
	(0.0559)
Trained with Peer	-0.0173
	(0.0560)
Constant	0.402***
	(0.0372)
Observations	944
R-squared	0.002
Spillover==T1	0.368
T1=Paired	0.623
Spillover=Paired	0.752

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Attrition is a binary variable equal to 1 if the individual is observed in the endline survey but not in the follow up. The above regression compares attrition rates among individuals in the spillover, T1, and paired treatment arms with those in the pure control group. Standard errors are robust and clustered at the village level.

B.4 Balance Tables

Table B.6: Balance Test for the Endline Sample

	(1) 0	(2) 1	(3) 2	(4) 3	(5) (1) vs. (2), p-value	(6) (1) vs. (3), p-value	(7) (1) vs. (4), p-value	(8) (2) vs. (3), p-value	(9) (2) vs. (4), p-value	(10) (3) vs. (4), p-value
Willing to Open Business	0.43	0.39	0.40	0.45	0.57	0.64	0.72	0.91	0.27	0.35
Income	24094.72	25317.20	25653.33	23907.65	0.74	0.63	0.95	0.92	0.69	0.56
Income Source- Agri.	0.92	0.87	0.84	0.91	0.27	0.12	0.88	0.49	0.18	0.07
Income Source- Business	0.11	0.06	0.10	0.10	0.31	0.88	0.85	0.20	0.18	0.91
Income Source- Job	0.02	0.02	0.01	0.03	0.99	0.23	0.64	0.33	0.66	0.36
Income Source- Remit.	0.01	0.01	0.01	0.02	0.67	0.80	0.37	0.56	0.11	0.54
Income Source- Other	0.04	0.03	0.04	0.01	0.65	0.83	0.16	0.76	0.27	0.07
Age	39.12	39.70	38.45	39.96	0.73	0.67	0.63	0.24	0.75	0.16
Elementary Education	0.18	0.15	0.19	0.16	0.36	0.67	0.68	0.33	0.70	0.49
Higher Education	0.09	0.09	0.05	0.07	0.99	0.37	0.66	0.29	0.43	0.54
Informal Education	0.32	0.31	0.30	0.33	0.89	0.77	0.82	0.81	0.63	0.54
Univeristy Education	0.01	0.01	0.01	0.01	0.74	0.75	0.82	0.99	0.92	0.92
Secondary Education	0.28	0.23	0.27	0.21	0.24	0.79	0.15	0.19	0.55	0.13
Degree	4.78	5.28	5.29	5.32	0.07	0.10	0.02	0.96	0.86	0.90
Brahmin	0.06	0.05	0.07	0.03	0.85	0.87	0.63	0.29	0.62	0.10
Chhetri	0.23	0.30	0.23	0.27	0.66	1.00	0.78	0.07	0.48	0.22
Dalit	0.06	0.01	0.04	0.01	0.21	0.72	0.22	0.31	0.92	0.23
Newar	0.46	0.41	0.50	0.45	0.86	0.87	0.99	0.16	0.39	0.42
Janjati	0.19	0.23	0.16	0.24	0.79	0.80	0.74	0.09	0.88	0.11

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The balance tests compare characteristics of the pure control group (0) with those in Treatment 1, Treatment 2, and Treatment 3 for the endline sample. The p-values in Columns 5-10 indicate if the difference is statistically significant.

Table B.7: Balance Test for the Follow-up Sample

	(1) 0	(2) 1	(3) 2	(4) 3	(5) (1) vs. (2), p-value	(6) (1) vs. (3), p-value	(7) (1) vs. (4), p-value	(8) (2) vs. (3), p-value	(9) (2) vs. (4), p-value	(10) (3) vs. (4), p-value
Income	25545.029	27694.207	25364.717	23304.280	0.585	0.946	0.507	0.612	0.410	0.396
Income Source- Agri.	0.865	0.869	0.842	0.968	0.934	0.601	0.003	0.672	0.010	0.033
Income Source- Business	0.024	0.000	0.017	0.000	0.123	0.689	0.123	0.171	0.171	
Income Source- Job	0.037	0.008	0.017	0.032	0.054	0.216	0.754	0.569	0.122	0.465
Income Source- Remit.	0.012	0.008	0.017	0.022	0.718	0.698	0.586	0.569	0.255	0.804
Income Source- Other	0.057	0.025	0.042	0.011	0.191	0.491	0.019	0.387	0.520	0.114
Age	39.861	39.459	38.975	38.430	0.781	0.501	0.361	0.626	0.307	0.650
Elementary Education	0.155	0.180	0.200	0.161	0.512	0.285	0.898	0.659	0.738	0.479
Higher Education	0.078	0.090	0.058	0.086	0.749	0.587	0.830	0.333	0.884	0.314
Informal Education	0.314	0.270	0.325	0.333	0.483	0.875	0.789	0.471	0.346	0.916
Univeristy Education	0.008	0.008	0.008	0.011	0.997	0.987	0.833	0.991	0.854	0.860
Secondary Education	0.314	0.221	0.242	0.269	0.050	0.190	0.458	0.650	0.265	0.648
Degree	4.807	5.438	5.350	5.226	0.009	0.036	0.045	0.743	0.466	0.594
Brahmin	0.073	0.041	0.050	0.022	0.531	0.610	0.233	0.757	0.559	0.160
Chhetri	0.257	0.311	0.225	0.301	0.616	0.760	0.700	0.102	0.857	0.146
Newar	0.196	0.238	0.158	0.237	0.586	0.604	0.661	0.052	0.986	0.264
Dalit	0.424	0.402	0.542	0.430	0.871	0.440	0.970	0.054	0.668	0.070
Janjati	0.049	0.008	0.025	0.011	0.264	0.553	0.283	0.535	0.858	0.327

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

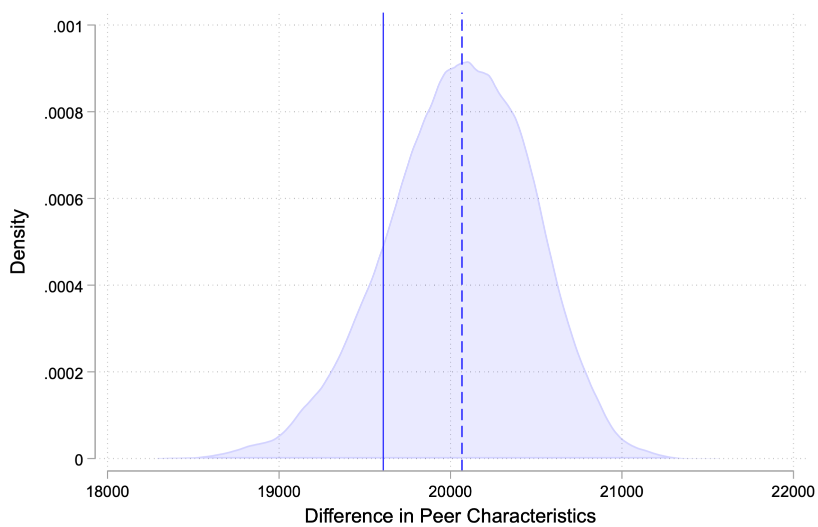
Notes: The balance tests compare characteristics of those in the pure control group (0) with those in the spillover group(1), Treatment 1 (2), and paired treatment arms (3) for the follow-up sample. The p-values in Columns 5-10 indicate if the difference is statistically significant.

B.5 Randomization Inference for Peer Assignment

We test whether the peer pairs created during the intervention are systematically more or less similar than would be expected under random assignment. To do this, we construct a statistic that captures the overall difference between peers across a set of characteristics including income, source of income, age, education level, caste, marital status, willingness to start a business, income aspirations, agricultural expenditure aspirations, overall degree, outdegree, and risk aversion. In each iteration of the simulation, we match individuals within each village into dyads, compute the sum of the quadratic differences in their baseline characteristics, and add this across the entire sample.

Comparing the observed statistic to the distribution obtained from the simulations allows us to assess whether actual peers are unusually similar or dissimilar. The distribution and the statistic from the actual assignment is plotted in Figure B.2, indicating that the observed pairings are not systematically different from what random assignment would generate.

Figure B.2: Randomization Inference on Peer Assignment



Notes: The histogram shows the distribution of peer similarity under 10,000 random re-assignments. The solid vertical line marks the statistic computed from actual peer pairs and the dashed vertical line plots the median difference computed from the re-assignments.

