DOI:
https://doi.org/10.1016/j.jebo.2023.12.026

Document Version
Peer reviewed version

Citation for published version (APA):
Dhillon, A., Afridi, F., & Sharma, S. (Accepted/In press). The Ties That Bind Us: Social Networks and Productivity in the Factory. *Journal of Economic Behavior and Organization*, *218*, 470-485. https://doi.org/10.1016/j.jebo.2023.12.026

The Ties That Bind Us: Social Networks and Productivity in the Factory*

Farzana Afridi† Amrita Dhillon‡ Swati Sharma§

January 2, 2024

Abstract

We use high frequency worker level productivity data from garment manufacturing units in India to study the effects of caste-based social networks on individual and group productivity when workers are complements in the production function. Using plausibly exogenous variation in the production lines' caste composition for almost 35,000 worker-days, we find that a 1 percentage point (pp) increase in the share of own caste workers in the line increases daily individual productivity by at least 0.09 pp. The least efficient worker's productivity, however, rises by almost 0.17 pp when the caste composition of the line becomes more homogeneous by 1 pp. These results are robust to unobservable heterogeneity in worker ability and line level trends. Production externalities, that induce greater effort through within-network peer effects, can potentially explain our findings.

KEYWORDS: caste, social networks, labor productivity, assembly lines, garment factory
JEL CLASSIFICATION: J24, J41, J46, O12, D86

*This paper has benefited from suggestions by Ashwini Deshpande, James Fenske, Karla Hoff, Seema Jayachandran, Kaivan Munshi, Anant Nyshadham, Atonu Rabbani, Chris Woodruff. Nikhil Bharadwaj and Karmini Sharma provided exceptional research assistance. We are grateful to two anonymous referees for their feedback. The Policy and Planning Research Unit (PPRU) at ISI, Delhi (Afridi) and the Centre for Competitive Advantage in the Global Economy (CAGE), Warwick University (Dhillon) provided financial support for this study. The usual disclaimers apply.

†Indian Statistical Institute, Delhi and IZA (Bonn) (fafridi@isid.ac.in)
‡King's College, London and CAGE, University of Warwick (amrita.dhillon@kcl.ac.uk)
§Indian Institute of Technology Delhi (Swati.Sharma@hss.iitd.ac.in)

1 Introduction

While much of the literature on the manufacturing sector has focused on productivity differentials across firms (Bloom et al. (2013)), in several industries production processes are organised in teams, such as assembly lines. Team productivity often varies significantly not just across firms but also within the same manufacturing units.[1] In our setting of the labor intensive garment industry in India, average team productivity differs by over 60% between the least and most productive teams or lines in the same manufacturing plant. This variation in productivity across teams is accompanied by equally large variation across workers within a team, with the least productive worker being more than 95% less efficient than the most productive worker.

Research providing micro econometric evidence on determinants of worker productivity under team production is, however, scarce. A majority of the existing studies estimate individual worker performance under either individual piece rate payments (performance pay) or team based incentives when workers are substitutes in the production function. Productivity of workers in large assembly lines within firms and the associated coordination problems have not been explored as much. We attempt to fill this gap by analysing the role of workers' social networks in explaining the large variation in individual and team output across production lines within garment manufacturing units in India. With millions of workers worldwide (Chang et al. (2016), GOI (2018)), labor-intensive garment manufacturing is a natural choice for advancing our understanding of worker performance within firms.

Given the nature of the production function in assembly lines, where complementarities between workers generate externalities in the production process and the total output of the team is constrained by the least efficient workers, the worker composition of these teams can play a significant role in determining both group and firm

[1] A recently concluded European Commission (ERC (2015-21)) project on garment manufacturing finds significant dispersion of productivity within factories in a sample of 100 factories in Bangladesh - production lines at the 90th percentile are 50% more efficient than those at the 10th percentile.

output. We use workers' caste as a proxy for their residence-based social networks.[2] Utilizing high-frequency data that include detailed information on the daily productivity of individual workers, their production lines, and the caste composition of the workers' lines on each production day in the stitching department of two garment factories in the National Capital Region of Delhi, we follow 1744 workers over 31 work days, giving us information for 34,641 worker-days. Our estimation strategy relies on variation in the daily worker composition of production lines due to unanticipated worker absenteeism (and the accompanying reallocation of workers across lines). Our identifying assumption is that conditional on controlling for individual worker unobservables and line-level trends, the daily variation in line composition is exogenous. Under this assumption, our analysis estimates the causal impact of the proportion of own-caste workers (i.e. same social network) in a production line on individual and line productivity on a work day.

Our findings suggest that a 1 pp increase in the strength of the workers' social network - the proportion of workers belonging to own caste - in the line on a work day, raises workers' own productivity by at least 0.09 pp. We calculate the caste-concentration index of the line and aggregate the data to the line level to find that the least efficient worker's productivity rises by almost 0.17 pp while the average line performance improves by more than 0.26 pp when the caste composition of the line becomes more homogeneous by 1 pp. Our findings are robust to a host of sensitivity checks, including worker ability, line specific unobservables and line-level trends in production.