C Endline Results

Table C.1: Types of Businesses that individuals are willing to open

VARIABLES	(1) Agricultural Business	(2) Sewing	(3) Shop/Parlor	(4) Other Business
Trained alone	0.0940	0.0166	-0.146	0.0356
	(0.0639)	(0.0528)	(0.0878)	(0.0243)
Trained with Peer	0.174***	-0.0441	-0.148*	0.0188
	(0.0509)	(0.0472)	(0.0823)	(0.0151)
Constant	0.468***	0.144***	0.369***	0.0180*
	(0.0414)	(0.0410)	(0.0791)	(0.00882)
Observations	413	413	413	413
R-squared	0.021	0.006	0.022	0.005

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the pure control group as the base category. Business type is equal to 1 if the individual reports wanting to open such a business and reports being ready to invest. Standard errors are robust and clustered at the village level.

Table C.2: Effects on Short Term Outcomes by Type of Peer (All)

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.774***	-0.150	0.229*	0.0754	0.0617
	(0.248)	(0.127)	(0.114)	(0.0854)	(0.0918)
Close x More Central	0.999***	0.261	0.387*	0.0544	0.178
	(0.288)	(0.200)	(0.199)	(0.212)	(0.110)
Close x Less Central	0.860***	-0.152	0.155	-0.0107	-0.0382
	(0.259)	(0.169)	(0.135)	(0.123)	(0.132)
Far x Central	0.819***	-0.321**	0.113	0.0802	0.107
	(0.257)	(0.124)	(0.184)	(0.168)	(0.103)
Far x Non Central	0.995***	-0.226*	0.277**	0.0483	0.0254
	(0.246)	(0.119)	(0.119)	(0.0905)	(0.0970)
Degree	0.0185	0.0149	0.0389*	0.0319*	0.0285*
	(0.0149)	(0.0213)	(0.0192)	(0.0178)	(0.0149)
Constant	-0.0883	-0.0765	-0.185	-0.145	-0.131
	(0.244)	(0.107)	(0.135)	(0.116)	(0.0851)
Observations	756	756	753	734	740
R-squared	0.169	0.022	0.028	0.023	0.023
p: Central and Close vs Less Central and Close	0.541	0.0557	0.275	0.806	0.0580
p: Central and Close vs Central and Distant	0.314	0.00657	0.308	0.929	0.377
p: Central and Close vs Less Central and Distant	0.981	0.0151	0.663	0.980	0.0908
p: Socially Close vs Distant	0.823	0.0108	0.680	0.782	0.954
p: Less Central vs More Central	0.879	0.198	0.815	0.766	0.0381

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the pure control group as the base category. Business type is equal to 1 if the individual reports wanting to open such a business and reports being ready to invest. Standard errors are robust and clustered at the village level.

Table C.3: Impact of training using alternative centrality measures

Centrality	Knowledge	Aspirations	Intention	Steps	Take-up
Treated Alone	0.773**	-0.152	0.228**	0.077	0.064
Degree:Close x More Central	0.999***	0.261	0.387*	0.054	0.178
Degree:Other Pair Types	0.911***	-0.238**	0.200*	0.044	0.035
More Central and Socially Close v/s Other Pairs	0.620	0.011	0.417	0.967	0.067
More Central and Socially Close v/s Trained Alone	0.213	0.039	0.426	0.910	0.266
Treated Alone	0.772**	-0.152	0.226**	0.074	0.062
Indegree:Close x More Central	1.019***	0.129	0.356*	0.012	0.082
Indegree:Other Pair Types	0.911***	-0.200*	0.208*	0.044	0.050
More Central and Socially Close v/s Other Pairs	0.497	0.083	0.534	0.886	0.748
More Central and Socially Close v/s Trained Alone	0.174	0.158	0.515	0.728	0.864
Treated Alone	0.777**	-0.153	0.237**	0.077	0.067
Eigenvector:Close x More Central	0.930***	0.125	0.340**	0.070	0.002
Eigenvector:Other Pair Types	0.930***	-0.232**	0.220*	0.043	0.077
More Central and Socially Close v/s Other Pairs	0.999	0.019	0.558	0.880	0.330
More Central and Socially Close v/s Trained Alone	0.192	0.087	0.537	0.957	0.511
Treated Alone	0.762**	-0.142	0.227**	0.080	0.068
Betweenness:Close x More Central	0.920***	-0.030	0.376**	0.139	0.161*
Betweenness:Other Pair Types	0.914***	-0.175*	0.198*	0.033	0.043
More Central and Socially Close v/s Other Pairs	0.962	0.298	0.219	0.469	0.068
More Central and Socially Close v/s Trained Alone	0.342	0.483	0.327	0.650	0.289

Notes: Each block of four rows corresponds to the results of a regression using the corresponding centrality measure (Degree, In-degree, Eigenvector, Betweenness). In each regression we estimate outcomes on: (i) an indicator for being trained alone (T1), (ii) an indicator for being paired with a socially close peer who is more central than the respondent ("Close $\times$ More Central"), and (iii) an indicator for all other pair types. The coefficients reported are relative to the pure control group. The bottom two rows in each block report p-values from tests comparing "Close $\times$ More Central" to (a) other pair types and (b) trained-alone individuals. Standard errors are clustered at the village level.

Table C.4: Impact of training using alternative definition of peer popularity

Centrality	Knowledge	Aspirations	Intention	Steps	Take-up
Treated Alone	0.773**	-0.148	0.229**	0.076	0.065
Degree:Popular x Close	0.893***	0.131	0.290*	0.039	0.155
Degree:Other Pair Types	0.935***	-0.244**	0.212*	0.048	0.030
Popular and Socially Close v/s Trained Alone	0.308	0.094	0.699	0.789	0.413
	0.741	0.027	0.671	0.962	0.165
Treated Alone	0.773**	-0.149	0.227**	0.073	0.062
Indegree:Popular x Close	0.952***	0.111	0.258*	-0.052	0.133
Indegree:Other Pair Types	0.918***	-0.227**	0.218*	0.063	0.034
Popular and Socially Close v/s Trained Alone	0.226	0.169	0.855	0.308	0.504
	0.780	0.052	0.816	0.433	0.224
Treated Alone	0.776**	-0.151	0.236**	0.075	0.067
Eigenvector:Popular x Close	0.790**	0.018	0.117	-0.088	0.054
Eigenvector:Other Pair Types	0.966***	-0.214**	0.275**	0.084	0.065
Popular and Socially Close v/s Trained Alone	0.916	0.200	0.515	0.300	0.927
	0.150	0.071	0.396	0.308	0.916
Treated Alone	0.762**	-0.141	0.228**	0.081	0.069
Betweenness:Popular x Close	0.830**	-0.176	0.251	0.109	0.193
Betweenness:Other Pair Types	0.933***	-0.141	0.229**	0.041	0.038
Popular and Socially Close v/s Trained Alone	0.702	0.841	0.919	0.867	0.341
	0.540	0.819	0.918	0.714	0.145

Notes: Each block of four rows corresponds to a separate regression using one centrality measure (Degree, In-degree, Eigenvector, Betweenness). For each outcome we regress on: (i) an indicator for being trained alone (T1), (ii) an indicator for being paired with a peer who is both *Popular* (equal to or above median centrality in the village network for that measure) and *Close* (network distance ≤ 2), and (iii) an indicator for all other pair types. Reported coefficients are relative to the pure control group. The last two rows in each block report p-values from tests comparing "Popular $\times$ Close" to (a) Other Pair Types and (b) Trained Alone. Standard errors clustered at the village level.

C.1 Comparison with Within-Village Controls and LATE

We report a set of robustness exercises that complement the main ITT estimates in Tables C.5- C.7. We begin by estimating ITT effects using within-village controls as the comparison group (Table C.5). Consistent with attenuation expected from compliance concerns within the village, these coefficients are generally small, with the exception of a limited effect of being trained alone on the take-up index. We then instrument actual treatment status with assignment (Table C.6). The first-stage F–statistics are high and the instruments are strong, yielding large treatment effects among compliers. However, the effect of pairing on average remains insignificant when compared to those trained alone, highlighting the importance of assessing heterogeneity. These local average treatment effects are similar when we use pure controls instead of within-village controls as the comparison group (Table C.7).

Table C.5: Impact of the training on immediate outcomes within treated villages

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.142	0.0895	0.120	0.131	0.201**
	(0.111)	(0.0784)	(0.112)	(0.108)	(0.0801)
Treatment with Peer	0.0565	0.0836	-0.0342	-0.0904	0.0684
	(0.111)	(0.0792)	(0.116)	(0.115)	(0.0799)
Treatment with Peer + Connections Module	0.151	0.0519	0.0772	0.0805	0.0954
	(0.112)	(0.0859)	(0.122)	(0.113)	(0.0801)
Constant	0.647***	-0.229***	0.133	-0.0294	-0.0961*
	(0.0879)	(0.0518)	(0.0829)	(0.0826)	(0.0567)
Observations	710	710	707	688	693
R-squared	0.004	0.002	0.003	0.007	0.010

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the control sample within treated villages as the base category. Robust standard errors in parentheses.

Table C.6: Impact of the training on immediate outcomes within treated villages (LATE)

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	5.502***	0.297	3.364***	3.025***	2.617***
	(1.195)	(0.331)	(0.813)	(0.791)	(0.647)
Trained with Peer	5.828***	0.342	3.645***	3.135***	2.728***
	(1.279)	(0.354)	(0.866)	(0.850)	(0.690)
Constant	-3.563***	-0.417	-2.490***	-2.315***	-2.015***
	(0.949)	(0.258)	(0.636)	(0.626)	(0.510)
Observations	710	710	707	688	693
R-squared	-6.037	-0.018	-1.944	-1.718	-2.239
Under-ID p(KP LM)	2.87e-06	2.87e-06	3.23e-06	5.49e-06	6.89e-06
First-stage F: Trained Alone	142.3	142.3	138.6	128.1	127.1
First-stage F: Trained with peer	131	131	128.4	115.9	118.5
p: Trained Alone = Trained with peer	0.388	0.710	0.313	0.662	0.581

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the control sample within treated villages as the base category. Local average treatment effects (LATE) are estimated by instrumenting actual treatment status with assigned treatment. Robust standard errors in parentheses.

Table C.7: Impact of the training on immediate outcomes (LATE)

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	1.450***	-0.138	0.539***	0.403**	0.379***
	(0.328)	(0.168)	(0.196)	(0.170)	(0.146)
Trained with Peer	1.354***	-0.190*	0.501***	0.270*	0.262*
	(0.331)	(0.113)	(0.169)	(0.145)	(0.135)
Constant	-0.435	0.0131	-0.228	-0.212	-0.191
	(0.325)	(0.0985)	(0.148)	(0.133)	(0.119)
Observations	768	768	765	746	751
R-squared	0.040	0.006	-0.013	-0.017	-0.020
Under-ID p(KP LM)	1.65e-05	1.65e-05	1.28e-05	9.67e-06	9.20e-06
First-stage F: Trained Alone	181.9	181.9	182.7	171.2	174.9
First-stage F: Trained with peer	239.9	239.9	240.6	228	230.2
p: Trained Alone = Trained with peer	0.286	0.694	0.829	0.278	0.324

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats individuals in the pure control villages as the base category. Local average treatment effects (LATE) are estimated by instrumenting actual treatment status with assigned treatment. Standard errors are clustered at the village level.

D Follow-up Results

Table D.1: Effects on Long Term Outcomes by Type of Peer

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Spillover	0.291	0.594***	0.289	0.148
	(0.192)	(0.196)	(0.201)	(0.149)
Trained Alone	0.185	0.447***	0.161	-0.0258
	(0.135)	(0.149)	(0.182)	(0.145)
Close x More Central	0.416	-0.0326	0.00385	0.278
	(0.516)	(0.188)	(0.405)	(0.242)
Other Pair Types	0.0548	0.167	0.0285	-0.00822
	(0.127)	(0.155)	(0.171)	(0.117)
Degree	0.0165	0.0186	0.0129	0.00782
	(0.0211)	(0.0246)	(0.0184)	(0.0238)
Constant	-0.0888	-0.0964	-0.0696	-0.0472
	(0.132)	(0.142)	(0.185)	(0.156)
Observations	566	570	570	570
R-squared	0.011	0.031	0.012	0.006
p: Central Friend vs Other Pairs	0.523	0.328	0.951	0.199
p: Central Friend vs T1	0.676	0.0439	0.708	0.189

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats individuals in the pure control villages as the base category and includes an indicator for the spillover group and those who were intended to be paired but remained unmatched. We additionally control for the individual's own degree centrality in regressions that include relative comparisons between own and peer centrality. Standard errors are robust and clustered at the village level.

Table D.2: Effects on Long Term Outcomes by Type of Link (All)

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Spillover	0.288	0.591***	0.288	0.148
	(0.193)	(0.196)	(0.201)	(0.149)
Trained Alone	0.191	0.453***	0.164	-0.0235
	(0.135)	(0.150)	(0.181)	(0.146)
Close x More Central	0.416	-0.0333	0.00358	0.278
	(0.517)	(0.191)	(0.405)	(0.243)
Close x Less Central	0.369	0.289	0.248	0.114
	(0.227)	(0.297)	(0.258)	(0.175)
Far x Central	-0.127	-0.175	-0.0268	-0.0904
	(0.165)	(0.147)	(0.193)	(0.161)
Far x Non Central	0.0438	0.348*	-0.0279	-0.00441
	(0.148)	(0.200)	(0.202)	(0.132)
Degree	0.00667	0.00674	0.00821	0.00367
	(0.0218)	(0.0256)	(0.0182)	(0.0251)
Constant	-0.0405	-0.0381	-0.0465	-0.0268
	(0.136)	(0.148)	(0.185)	(0.162)
Observations	566	570	570	570
R-squared	0.016	0.039	0.014	0.007
p: Socially Close vs Distant	0.115	0.811	0.565	0.165
p: Less Central vs More Central	0.853	0.0463	0.606	0.803
p: Central and Close vs Less Central and Close	0.942	0.324	0.552	0.508
p: Central and Close vs Central and Distant	0.347	0.471	0.942	0.103
p: Central and Close vs Less Central and Distant	0.500	0.122	0.943	0.272

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats individuals in the pure control villages as the base category and includes an indicator for the spillover group and those who were intended to be paired but remained unmatched. We additionally control for the individual's own degree centrality in regressions that include relative comparisons between own and peer centrality. Standard errors are robust and clustered at the village level.

Table D.3: Effects on Long Term Outcomes (PDS-Lasso)

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Spillover	0.185	0.604***	0.270	0.150
	(0.189)	(0.189)	(0.190)	(0.146)
Trained Alone	0.166	0.500***	0.177	0.00308
	(0.130)	(0.159)	(0.170)	(0.142)
Trained with Peer	0.0654	0.204*	0.0167	-0.00960
	(0.108)	(0.121)	(0.158)	(0.116)
Constant	-0.0347	0.743***	0.413	-0.0118
	(0.0751)	(0.164)	(0.278)	(0.0888)
Observations	575	579	579	579
Number of groups	0	0	0	0
p: Trained Alone v/s Spillover	0.920	0.585	0.455	0.318
p: Trained Alone v/s Trained with Peer	0.360	0.0489	0.0457	0.905
p: Trained with Peer v/s Spillover	0.463	0.0217	0.0343	0.191

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats individuals in the pure control group as the base category and includes an indicator for the spillover group. Standard errors are robust and clustered at the village level. We implement PDS Lasso (Belloni et al. 2014) including baseline controls for income (level and sources), age, education (elementary, secondary, higher, university, informal), and caste (categories 1–6).

Table D.4: Impact of the training on follow-up outcomes (LATE)

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Trained Alone	0.102	0.521**	0.243	0.00955
	(0.172)	(0.240)	(0.225)	(0.178)
Trained with Peer	0.121	0.143	0.0715	0.00566
	(0.128)	(0.214)	(0.181)	(0.154)
Constant	0.0159	0.0241	-0.0451	-0.0186
	(0.0807)	(0.116)	(0.164)	(0.103)
Observations	472	475	475	475
R-squared	0.002	0.019	0.004	-0.000
Under-ID p(KP LM)	1.18e-05	1.17e-05	1.17e-05	1.17e-05
First-stage F: Trained Alone	99.89	101.7	101.7	101.7
First-stage F: Trained with peer	149.8	149.6	149.6	149.6
p: Trained Alone = Trained with peer	0.890	0.160	0.340	0.979

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the pure control group as the base category and excludes within village controls. Robust standard errors in parentheses.