Given the absence of explicit group-based incentives, it is puzzling that individual productivity, and especially minimum productivity in the line, improves when teams are more socially connected. In our context, workers receive a fixed, monthly salary but their total earnings depend on their skill grade (with wage differential between grades of about 10-12%) and overtime wages (at higher than regular hourly wage

[2]See Munshi (2019), Afridi et al. (2015) for evidence on the importance of caste networks in labor markets in India, in contrast to social networks represented by friends or acquaintances.

rate). Workers who are more productive have a higher probability of being promoted to higher grades or receiving overtime due to recommendations by their line supervisor whose remuneration is tied to the line output. Thus there exist *implicit* individual financial incentives linked to higher team production. Since promotions are much more likely for high ability workers, they have strong incentives to monitor (or mentor) poorly performing co-workers and enforce higher effort from those who are holding up line output.[3]

If poor performance at work lowers earnings of co-workers in the line due to the production externality, workers are induced to put in greater effort when more of their co-workers in the line belong to their own-caste network through monitoring (or mentoring) and enforcement using network based rewards or punishments.[4]

We document a few pieces of evidence that are consistent with our leading explanation. First, our worker level data suggest strong socio-economic interdependence and benefits from one's networks, formed in caste-based residential neighborhoods, as sources of information for job openings as well as for referrals.[5] Second, we conjecture that social pressures to increase effort are higher the lower is the initial productivity of the worker, as these workers are most likely to be holding up line output and more likely to need network resources in the future. Indeed, the effect size of increase in own caste share on productivity of least efficient workers in the line is large. Third, we find that the higher productivity of the least efficient workers is driven by an increase in the proportion of own caste high experience workers in the line.

[3]In a highly competitive product market such as the garment industry, firms are constrained in offering employees explicit monetary incentives, nor can they punish low productivity employees due to minimum wage constraints and sufficient availability of outside job options (at minimum wages) (McKinsey (2016), Chang et al. (2016)).

[4]Greif (1993) show how informal sanctions enforced by the network of traders can help in reducing moral hazard.

[5]For instance, 75% of the workers obtained information on their current job through their social network while 64% of the informants were employed in the factory at the time of the job opening. Almost a third of these informants were still employed at the time of our survey (conditional on informal flow of information), the majority of whom were line level worker (62%) and/or neighbors (52%) who were known to the respondent for over 7 years. Not only did these social contacts provide information on job openings, 42% of them also referred the worker to the management for jobs. 77% of these workers also say that they would be able to borrow money from this informant in an emergency.

We do not find evidence supportive of other mechanisms, such as altruism. Workers may simply have pro-social preferences towards other network members, which can lead to higher effort by the lowest ability workers leading to an increase in line output (as shown in our companion paper, Afridi et al. (2020)). However, altruism is not consistent with some of our findings, such as the positive impact of the presence of the job informant and higher proportion of experienced workers in the line, on the effort of low productivity workers.

The literature on worker productivity primarily focuses on peer effects as an explanation for variation in worker performance under production functions in which workers are substitutes and effort is observable. Knowledge spillovers or having a more productive co-worker improves worker productivity due to strategic complementarities (Falk and Ichino (2006), Mas and Moretti (2009), Lindquist et al. (2015)). We do not find evidence of non network mediated knowledge spillover effects. Peer effects on productivity mediated through social networks that create pressures to conform to a social norm, have been studied in the context of workers being substitutes in production (Bandiera et al. (2010)).[6] We do not find evidence that higher ability workers reduce their productivity. Identity motivations may also impact worker performance (Eckel and Grossman (2005)). Field experiments, however, indicate that the effect of identity on worker performance is contingent on the nature of financial incentives (Hjort (2014), Kato and Shu (2016)).[7]

Unlike almost all of the above literature, [8] we examine the role of social networks in affecting co-worker productivity when there are production complementarities and

[6]Bandiera et al. (2010) find that having a more able, self-reported friend as a co-worker increases productivity of lower ability workers but decreases productivity of higher ability workers in a UK based soft fruit producing firm.

[7]Hjort (2014) finds that ethnic homogeneity can lead to higher team output as compared to heterogeneous teams at a flower processing plant in Kenya, where workers are both substitutes and complements in the production process, and when payoffs are based on individual output. Shifting from fixed pay to performance pay based on group output, however, reduces allocative inefficiencies in multi-ethnic teams. In contrast, however, Kato and Shu (2016) show that migrant social identities mitigate competition among in-group members thereby reducing productivity in homogeneous groups when wages are relative, in a cloth manufacturing firm in China. We provide evidence against such taste based discrimination.

[8]Hjort (2014), discussed above, is an exception.

consequent financial spillovers. The only paper we are aware of that focuses on complementarity in production, and assembly lines in particular, is a lab-in-the field experiment with garment factory workers in India. Afridi et al. (2020) identify pro-social motivations between socially connected co-workers as a determinant of higher group output and better coordination, especially relevant when workers cannot observe each other's output or communicate. The O-Ring theory (Kremer (1993)) would predict that the optimal composition of lines should be based on positive assortative matching - but this is not what we find in our setting. Instead, there is a positive effect from mixing different ability level workers on the same line as long as they belong to the same social network. Our paper can therefore provide one plausible explanation of why we do not observe assortative matching by ability in garment assembly lines.