E Mechanisms

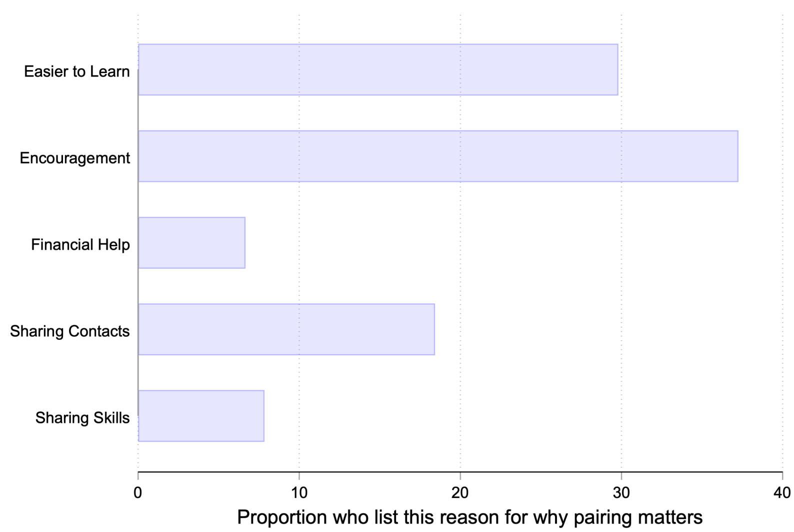


Figure E.1: The figure shows participant responses when asked to select one of the options explaining why pairing was beneficial for them.

Table E.1: Impact of being trained with a Central Friend after controlling for Similarity

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.773***	-0.152	0.228*	0.0765	0.0636
	(0.248)	(0.127)	(0.114)	(0.0844)	(0.0910)
Friend x More Central	0.987***	0.241	0.659**	0.190	0.346*
	(0.321)	(0.210)	(0.319)	(0.257)	(0.176)
Other Pair Types	0.898***	-0.260	0.485**	0.185	0.211
	(0.288)	(0.176)	(0.230)	(0.185)	(0.168)
Degree	0.0207	0.0192	0.0409**	0.0297*	0.0247*
	(0.0134)	(0.0193)	(0.0177)	(0.0168)	(0.0136)
Similarity Index	0.0226	0.0378	-0.504	-0.251	-0.310
	(0.289)	(0.258)	(0.424)	(0.292)	(0.300)
Constant	-0.0986	-0.0974	-0.194	-0.134	-0.112
	(0.242)	(0.103)	(0.128)	(0.111)	(0.0808)
Observations	756	756	753	734	740
R-squared	0.167	0.021	0.030	0.023	0.024
p: Central Friend vs Other Pairs	0.618	0.0113	0.438	0.982	0.0835
p: Central Friends vs T1	0.382	0.0745	0.173	0.649	0.127

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the pure control group as the base category. We additionally control for the "Similarity Index" which is a weighted index of similarity along characteristics including income, age, caste, marital status, and education and also control for degree-centrality of the individual.

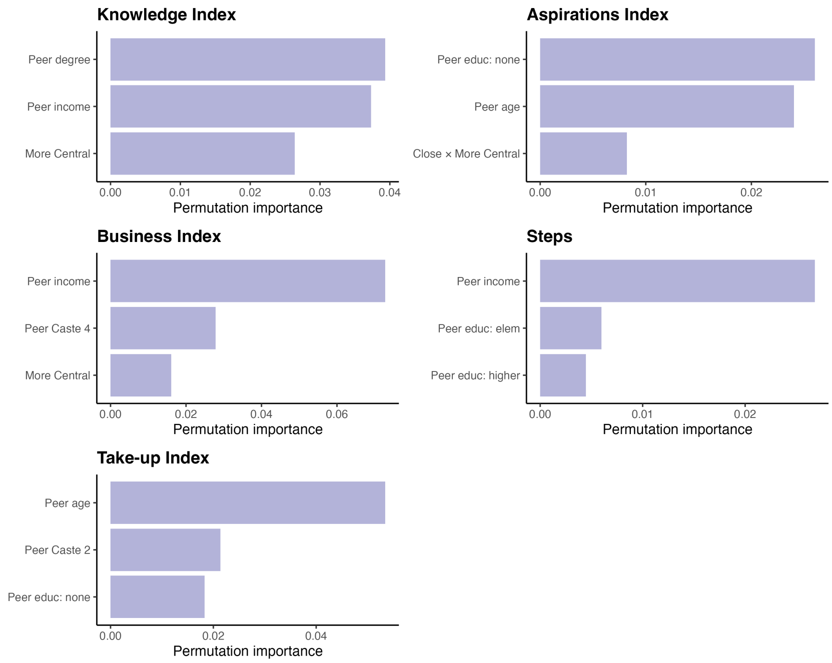
Table E.2: Effect of being matched with a Central Friend after controlling for Other Characteristics

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Friend x More Central	1.227***	0.313	0.370	-0.194	0.00929
	(0.420)	(0.511)	(0.476)	(0.535)	(0.317)
Other Pair Types	1.168***	-0.161	0.200	-0.224	-0.112
	(0.385)	(0.439)	(0.470)	(0.432)	(0.304)
Constant	-0.100	-0.0971	-0.193	-0.139	-0.110
	(0.244)	(0.105)	(0.130)	(0.111)	(0.0826)
Observations	755	755	752	733	739
R-squared	0.175	0.033	0.036	0.035	0.029
p: Central Friend vs Other Pairs	0.738	0.00889	0.467	0.900	0.0930
p: Central Friends vs T1	0.218	0.370	0.774	0.611	0.868

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the pure control group as the base category. We additionally control for peer characteristics such as peer age, caste, whether they are married, baseline income, and education. Standard errors are robust and clustered at the village level.

(A) Endline



(B) Follow-up

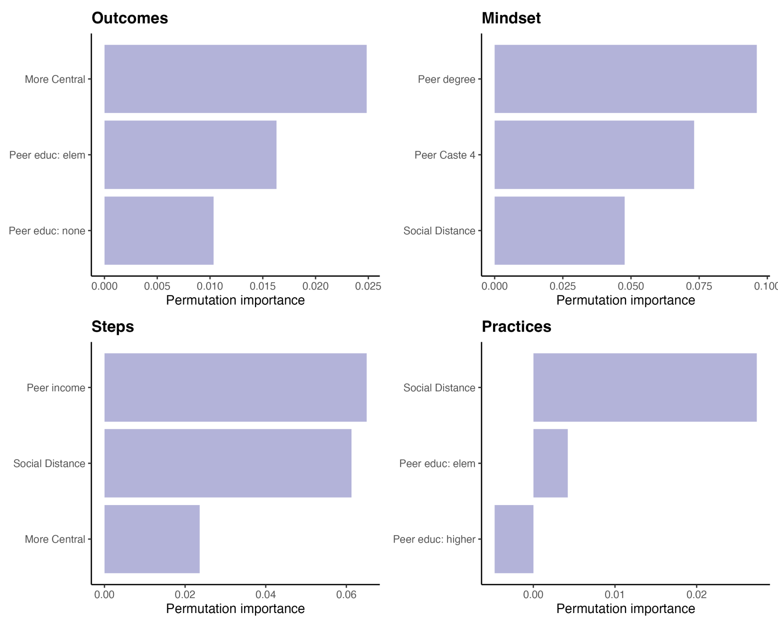


Figure E.2: Notes: The bars show the top three predictors of each outcome at endline (top) and follow-up (bottom), estimated using random forest models. Variable importance is measured by the increase in root mean squared error (RMSE) when the predictor is permuted. Outcomes are indices pre-specified in the analysis plan.

Table E.3: Effect on Learning during and after the training.

VARIABLES	(1) Knowledge Index	(2) Profit (Game)	(3) Profit (Business Plan)
Trained alone	0.789***		
	(0.247)		
Trained with Peer	0.748***	52.51	203,402
	(0.244)	(41.13)	(263,911)
Constant	0	273.3***	534,159***
	(0.234)	(34.06)	(168,030)
Observations	768	437	436
R-squared	0.122	0.004	0.001
Treatment 1==Pair	0.466		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the pure control group as the base category in Column 1 and Treatment 1 as the base category in Column 2. Standard errors are robust and clustered at the village level.

Table E.4: Learning by whether pairs are friends or not

VARIABLES	(1) Knowledge Index	(2) Profit (Game)	(3) Profit (Business Plan)
Trained alone	0.789***		
	(0.248)		
Paired (Close)	0.946***	55.71	1.011e+06*
	(0.249)	(58.02)	(566,051)
Paired (Far)	0.934***	48.64	-111,855
	(0.246)	(44.11)	(254,422)
Constant	-0	273.3***	534,159***
	(0.235)	(34.14)	(168,425)
Observations	763	432	432
R-squared	0.165	0.005	0.023
p: Close vs Far	0.894	0.897	0.0508
p: Close vs T1	0.213	0.213	0.213

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Column 1 uses the pure control group as the base category, while Columns 2 and 3 use those treated alone as the base category. Standard errors are robust; in Column 1 they are clustered at the village level.

Table E.5: Learning by type of pair

VARIABLES	(1) Knowledge Index	(2) Profit (Game)	(3) Profit (Business Plan)
Trained alone	0.773***		
	(0.248)		
Close x More Central	0.999***	46.14	1.712e+06
	(0.288)	(81.27)	(1.125e+06)
Other Pair Types	0.911***	54.30	21,784
	(0.244)	(43.04)	(247,688)
Degree	0.0207	12.84	55,043
	(0.0134)	(8.028)	(53,121)
Constant	-0.0985	205.2***	238,772
	(0.242)	(53.61)	(351,553)
Observations	756	427	427
R-squared	0.167	0.011	0.028
p: Central and Socially Close Peer vs Other Pairs	0.620	0.917	0.134
p: Central and Socially Close Peer vs T1	0.213		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The regression in Column 1 treats the pure control group as the base category while regressions in Columns 2 and 3 treat Treatment 1 as the base category since the games were only conducted for trainees. Standard errors are robust and clustered at the village level.

Table E.6: Correlation of Number of Contacts Pooled and Similarity with Peer

VARIABLES	(1) Number of contacts pooled	(2) Number of contacts pooled	(3) Number of contacts pooled	(4) Number of contacts pooled
Same age group		-0.202		-0.548
		(0.738)		(0.670)
Same caste		1.299**		1.476***
		(0.526)		(0.438)
Same education		0.370		0.654
		(0.503)		(0.480)
Same income group		-0.745		-0.363
		(0.513)		(0.615)
Quadratic diff. in connections	0.0120	0.0116	0.0142	0.0144
	(0.00981)	(0.00959)	(0.00857)	(0.00881)
Network Distance	0.113	0.0943	0.209**	0.0516
	(0.134)	(0.127)	(0.0864)	(0.112)
Constant	5.850***	5.052***	5.389***	5.000***
	(0.547)	(0.639)	(0.326)	(0.301)
Number of Dyads	68	68	92	92
R-squared	0.030	0.147	0.070	0.198

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression compares dyads in Treatment Arm 3. Columns 1 and 2 only include pairs assigned to Treatment 3 and attended it while Columns 3 and 4 include pairs who attended the connections module, regardless of assignment. Robust standard errors in parentheses.

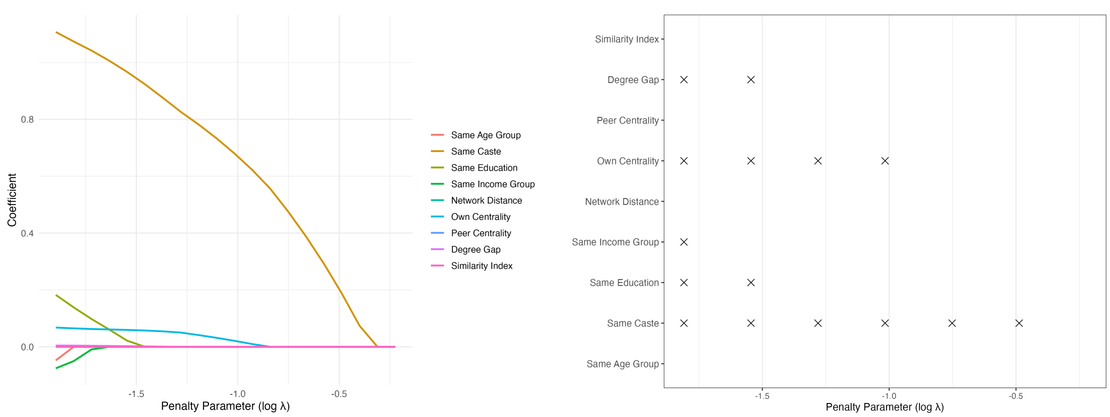
Table E.7: Perceived Future Interactions by Number of Contacts Shared

VARIABLES	(1) Pairs will start Business Together	(2) Pairs will meet in Future
Number of links pooled in T3	0.0526**	-0.00918
	(0.0229)	(0.0175)
Paired (Socially Close)	0.269**	0.0992
	(0.106)	(0.0656)
Paired (High Centrality)	-0.0252	-0.0635
	(0.105)	(0.0847)
Constant	-0.0309	0.905***
	(0.157)	(0.110)
Observations	90	90
R-squared	0.097	0.031

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: All columns include individuals who were assigned to Treatment 3 and attended it. Robust standard errors in parentheses.

Figure E.3: LASSO predictors of number of pooled network links



Notes: The left panel shows LASSO coefficient paths for predicting the number of links pooled within each pair who attended Treatment 3, regardless of assignment; the right panel shows when each variable enters the model as the penalty parameter (λ) decreases.

Table E.8: Relationship Between Caste Homophily and Number of Contacts Pooled

VARIABLES	(1) Number of contacts pooled	(2) Number of contacts pooled	(3) Number of contacts pooled	(4) Number of contacts pooled
Homophily	-0.267***	-0.0962*	-0.255***	-0.0517
	(0.0810)	(0.0573)	(0.0912)	(0.0840)
Same caste		1.557***		2.092***
		(0.433)		(0.673)
Homophily x Same Caste		-1.729*		-4.785***
		(0.906)		(1.100)
Constant	6.644***	5.523***	6.613***	5.053***
	(0.196)	(0.353)	(0.222)	(0.617)
Observations	123	121	96	94
R-squared	0.047	0.155	0.044	0.212

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes:Notes: Each column reports OLS estimates of the relationship between caste homophily and the number of unique contacts pooled by two peers. Homophily is defined as $\frac{\text{share of same-caste links} - \text{share of own caste in the population}}{\text{share of own caste in the population}}$. Columns (1) and (2) includes all participants who attended the connections module (T3) sessions, while columns (3) and (4) restricts the sample to those randomly assigned to T3. Robust standard errors are in parentheses.

Table E.9: Heterogeneous Treatment Effects by Number of Contacts Shared

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.789***	-0.139	0.253**	0.101	0.105
	(0.248)	(0.127)	(0.116)	(0.0879)	(0.0886)
Treatment with Peer	0.886***	-0.121	0.208	0.0146	0.0811
	(0.251)	(0.105)	(0.123)	(0.105)	(0.100)
Treatment with Peer + Connections Module	0.994***	0.0119	0.437***	0.367*	0.265
	(0.289)	(0.167)	(0.155)	(0.193)	(0.156)
T3 X Number of contacts pooled	-0.0746	-0.400	-0.287	-0.436*	-0.319
	(0.297)	(0.300)	(0.296)	(0.250)	(0.272)
Constant	-0	0	0	0	-0
	(0.235)	(0.0934)	(0.0794)	(0.0617)	(0.0621)
Observations	743	743	740	723	727
R-squared	0.158	0.009	0.020	0.024	0.020

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the pure control group as the base category. Standard errors are robust and clustered at the village level.