Our findings demonstrate the importance of pre-existing social connections in the form of caste-based networks, amongst workers as another channel through which economically interdependent workers can influence each other's performance and thereby affect the group output.[9] Even though our analysis is based on garment factory production lines, it is applicable to situations where the production process is organised into teams with fixed, individual wages. Our results indicate that deep social interactions outside the workplace carry implications for productivity - with corresponding implications for optimal design of production schedules and composition of teams - within the firm. Firms can, thus, leverage social networks amongst workers to relax financial constraints on worker compensation, as the insights from the microfinance literature and it's applications in labor economics have shown in different contexts (Varian (1990), Ghatak and Guinnane (1999), Bryan et al. (2015)), Heath (2018), Dhillon et al. (2019)).

The remainder of the paper is organized as follows. Section (2) describes the background of our study - the caste-based residential neighborhoods in urban India,

[9]Indeed it is these social networks that are most relevant in labor market outcomes rather than any friendships made in the workplace. We did not find significant work-based-friendships in our sample.

and the production process and worker incentives in garment factories. Section (3) summarizes the observed data regularities. We discuss our empirical methodology, report the results of our analysis in Section (4) and conduct robustness checks in Section (5). We explore mechanisms that can explain our findings in Section (6) and conclude in Section (7).

2 Background

2.1 Caste as a proxy for social networks

Workers' social networks play a significant role in the functioning of labor markets (Afridi et al. (2015)) and in ensuring migrants' economic mobility, more so in low income countries (Munshi (2014), Munshi (2019)). Historical data highlight the salience of social networks based on caste and homophily in India (Munshi (2019)).[10] Chandavarkar (1994) documents historical migration to industrial hubs within the framework of caste, kinship and village connections from India's rural areas. The rural migrants not only resided with their co-villagers, caste-fellows and relatives in the city but also obtained work with their assistance (Burnett-Hurst (1925), Gokhale (1957)). Today caste and kinship continue to be integral to individuals social networks in urban areas, particularly amongst rural migrants in the city's working-class neighborhoods.[11]

In our study we focus on India's garment manufacturing sector, which is amongst the largest providers of employment for low skilled workers offering work opportunities to rural migrants from diverse caste groups. Migrants tend to find employment through information about job openings and referrals from their caste-based networks, and may also depend on their support to weather socio-economic shocks and for risk-sharing. In our data we find that a majority (74.5%) of the garment factory

[10] Caste, a unique feature of Indian society, is inherited at birth. The caste system classifies Hindu society into hierarchical groups, based historically on occupation.

[11] 30% of the Indian population has migrated from another part of the country at some point, of which almost 15% migrate for employment (GOI (2011)).

workers obtained information about their current job from someone in their network. Informants of 65% of these workers were employees in the same factory. Conditional on the informant being still employed in the same factory as our survey respondent (50% of workers who received job information from same factory employee), we find that 42% of these workers were referred to the management by the informant, who was most likely a co-worker in the same production team or line (61.6%) and/or a neighbor (52.1%) whom they knew for some time (7.4 years).[12]

While our data show that job informants typically live close to or within the worker's residential units or migrant colonies, they often belong to the same caste groups as well. Of the workers residing in the same town in our sample, 53.5% shared the same caste category. Residential segregation by caste becomes stronger as we move from towns to clusters, colonies and lanes (63.2%, 66.3% and 83.2%, respectively, belonged to the same caste category, conditional on both caste and current residence information being available for a worker in our data). Thus, own-caste neighborhoods represent the social networks that workers derive economic benefits from.

We use workers' affiliation to broad caste categories, viz. Scheduled Caste (SC), Scheduled Tribes (ST) and Other Backward Castes (OBC), to define social networks for two reasons. First, existing evidence suggests that there is strong residential segregation by caste in urban India. While Vithayathil and Singh (2012) show high levels of residential segregation by caste at the ward level in the large metropolitan cities in contemporary India, higher than segregation by socio-economic status, Bharathi et al. (2019) find that at the census enumeration block level (smaller than a ward, with about 100-125 households) there is an even higher degree of residential segregation by caste *categories*. More recently Adukia et al. (2022) make a similar observation on residential segregation of SCs and Muslims based on data for 1.5 million neighborhoods in India. Second, Munshi (2014) points out that depending on the context,

[12]Since our survey objective was to map workplace relationships, subsequent questions on the informant's line number, relationship with the worker and referrals for her, was conditional on the informant being a current co-worker.

social networks can be organized around residential location, such as the origin village or the destination neighborhood. Thus, broad caste categories are suitable proxy for social networks in our context - relevant for residential location decisions (e.g. areas are often classified as *harijan* or low caste) and in fostering shared experiences. Narrow caste groups, viz. sub-caste or *jati*, on the other hand, are also likely to represent identity which might suggest taste based discrimination. We show later that taste based discrimination is not a likely explanation for our results. To the extent that *jati's* represent social networks, our results are fully consistent with it, and we show that our findings are robust to *jati* classification later.

2.2 Garment production and worker incentives

The manufacturing process in a garment factory encompasses multiple departments. We focus on the production department, responsible for the stitching of garments. A single factory can have multiple production or stitching floors. On each floor there are multiple production lines in which stitching machines placed one behind the other are operated by workers (see Figure A.1 in the Online Appendix).[13] Each line is assigned a particular style of garment to be produced over certain days until the production target for that garment-style is met.