Table E.10: Effect of Peer Risk Aversion

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.596***	-0.0324	0.207	0.00662	0.105
	(0.135)	(0.170)	(0.194)	(0.101)	(0.145)
Trained with Peer	0.658***	-0.0263	0.107	-0.118	0.111
	(0.133)	(0.147)	(0.169)	(0.113)	(0.118)
Risk Averse (Binary)	-0.0899	0.0280	-0.177	-0.109	-0.127
	(0.162)	(0.184)	(0.219)	(0.140)	(0.169)
T1 x Risk Averse	-0.0476	-0.119	-0.0522	-0.00635	0.0428
	(0.190)	(0.234)	(0.284)	(0.197)	(0.221)
T2/T3 x Risk Averse	0.105	-0.189	0.103	0.190	-0.0251
	(0.209)	(0.219)	(0.263)	(0.183)	(0.191)
Constant	0.256***	-0.0441	0.168	0.139**	0.0456
	(0.0812)	(0.118)	(0.126)	(0.0544)	(0.0788)
Observations	673	673	670	651	656
R-squared	0.110	0.008	0.019	0.019	0.027

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the pure control group as the base category. Standard errors are robust and clustered at the village level.

Table E.11: Treatment Effects on Network-Based Measures of Saving

VARIABLES	(1) Number of Savings Groups	(2) Joined Cooperative in last year
Spillover	0.0310	0.0786
	(0.312)	(0.0533)
Trained Alone	0.144	0.112**
	(0.308)	(0.0478)
Trained with Peer	0.136	0.00104
	(0.298)	(0.0385)
Constant	1.864***	0.0929***
	(0.283)	(0.0329)
Observations	579	580
R-squared	0.002	0.020

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats the pure control group as the base category. Standard errors are robust and clustered at the village level.

Table E.12: Determinants of reporting encouragement as a benefit of pairing

VARIABLES	(1) Encouragement	(2) Encouragement
Peer caste = 2	-0.161	-0.151
	(0.128)	(0.123)
Peer caste = 3	-0.00793	-5.77e-05
	(0.125)	(0.120)
Peer caste = 4	-0.300**	-0.274**
	(0.135)	(0.130)
Peer education = 1	-0.239***	-0.240***
	(0.0872)	(0.0878)
Peer education = 2	-0.109	-0.112
	(0.0850)	(0.0854)
Peer education = 3	-0.147	-0.135
	(0.121)	(0.121)
Peer income	-8.61e-07	-8.96e-07
	(8.25e-07)	(8.04e-07)
Peer marital status = 1	0.0188	0.0305
	(0.131)	(0.126)
Peer marital status = 3	0.333	0.285
	(0.288)	(0.279)
Peer age	0.00241	0.00254
	(0.00402)	(0.00401)
Own network degree	-1.08e-05	
	(0.0130)	
Peer is more central	0.134*	
	(0.0707)	
Peer network degree		0.0302**
		(0.0122)
Constant	0.411	0.282
	(0.256)	(0.250)
Observations	247	247
R-squared	0.112	0.117

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable is an indicator equal to 1 if the respondent reported encouragement as a reason why pairing was beneficial. The sample is restricted to individuals assigned to the paired treatment arm and were actually trained with a peer. Robust standard errors in parentheses.

Table E.13: Effects on Network Interactions during the year after training

VARIABLES	(1) Talk to anyone about business	(2) Talk to anyone about business	(3) Talk to anyone about business
Spillover	0.0723**	0.0730**	
	(0.0290)	(0.0297)	
Trained Alone	0.0513	0.0494*	
	(0.0328)	(0.0253)	
Close x More Central		0.0911	
		(0.0718)	
Close x Less Central		0.155**	
		(0.0746)	
Far x Central		0.0636	
		(0.0424)	
Far x Less Central		0.0614*	
		(0.0372)	
Degree		0.00423	
		(0.00523)	
Trained with Peer	0.0655**		
	(0.0277)		
Peer Centrality			-0.00752
			(0.00873)
Social Distance			-0.0353***
			(0.0132)
Constant	0.0143	-0.00634	0.261***
	(0.0123)	(0.0264)	(0.0917)
Observations	579	569	201
R-squared	0.013	0.027	0.045
Paired=Nonpaired	0.689		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

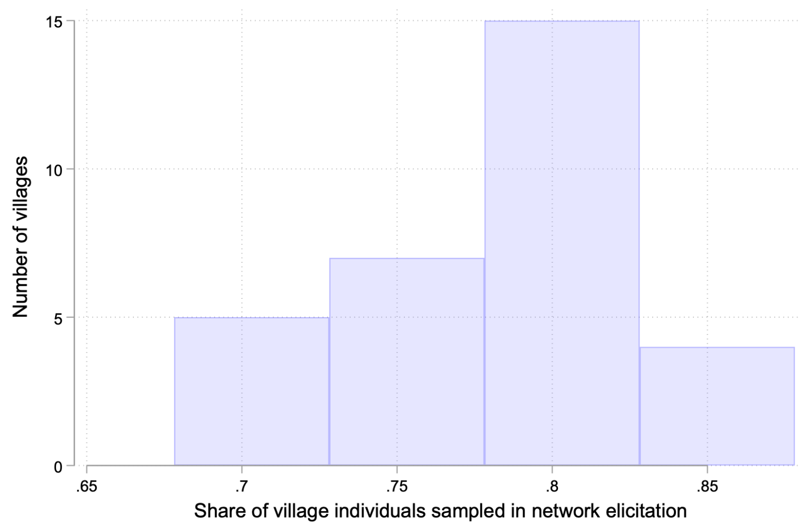
Notes: Column 1 treats the pure control group as the base category while Column 2 treats those paired without the connections module as the base category. Both the variables are self-reported and show whether an individual talked to anyone or the matched peer about businesses during the one year after the intervention. Standard errors are robust and clustered at the village level.

F Network Sampling and Robustness Checks

Figure F.1 shows the proportion of village populations for which we elicited network data. On average, this rate is 78%. We then correct our measure of degree centrality to account for unobserved links by dividing the observed degree by the village-specific sampling proportion. The intuition is as follows: if individual i has d_i links in an observed network with x nodes, then in a village with population y they are predicted to have $\frac{d_i}{x} \times y$ links. Figure F.2 shows the distribution of village densities based on this corrected measure. Density is higher than the earlier conservative estimate but remains low overall, indicating sparse networks.

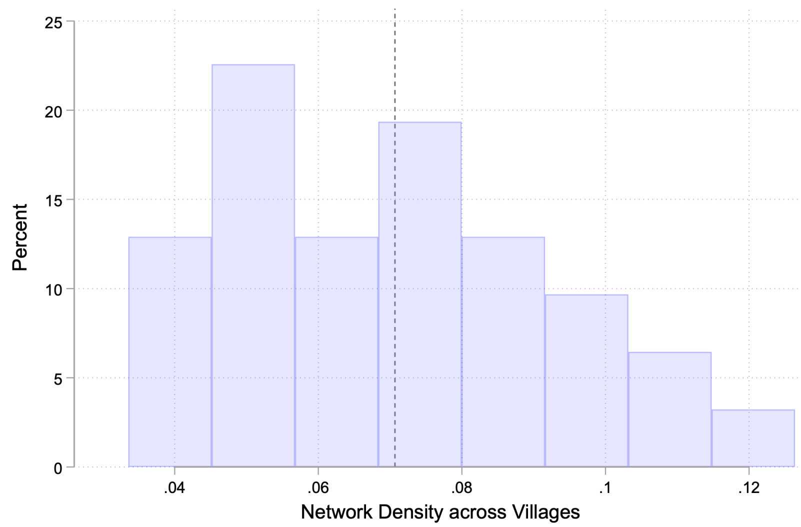
Next, we present endline and follow-up results using this corrected measure of centrality in Table F.1 and Table F.2. The results show that the positive effects of close and central peers on aspirations at endline, and of close and less central peers at follow-up, are robust to this correction.

Figure F.1: Fraction of Individuals per Village Sampled in Network Elicitation



Notes: The figure plots the distribution of the share of each village's population for which network data were elicited. Village population is defined as the union of all individuals who had network data collected or were listed as contacts by others but do not have network data. The sampling rate for each village is the ratio of individuals with network data to this constructed village population.

Figure F.2: Network Density across Villages



Notes: The figure shows the distribution of village-level network density computed using the degree measure corrected for sampling coverage. For each village, we first adjust individual degrees accounting for individuals who are not included in the network elicitation, and then calculate density based on these corrected degrees.