There are two types of production lines: assembly and non-assembly. In an assembly line each worker contributes to the production of the garment by performing different assigned operations. She receives bundles containing cut pieces of parts of a garment at the beginning of every work hour. The stitched garment is then assembled at the front of the line.[14] The production process is, thus, both simultaneous and complementary. Hence there exist strong production externalities in the assembly line - the total number of finished garments produced by the line on a day would

[13]Besides the machine operator, who is responsible for stitching, the production line includes helpers who assist with specific operations (fold, cut, match or iron parts of garments) - about 16% of workers in a line on a day. We use the term 'worker' to denote both operators and helpers.

[14]Figure A.2 in the Online Appendix illustrates the different operations in the production of a shirt in an assembly line, e.g. attaching cuffs and stitching armholes.

depend on the productivity of the least efficient worker. Indeed, we find a significant, positive correlation between the line efficiency recorded by the factory management and that of the least efficient worker in that line in our data.

Observability of co-worker effort is imperfect. Workers can see who is sitting in their line immediately ahead and behind but they cannot directly observe each other's output. Thus a worker is more likely to be aware of those co-workers' efficiency who are seated in close proximity, even though they can see who is on the line - especially if they are known to them. On the other hand, in the less ubiquitous non-assembly lines the entire line is responsible for producing only one part of the garment, e.g. collars. Thus, all workers perform the same operation.

The management monitors workers' performance via production line supervisors. It is the supervisor's responsibility to ensure that the line meets its production targets for the work day. His (all supervisors in our sample are men) financial incentives - bonus and promotions - are linked to his line's performance, as per our discussion with the factory management. Supervisors receive a monthly bonus if their line's efficiency (averaged across workdays) in that month crosses a threshold, with a higher bonus at higher threshold.[15] Although workers receive a fixed, minimum wage per month, there are different grades of workers classified according to their skill level. The wage differential between grades is about 10 -12%. During the period of our study, workers were not offered any performance linked bonuses.

A worker's chances of being promoted to a higher grade improves with supervisor goodwill, which itself is linked to line productivity. Since the management maintains records of operational efficiency at the line level, supervisors are aware of which worker-operations are holding up the line output. Data suggest that an overwhelming majority (almost 84%) of recommendations for worker promotions in the factory are from the same line as the recommending supervisor. Moreover, the recommendations of promotion by the supervisor is positively correlated with the skill-level of the worker

[15]Supervisors receive a fixed monthly salary (higher than the workers') as well as a monthly bonus linked to line level performance, which varies between 8-14% of the salary depending on the efficiency of his line.

(ρ=0.52, p<0.01). Supervisor recommendations for worker promotion also influence the recommendations made by higher-up management – there is positive correlation between supervisor and floor-manager recommendation (ρ=0.50, p<0.01).[16] Supervisors also allot limited overtime positions to workers, which typically pay a higher hourly wage. Since overtime positions are few, more productive workers have a higher probability of receiving over time work. In essence, therefore, there exist implicit individual financial incentives linked to being a more productive worker in a line. Given the production externalities in the assembly line, the performance of co-workers in an assembly line can impact the earnings of a worker.

Workers are assigned to specific lines by the management at the time they join the factory. However, given the constrained supply of skilled workers and the high proportion of migrant laborers in this industry, worker absenteeism (and attrition) is significant (GOI (2018)).[17] The number of observed workers in a line on a workday deviates and varies day-to-day from the allocated line strength - an average daily deviation of 31%. This implies an average change in line strength of over 15 workers per day. Hence workers are moved across lines to address unanticipated absenteeism to meet production targets. Any reassignment of workers across the lines is controlled by floor or line in-charge according to the supply of and demand for workers, the relevant skill requirement and production deadlines.[18] Thus, the daily caste composition of a line varies due to unanticipated worker absenteeism and the resulting worker reallocation across lines. Our identification strategy, discussed in detail later, takes advantage of plausibly exogenous variation in the daily worker composition of production lines, and thereby the size of the caste-based network of a worker in her

[16]The data (N=431) consist of all independent recommendations from supervisors and floor managers of the production department for different types of promotions and a random sample of non-recommended workers (approx. 15% of workers per line) from one of our sampled factories.

[17]Average reported weekly absenteeism is about 10% in our sample, but is likely an underestimate. Workers switch jobs frequently in the garment industry. A typical worker in our sample was employed in the current job for 2 years but had been in the garment industry for almost 4 years. Poaching of workers is common, especially during the peak demand season. Even during our survey period, which was a normal production period, more than 8% workers exited while over 5% joined the factory.

[18]Adhvaryu et al. (2019) document limited relational trading between supervisors inside garment factories to reallocate workers in order to address worker absenteeism.

line, to estimate the impact of social network strength on productivity.

3 Data

Our data come from two factories located in the industrial hubs of Faridabad and Gurugram (both in the National Capital Region, NCR) in the state of Haryana, India. While the former factory caters to foreign buyers (89% of our sample), the latter manufactures garments for the domestic market. There are two main data sources: (1) own survey of factory workers and (2) administrative data from the factory management.

3.1 Survey data

We conducted a census of workers employed in the two factories during a regular production season from 2$^{\text{nd}}$ August 2015 - 15$^{\text{th}}$ October 2015 (approximately 61 continuous work days) to obtain information on their demographic and other individual characteristics. The resulting data on 1916 workers and 73 supervisors include all workers and supervisors in the stitching department of the sampled factories.[19] The survey gathered information on individual characteristics, including current residence, native state of residence and caste, years of experience in the garment industry, and the process of obtaining the current job, particularly referrals.

Using each state government's administrative list of Scheduled Castes (SC), Scheduled Tribes (ST) and Other Backward Castes (OBC) and the native state reported by the worker (or supervisor), we mapped the reported sub-caste or *jati* of each worker (supervisor) into 3 categories: (1) **L**, i.e. SC or ST (2) **M**, i.e. OBC and (3) **H** or high castes who do not benefit from affirmative action policies.

[19] Any new worker recruited during our study period was included in our survey.

3.2 Worker productivity and attendance data

The factory management records hourly, line level productivity by worker-operation within a line. For the purposes of our study the management also recorded the unique ID of each worker. This allowed us to obtain disaggregated worker level output, and also follow workers across lines and work days. These data were obtained for a period of 31 (continuous) working days between 8$^{\text{th}}$ September- 15$^{\text{th}}$ October 2015, a sub-set of the 61 days during which the worker census was conducted. Note that there were no major festivals that could result in group or caste specific absenteeism during our study period.

One obvious challenge in comparing worker productivity is the difference in the operations they perform. However, each style-operation combination has a specific daily target output associated with it which is set by the industrial engineer of the factory – SAM (standard allowable minutes) based on a standardized global database that includes information on the universe of garment-styles.[20] Dividing the recorded total daily output (summed over 8 hours in a work day) by the target daily output according to the SAM per worker-operation, we end up with a normalised measure of worker productivity for each style-operation. Thus, the closer the worker's actual output is to the target output, the more efficient or productive is the worker.[21] Another challenge in measuring worker productivity in an assembly line is the production externalities imposed by other workers or own productivity - this is relevant for tasks such as assembling where the availability of all pieces constrains total output. However, the target output for each worker gets adjusted according to the available work-in-progress pieces, so we are reassured that the SAM measures the true worker productivity. We calculate a worker's efficiency, therefore, as follows:

[20] The SAM is the time it takes in minutes to conduct a particular operation under ideal conditions. It is, thus, higher for more complex operations. Using the SAM for the style-operation, we calculate the target output per worker per style operation.

[21] After normalization, about 1.2% of person days had efficiency>1 (mapping into 149 workers). *t*-test shows that these 149 workers have significantly higher efficiency on other working days as well. We keep these observation in our analysis and approximate their efficiency to 1.

Daily worker efficiency = Daily output of worker/Daily target output of worker

We measure workers' line level performance in two ways. First, as the average efficiency of all workers in a line on a day and second, as the efficiency of the least efficient worker in the line on a workday since the lowest effort determines the total output (or units of complete garment) in the assembly line.[22] Data on workers' and supervisors' daily attendance was obtained from the Human Resource (HR) departments of the two factories.[23] We match workers across the survey, production and attendance data using unique worker IDs to obtain a panel of 1916 workers. Taking into account missing information across the three data sources, our final dataset consists of 1744 workers and 34,641 worker-days.[24]

Table 1, column 1, summarizes the characteristics of our sample. More than 66% of the factory workers are migrants from two large north-Indian states of U.P. and Bihar. On average, a worker has been in the garment sector for over 3.5 years and 74.5% of them obtained their current job through information from their social network. Conditional on the job informant being still employed at the factory, 42.1% of workers were referred to the job by the informant. The majority of supervisors were from M category unlike workers who were more likely to belong to H category. Almost 35% of workers belong to the same caste category as their line supervisor. These worker characteristics are described by their caste category in columns 2-4 in Table 1. The largest proportion of workers belong to the H caste category (47%) followed by M (31%) and L caste (22%), in our sample. The characteristics of workers are

[22]Management data on line efficiency is missing for 18% of our line-day observations. The correlation between line efficiency in management records and our measures of minimum (average) worker efficiency in the line is 0.19 (0.51) (both at $p<0.01$).

[23]Workers reported their unique IDs in the survey data which were cross checked using the HR data. In the export factory a card punching system was used for recording attendance. In the domestic factory, workers were required to submit their ID cards to the HR representative who would then enter their unique IDs into the computer records at the beginning of the work day.

[24]We do not have production data for 112 surveyed workers who exited the factory before we started collecting the output data. 6 workers for whom we have HR records are missing from the production data. Information on native state or *jati* or both is missing for 52 workers. We drop 2 workers for whom we have only half-day attendance information. In total, therefore, we lose 172 workers from our original sample of 1916. We do not find any significant differences in the characteristics of workers who attrited from our sample and those who were on the rolls during the collection of the production data. See Table A.1 in the Online Appendix for details.

largely similar across caste categories - in particular we find no evidence of systematic productivity differences between workers of different caste groups.[25]

Table 2, Panel A, shows the average efficiency of a worker and across worker-days on the stitching floor. Workers typically achieve only around 0.312 of their target output, on average. Worker efficiency is not statistically significantly different across caste categories. The average network strength or "proportion own caste", measured by the number of workers belonging to the caste category of the worker divided by the total number of workers in the line on a workday, is 0.395. Panel B shows the performance of a line across the sampled period. The average efficiency of a line is about 0.298 and the average minimum efficiency of line is 0.051, indicating that least performing worker is meeting only 5% of the target output. We find similar productivity statistics by line-days. The network strength in Panel B is measured by the sum of square of the shares of each caste category in a line on a day.