Table F.1: Effects on Endline Outcomes by Peer Type

(a) By Social Distance and Peer Centrality

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.789***	-0.139	0.253**	0.101	0.105
	(0.248)	(0.127)	(0.116)	(0.0879)	(0.0886)
Trained with Peer	1.195***	0.208	0.321	0.0322	-0.172
	(0.281)	(0.172)	(0.252)	(0.153)	(0.196)
Trained with Peer × Peer distance	-0.0304	-0.0811***	-0.0227	0.00355	0.0220
	(0.0263)	(0.0219)	(0.0450)	(0.0301)	(0.0214)
Trained with Peer × Peer degree	-0.0227	-0.0147	0.00164	0.00554	0.0276
	(0.0150)	(0.0134)	(0.0212)	(0.0124)	(0.0174)
Constant	0	0	0	-0	-0
	(0.235)	(0.0933)	(0.0794)	(0.0617)	(0.0621)
Observations	753	753	750	731	737
R-squared	0.166	0.014	0.019	0.019	0.021

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) By Social Closeness and More/Less Central

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.772***	-0.154	0.220*	0.0694	0.0538
	(0.248)	(0.127)	(0.114)	(0.0847)	(0.0924)
Trained with Peer	0.952***	-0.301**	0.212	0.0375	-0.00403
	(0.247)	(0.125)	(0.130)	(0.103)	(0.0972)
Trained with Peer (Close)	-0.00390	0.278**	0.0348	-0.0461	-0.0141
	(0.0975)	(0.113)	(0.172)	(0.130)	(0.0670)
Trained with Peer (More Central)	-0.0658	0.0974	-0.0115	0.0446	0.142*
	(0.105)	(0.115)	(0.139)	(0.150)	(0.0704)
Degree (Alternative)	0.0125	0.0132	0.0325**	0.0248*	0.0252**
	(0.0112)	(0.0163)	(0.0152)	(0.0138)	(0.0117)
Constant	-0.0732	-0.0831	-0.190	-0.138	-0.143
	(0.239)	(0.108)	(0.132)	(0.113)	(0.0848)
Observations	756	756	753	734	740
R-squared	0.167	0.017	0.027	0.023	0.024

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(c) By Social Closeness Interacted with More/Less Central

VARIABLES	(1) Knowledge Index	(2) Aspirations Index	(3) Business Index	(4) Additional Steps	(5) Take-up Index
Trained alone	0.770***	-0.157	0.218*	0.0709	0.0565
	(0.248)	(0.127)	(0.114)	(0.0835)	(0.0916)
Close x More Central	0.995***	0.257	0.379*	0.0484	0.173
	(0.288)	(0.200)	(0.200)	(0.212)	(0.108)
Other Pair Types	0.909***	-0.245**	0.189*	0.0384	0.0270
	(0.244)	(0.113)	(0.110)	(0.0869)	(0.0941)
Degree (Alternative)	0.0147	0.0171	0.0344**	0.0231*	0.0220*
	(0.0100)	(0.0149)	(0.0138)	(0.0131)	(0.0107)
Constant	-0.0863	-0.106	-0.201	-0.128	-0.124
	(0.238)	(0.104)	(0.125)	(0.108)	(0.0808)
Observations	756	756	753	734	740
R-squared	0.167	0.022	0.028	0.023	0.023
p: Central and Close vs Other Pairs	0.627	0.0104	0.412	0.967	0.0641
p: Central and Close vs T1	0.217	0.0388	0.424	0.909	0.260

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The above regressions treat individuals in the pure control villages as the base category and include an indicator for those who were intended to be paired but remained unmatched. We additionally control for the individual's own degree centrality in regressions that include relative comparisons between own and peer centrality. Standard errors are robust and clustered at the village level.

Table F.2: Effects on Long-Term Outcomes

(a) By Social Distance and Peer Centrality

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Trained Alone	0.194	0.490***	0.173	-0.0201
	(0.137)	(0.164)	(0.188)	(0.146)
Trained with Peer	0.565*	0.427	0.511	0.326
	(0.324)	(0.387)	(0.342)	(0.222)
Trained with Peer × Peer distance	-0.0994**	0.00975	-0.102*	-0.101**
	(0.0466)	(0.0707)	(0.0524)	(0.0454)
Trained with Peer × Peer degree	-0.0163	-0.0489*	-0.0203	0.00598
	(0.0230)	(0.0239)	(0.0232)	(0.0183)
Constant	0	-0	-0	-0
	(0.0702)	(0.0723)	(0.158)	(0.0860)
Observations	564	568	568	568
R-squared	0.013	0.035	0.018	0.012

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) By Social Closeness and More/Less Central

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Trained Alone	0.196	0.455***	0.165	-0.0224
	(0.135)	(0.149)	(0.180)	(0.145)
Trained with Peer	0.0245	0.326	0.00508	-0.0332
	(0.174)	(0.195)	(0.204)	(0.126)
Trained with Peer (Close)	0.424*	0.0252	0.182	0.221
	(0.218)	(0.183)	(0.250)	(0.173)
Trained with Peer (More Central)	-0.121	-0.467**	-0.0837	-0.0129
	(0.229)	(0.182)	(0.196)	(0.136)
Degree (Alternative)	-0.00164	0.00244	0.00337	0.000900
	(0.0178)	(0.0204)	(0.0157)	(0.0199)
Constant	0.00216	-0.0198	-0.0266	-0.0143
	(0.137)	(0.147)	(0.185)	(0.159)
Observations	566	570	570	570
R-squared	0.015	0.038	0.013	0.006

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(c) By Social Closeness Interacted with More/Less Central

VARIABLES	(1) Outcomes Index	(2) Steps Index	(3) Mindset Index	(4) Business Practices
Spillover	0.287	0.591***	0.286	0.147
	(0.192)	(0.197)	(0.201)	(0.148)
Trained Alone	0.193	0.447***	0.165	-0.0237
	(0.135)	(0.149)	(0.180)	(0.145)
Close x Less Central	0.380*	0.270	0.256	0.114
	(0.223)	(0.294)	(0.259)	(0.176)
Other Pair Types	0.0442	0.113	-0.0232	0.0100
	(0.135)	(0.141)	(0.168)	(0.126)
Degree (Alternative)	0.00116	0.0106	0.00324	0.00233
	(0.0174)	(0.0185)	(0.0165)	(0.0191)
Constant	-0.0148	-0.0692	-0.0258	-0.0229
	(0.135)	(0.136)	(0.187)	(0.155)
Observations	566	570	570	570
R-squared	0.011	0.031	0.014	0.004
p: Central Friend vs Other Pairs	0.182	0.555	0.203	0.562
p: Central Friends vs T1	0.393	0.542	0.708	0.393

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This regression treats individuals in the pure control villages as the base category and includes an indicator for the spillover group and those who were intended to be paired but remained unmatched. We additionally control for the individual's own degree centrality in regressions that include relative comparisons between own and peer centrality. Standard errors are robust and clustered at the village level.

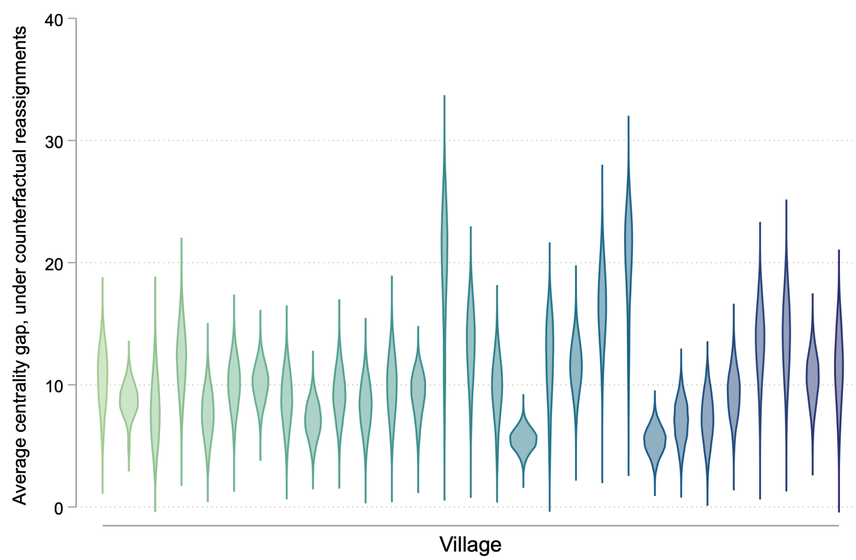
G Peer Effects and Feasibility of Peer Pairings

Table G.1: Identification of Peer Effects (Robustness)