Figure A.3 in the Online Appendix exhibits the large variation in the line level performance cross-sectionally, averaged across work days, in terms of minimum (left panel) and average worker efficiency (right panel). The variation in performance across production lines is accompanied by wide variation in both the strength (number of workers) of a line (Online Appendix Figure A.4a) and its performance across workdays (Online Appendix Figure A.4b).[26] The binscatter plot in Figure A.5 in the Online Appendix shows that the higher the proportion of own caste workers in the line, the higher the efficiency of the worker in the line on that day. This suggests that social connections amongst co-workers, mediated through caste, may have a significant impact on individual productivity.

[25]The p-values for each pairwise t-tests of efficiency varies from 0.06 to 0.37. Using the median worker efficiency calculated for workers observed number of days, we further divide workers into low (those below median) and high ability (equal to or above median) and run a probit model regressing ability type on worker characteristics. The coefficients on caste group (L being the benchmark category) are insignificant, thus, validating the claim that productivity is not systematically correlated with caste groups.

[26]The caste composition of the Indian population is 28.2% SC or ST, 41.1% OBC and 30.8% high castes (Census 2011).

4 Methodology and Results

4.1 Identification

If workers self-select or are sorted into production lines by caste, then any relationship between worker efficiency and composition of a line may be endogenous.[27] To test our claim that the caste of a worker and the worker's observed line on a work day are independent we follow Hjort (2014) in conducting the Pearson's chi-square test. Specifically, if $P(C_i)$ denotes the probability of worker i belonging to the caste category C, and $P(L_i)$ denotes the probability of worker i being observed in line L, then $P(C_i \cap L_i)$ is the joint probability of worker in caste C sitting in line L. If the two events are truly independent then we should find that $P(C_i \cap L_i) = P(C_i) \cap P(L_i)$ holds on average. From the production data we have information on the caste composition of each line on a day, $P(C_i \cap L_i)$, and on $P(L_i)$. We perform this test for each line and each work day for both the factories in our sample. Table A.2 in the Online Appendix gives a snapshot of the caste distribution of workers in production lines on a randomly selected work day for the export factory and Online Appendix Table A.3 shows the same analysis for the domestic factory. We fail to reject the null hypothesis at 5% level of significance for all 1043 line days, except 2 (3) work days in the export (domestic) factory. Hence, we argue that worker allocation or reallocation to production lines is independent of their caste affiliation.

Our claim is further substantiated by the fact that worker absenteeism is not systematically correlated with workers' caste category (Table A.4, columns 1-2 in the Online Appendix). In addition, there is no correlation between the average number of lines a worker is observed in and her caste in our production data. Thus the changes in line-level caste composition emanating from worker reallocation due to unanticipated absenteeism are independent of own caste. Further, we do not find a systematic relationship between either a line's daily or lagged production target

[27] In our study the management did not collect information on workers' caste at the time of recruitment.

and its caste concentration on a work day (Table A.4, columns 3-4 in the Online Appendix), indicating that supervisors do not strategically adjust line caste composition in response to productivity targets. In our empirical analysis, therefore, we contend that the observed variation in the caste composition of workers in a line across work days is exogenous, conditional on worker ability and line-specific (production) trends.

4.2 Estimation methodology

Our baseline specification exploits the panel structure of our data and is given by:

$$\mathrm{Y}_{ilt} = \alpha + \beta network\_strength_{ilt} + \gamma \mathrm{X}_i + \epsilon_{ilt} \tag{1}$$

where, Y_{ilt} is the efficiency of *i-th* worker sitting in the *l-th* line on *t-th* work day, $network\_strength_{ilt}$ is defined as the number of workers belonging to *i-th* workers caste category (H, M or L) divided by the total number of workers in the line on that work day. It reflects the strength of caste based social connections a worker can have in a line on a given day. Both Y_{ilt} and $network\_strength_{ilt}$ range from 0 to 1. $\mathbf{X_i}$ is a vector of worker characteristics such as caste category, age, marital status, religion, native state, experience, education and number of reported friends in the factory. Throughout, we control for the number of workers (i.e. line strength) in the line on each workday. Standard errors are clustered at the factory-line level. β is our main coefficient of interest. If β >0 then it would suggest that having more workers of one's own caste category in the line has a positive effect on worker's productivity.

Equation (1) ignores unobserved, time invariant individual heterogeneity, such as ability, which may be correlated with the line's caste composition and also affect individual productivity. We, therefore, include individual fixed effects (FE) in subsequent specifications, besides factory floor or line FE to account for floor and line level unobservables (e.g. floor managers' and line supervisors' characteristics) as well as line-specific trends that may influence line composition as well as productivity.

To analyze line level productivity we estimate equation (1) at the line level and measure social connections amongst workers in the line by the caste concentration index (CCI) which is the sum of the square of proportion of each of the three caste categories in a line on a day. The higher the caste concentration index of a line the higher would be the caste homogeneity in that line. Hence workers in that line are more likely to belong to the same social network and be more connected. We also include the average worker level characteristics in the line, included in vector $\mathbf{X_i}$ in equation (1), and all other controls specified above, including line level fixed effects and line-level trends. The standard errors are clustered at factory-line level, as in the individual level analysis.

We conduct a host of robustness checks of our results, including accounting for unobservable heterogeneity in the style of garment produced by a line, to strengthen the causal interpretation of our main coefficient of interest.[28]

4.3 Results

4.3.1 Line composition and worker performance

The results of the analysis using equation (1) are presented in Table 3. In the top panel the sample consists of all production lines - assembly and non-assembly. Column 1 shows estimates of equation (1), where 'Network strength' is as defined in equation (1). The coefficient β is positive, suggesting that a 1 pp increase in the proportion of workers of one's own caste increases an individual worker's efficiency by 0.137 pp. In column 2 we include individual FE. The coefficient of interest remains significant at 1% level, and is comparable in magnitude. In subsequent columns we include floor (column 3) and line (column 4) FE. The magnitude and significance of the estimate is robust.

As discussed previously, we do not find a correlation between trends in line characteristics that affect productivity and its caste composition. Nevertheless we account

[28] We find that line level productivity and absenteeism are not systematically correlated when we regress the dummy $Y = 1$ if average efficiency of the line $\geq$ median average efficiency across line-days on average line-day absenteeism in a probit model.

for variation in output targets due to changes in production styles that may be correlated with both line composition and worker performance. On average, we observe 2.8 unique production styles per line over the 31 day sample period. Hence one production style runs for over 10 days in a line. Thus in column 5 we include secular trends through work week FE and line specific work week trends. β remains significantly positive, suggesting an 0.089 pp increase in worker productivity for a 1 pp rise in share of own caste workers in the line. Our results are unchanged when we account for secular work day trends in column 6 along with individual and line FE.[29]

Since the production procedure followed in assembly lines is subject to productivity spillovers unlike non-assembly lines, we separate the sample of assembly lines where each worker performs a different operation in the line in the bottom panel of Table 3. The coefficient β is comparable, suggesting 0.098 to 0.195 pp higher worker efficiency when the proportion of own caste workers in the line rises by 1 pp. This indicates that the overall effects we observe in columns 1-6 are driven by assembly lines.

To elaborate on what the estimates imply, recall that workers receive bundles of cut sub-parts of a garment at the beginning of the each work hour. Now suppose a worker receives 4 bundles of 20 pieces each, and her hourly target output is 80 stitched pieces while her daily target is 640 pieces (8 hours x 80 pieces). Given the average efficiency of 0.31, assume she manages to complete only 198 pieces. An increase of 9 pp in her daily efficiency implies that her daily output increases by approximately 58 pieces or, on average, around 7 additional stitched pieces per hour when her line becomes completely homogeneous in her caste, i.e. from having no worker of her caste to every worker belonging to her caste (0 to 1) in the line. Since the mean worker efficiency is 0.31 the most conservative estimate in column 5 suggests that worker efficiency can rise by approximately 28.7 - 31.6% when the line is perfectly

[29]Note that since we identify the effect of network strength through line-day variations, we cannot include line specific work day FE along with line and individual FE.

homogeneous, relative to perfectly heterogeneous.[30]

4.3.2 Line composition and line level performance

In Table 4 we estimate the minimum worker efficiency using equation (1) for all lines (Panel A) and only assembly lines (Panel B). In Table 4, column 1 we include only line level characteristics as controls and subsequently augment the specification with factory floor (column 2), line (column 3), week and line specific work week FE (column 4) and work day FE (column 5). A 1 pp increase in the network strength as measured by the CCI causes a 0.105 pp (column 2) to 0.168 pp (column 4) increase in the least productive worker's efficiency in the full sample. Restricting the sample to assembly lines alone, the sample size falls from 1043 to 868 but does not change our estimates much. Note that the average minimum line efficiency is just 5%, hence the estimated impact of network strength is economically meaningful. In the strictest specification with line specific weekly trends, the results suggest that the minimum efficiency of the line or the least productive workers performance can increase by 128-336% when all the workers in her line belong to her own caste network, relative to none from her network.

In Table 5 we show the results of the same analysis but when the dependent variable is the average efficiency of a worker in the line. Columns 1-5 indicate a 0.256 to 0.398 pp improvement in a line's average efficiency when the CCI increases by 1 pp, in Panel A. We restrict the sample to only assembly lines and redo the analysis in Panel B. The point estimates are similar to those shown in the top panel. Our preferred specification with line specific weekly trends suggests 88.3 - 108.7% higher average efficiency when the line is completely homogeneous, i.e. all line workers belong to same caste relative to none, given the mean average line efficiency of 30%.

[30] Alternatively, to interpret the estimates in terms of one additional worker from own caste group in the line, consider the average strength of 33 workers in a line with approximately 15 H caste workers. With one more H caste worker, the share of H increases from 46% to 48.48% (16 H of 33 workers) in the line, or by 2.48 pp. Thus, our estimates from Table 3 would indicate an increase of 0.22 - 0.35 pp in an H worker's efficiency when an additional worker from her caste group is in the line, keeping the line strength fixed.

To sum, the larger the strength of own caste network, the higher is the performance of the worker and the group. Hence the observed effects on social network strength may hold when a critical mass of socially connected co-workers constitute a line, enabling physical proximity of co-workers belonging to the same caste group. This would improve information on individual worker's ability type, observed effort and its transmission to others in the line. Not surprisingly, therefore, our overall results are driven by H caste who comprise 46.5% of workers in a line on a workday, on average (Online Appendix Table A.5).