VARIABLES	(1) Outcome (Follow-up)	(2) Outcome (Follow-up)	(3) Outcome (Follow-up)
Peer Outcome (Endline)	0.276*	0.343*	0.381*
	(0.163)	(0.183)	(0.210)
Peer Outcome (Follow-up)	0.141*	0.151*	-0.0829
	(0.0762)	(0.0776)	(0.0776)
Close X Peer Outcome	0.0238	-0.0467	-0.117
	(0.188)	(0.188)	(0.200)
Peer Degree X Peer Outcome	-0.0646**	-0.0581*	-0.0594*
	(0.0306)	(0.0321)	(0.0347)
Own Outcome (Endline)	0.295***	0.101	0.107
	(0.0757)	(0.0817)	(0.100)
Constant	-0.0843	0.736	0.539
	(0.0658)	(0.745)	(0.842)
Observations	186	186	186
R-squared	0.083	0.192	0.360

Notes: The table reports the effect of the matched peer's endline outcome on the individual's own follow-up outcome, additionally controlling for the peer's own outcome in the follow up. Close × Peer Outcome interacts the peer outcome with an indicator for whether the peer was socially close. Peer Degree × Peer Outcome interacts the peer outcome with the peer's degree centrality. Columns (2) and (3) control for the individual's and peer's age, income, education, caste, and network degree. Column (3) additionally includes village fixed effects.

Figure G.1: Feasibility of Counterfactual Pair Reassignments



Notes: The figure plots the distribution of the *average within-village degree-centrality gap* across 10,000 random reassignments of peer dyads (without replacement), holding the network structure fixed.

References

Anderson, M. L. (2008), 'Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects', *Journal of the American statistical Association* **103**(484), 1481–1495.

Asiedu, E., Lambon-Quayefio, M., Truffa, F. & Wong, A. (2023), 'Female entrepreneurship and professional networks', *PEDL research paper* .

Banerjee, A., Breza, E., Chandrasekhar, A. G., Duflo, E., Jackson, M. O. & Kinnan, C. (2024), 'Changes in social network structure in response to exposure to formal credit markets', *Review of Economic Studies* **91**(3), 1331–1372.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. (2013), 'The diffusion of microfinance', *Science* **341**(6144), 1236498.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. (2019), 'Using gossips to spread information: Theory and evidence from two randomized controlled trials', *The Review of Economic Studies* **86**(6), 2453–2490.

Beaman, L., BenYishay, A., Magruder, J. & Mobarak, A. M. (2021), 'Can network theory-based targeting increase technology adoption?', *American Economic Review* **111**(6), 1918–1943.

Belloni, A., Chernozhukov, V. & Hansen, C. (2014), 'Inference on treatment effects after selection among high-dimensional controls', *Review of Economic Studies* **81**(2), 608–650.

Bernard, T. & Seyoum Taffesse, A. (2014), 'Aspirations: An approach to measurement with validation using ethiopian data', *Journal of African economies* **23**(2), 189–224.

Bhattarai, S. (2006), The bola or parma of the newar in manamaiju village. the significance of a farm labor exchange system among indigenous peasants in nepal., Master's thesis, Universitetet i Tromsø.

Boucher, V., Rendall, M., Ushchev, P. & Zenou, Y. (2024), 'Toward a general theory of peer effects', *Econometrica* **92**(2), 543–565.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2020), 'Peer effects in networks: A survey', *Annual Review of Economics* **12**(1), 603–629.

Breiman, L. (2001), 'Random forests', *Machine learning* **45**(1), 5–32.

Breza, E. & Chandrasekhar, A. G. (2019), 'Social networks, reputation, and commitment: evidence from a savings monitors experiment', *Econometrica* **87**(1), 175–216.

Brooks, W., Donovan, K. & Johnson, T. R. (2018), 'Mentors or teachers? microenterprise training in kenya', *American Economic Journal: Applied Economics* **10**(4), 196–221.

Cai, J., Janvry, A. D. & Sadoulet, E. (2015), 'Social networks and the decision to insure', *American Economic Journal: Applied Economics* **7**(2), 81–108.

Cai, J. & Szeidl, A. (2018), 'Interfirm relationships and business performance', *The Quarterly Journal of Economics* **133**(3), 1229–1282.

Calvó-Armengol, A., Patacchini, E. & Zenou, Y. (2009), 'Peer effects and social networks in education', *The review of economic studies* **76**(4), 1239–1267.

Campos, F. M. L., Coleman, R. D., Conconi, A., Donald, A. A., Gassier, M., Goldstein, M. P., Chavez, Z. L., Mikulski, J., Milazzo, A., Paryavi, M., Pierotti, R. S., O'Sullivan, M. B. & Vaillant, J. (2019),

'Profiting from parity: Unlocking the potential of women's businesses in africa : Main report'.
URL: *http://documents.worldbank.org/curated/en/501971553025918098/Main-Report*

Carranza, E., Dhakal, C. & Love, I. (2018), 'Female entrepreneurs: How and why are they different?'.

Chandrasekhar, A. G., Kinnan, C. & Larreguy, H. (2018), 'Social networks as contract enforcement: Evidence from a lab experiment in the field', *American Economic Journal: Applied Economics* **10**(4), 43–78.

De Mel, S., McKenzie, D. & Woodruff, C. (2014), 'Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from sri lanka', *Journal of Development Economics* **106**, 199–210.

Duflo, E., Dupas, P. & Kremer, M. (2011), 'Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya', *American economic review* **101**(5), 1739–1774.

Durlauf, S. N. & Fafchamps, M. (2005), Social capital, *in* P. Aghion & S. N. Durlauf, eds, 'Handbook of Economic Growth', Vol. 1B, Elsevier, Amsterdam, pp. 1639–1699.

Fafchamps, M. & Quinn, S. (2018), 'Networks and manufacturing firms in africa: Results from a randomized field experiment', *The World Bank Economic Review* **32**(3), 656–675.

Field, E., Jayachandran, S. & Pande, R. (2010), 'Do traditional institutions constrain female entrepreneurship? a field experiment on business training in india', *American Economic Review* **100**(2), 125–129.

Field, E., Jayachandran, S., Pande, R. & Rigol, N. (2016), 'Friendship at work: Can peer effects catalyze female entrepreneurship?', *American Economic Journal: Economic Policy* **8**(2), 125–153.

Glaeser, E. L., Laibson, D. & Sacerdote, B. (2002), 'An economic approach to social capital', *The economic journal* **112**(483), F437–F458.

Granovetter, M. S. (1973), 'The strength of weak ties', *American journal of sociology* **78**(6), 1360–1380.

Lerner, J. & Malmendier, U. (2013), 'With a little help from my (random) friends: Success and failure in post-business school entrepreneurship', *The Review of Financial Studies* **26**(10), 2411–2452.

Manski, C. F. (1993), 'Identification of endogenous social effects: The reflection problem', *The review of economic studies* **60**(3), 531–542.

McKenzie, D., Woodruff, C., Bjorvatn, K., Bruhn, M., Cai, J., Gonzalez-Uribe, J., Quinn, S., Sonobe, T. & Valdivia, M. (2021), 'Training entrepreneurs', *VoxDevLit* **1**(2), 3.

Messerschmidt, D. A. (1981), '" nogar" and other traditional forms of cooperation in nepal: Significance for development', *Human organization* pp. 40–47.

Putnam, R. D. (2000), *Bowling alone: The collapse and revival of American community*, Simon and schuster.

Sacerdote, B. (2011), Peer effects in education: How might they work, how big are they and how much do we know thus far?, *in* 'Handbook of the Economics of Education', Vol. 3, Elsevier, pp. 249–277.

Sherpa, N. S. (2005), Indigenous peoples of nepal and traditional knowledge, *in* 'International workshop on Traditional Knowledge', pp. 21–23.

Vasilaky, K. N. & Leonard, K. L. (2018), 'As good as the networks they keep? improving outcomes through weak ties in rural uganda', *Economic Development and Cultural Change* **66**(4), 755–792.

Vega-Redondo, F., Pin, P., Ubfal, D., Benedetti-Fasil, C., Brummitt, C., Domínguez, M., Rubera, G., Hovy, D. & Fornaciari, T. (2023), 'Networking entrepreneurs', *Working Paper.* .

Zárate, R. A. (2023), 'Uncovering peer effects in social and academic skills', *American Economic Journal: Applied Economics* **15**(3), 35–79.