5 Robustness

5.1 *Jati* level analysis

Administrative caste categories contain thousands of *jati* or sub-castes. *Jatis* are the narrowest endogamous social groups, historically associated with a particular type of occupation resulting in strong *jati* based labor-market networks especially within the traditional occupation of the respective *jati* (Munshi (2019)). Even though our focus is on social connections based on residential segregation, we check our main results using caste networks defined along *jati* lines with the same specification in Table A.6, Table A.7, and Table A.8 in the Online Appendix. Our results are not only robust to this alternative measure of social networks, the size effects are larger than for broader caste categories shown in Tables 3 - 5.[31]

5.2 Sample selection

Even though we find no statistical difference in workers' performance by caste, results can be biased if absenteeism or the probability that a worker is observed in the

[31] We use self reported *jatis* of the workers collected during the survey of the workers. After standardizing *jati* spellings, we utilize Risley (1908) and Singh (1996) to account for spellings, colloquial, regional verbatim differences in *jatis*. We supplemented this process with state wise SC/ST/OBC lists along with open web sources. Thus, 1744 workers map into 133 *jati(s)* with an average own *jati* proportion of 10% of the line strength at person-days level.

data is systematically correlated with worker productivity or ability. We, therefore, analyze the relationship between worker productivity and attendance for the 31 continuous work days in our sample. As shown previously in Online Appendix Table A.4, there is no systematic relationship between caste category and worker presence, but experienced workers are more likely to be observed working.[32]

Suppose, however, that more productive workers replace the less productive, absent workers in a line on a day. If this is systematically correlated with the caste composition of co-workers in a line our results above would be biased upwards. We use inverse probability weights (IPW) suggested by Moffitt et al. (1999) and Baulch and Quisumbing (2011), which intuitively gives greater weightage to workers who are more likely to be absent (and of lower productivity) on a given work day, to check the robustness of our results to this potential selection bias. We do not find any difference in the significance of the estimates in Online Appendix Table A.9, suggesting that selection on worker characteristics is not driving our results. Our results are also robust to dropping outlier workers (Online Appendix Table A.10) and workdays (Online Appendix Table A.11).[33]

5.3 Trends

Supervisors and managers may reallocate workers across or within lines purposively to meet production targets which could vary across days and may be correlated with caste categories of workers. Note that line productivity is measured at the workday level which does not allow us to control for line specific workday FE in the line level analysis. However, in addition to line-work week FE reported in Online Appendix

[32] Unbalanced panel at the line level can be an issue if the caste composition differs systematically across lines which are observed less versus those that are observed more often. However, the *t*-test suggests that the caste concentration across days doesn't differ significantly for assembly lines which are observed more versus those observed less than the median number of working days.

[33] We account for outliers by dropping the bottom and top percentile of the distribution of worker efficiency, minimum and average worker efficiency in the line (Online Appendix Table A.10). We also drop work days 18 – 24 as outliers with high average efficiency and/or high variation in line strength, relative to the average (Online Appendix Table Table A.11). Dropping two festival days – Sept 9^{th} *(Ganesh Chaturthi)* and Sept 16^{th} *(Id-ul-Zuha)* to account for any systematic absenteesim by caste (although factories were fully operational on the aforementioned days) also does not change our results.

Table 3, we include line specific garment style FE in Online Appendix Table A.12. Since we have data on production styles only for the large export oriented factory, we restrict our sample to assembly lines in this factory. In Table A.13 we also report the results for line performance analysis with line level-garment style FE, since production styles can affect line characteristics, targets and productivity. Our results are robust to secular and line specific trends throughout.

5.4 Number of clusters

A concern with our estimates in Tables 3-5 is that high intra-cluster correlation, coupled with the small number of clusters (30 production lines) in our study, would lead to incorrect standard errors. We, therefore, show bootstrapped standard errors in Online Appendix Table A.14.[34] Our standard errors are marginally higher but the main coefficient of interest remains significant, consistent with results reported in Tables 3 - 5.[35]

6 Potential mechanisms

As we discuss in the Introduction, our explanation for the findings centers around the ability of social networks to provide network based rewards or punishments when workers help their (higher ability) peers to get overtime or promotions. We discuss other possible explanations later.

The reputation based channel builds on the insights from the microfinance literature (Varian (1990), Ghatak and Guinnane (1999), Bryan et al. (2015)) and applications in labor economics (Heath (2018), Dhillon et al. (2019)) to show how social networks can solve enforcement problems when formal institutions cannot. We apply

[34]We show pair-wise bootstrapped standard errors, with (column 1) and without clustering at the line level (columns 2-4, 5 and 9), respectively. We report wild-cluster bootstrapped standard errors (Cameron et al. (2008)) in columns 6-8 and 10-12. See table notes for explanation of choice of bootstrap procedures.

[35]We also drop outlier observations, i.e. line-days on which worker strength falls in the lowest one percentile of the distribution of strength and work days on which the number of factory lines is less than 30. In this sample of 944 line-days we wild-cluster bootstrap the standard errors. Our results remain significant.

these to a context where workers are complementary in the production process. Below we provide the intuition for the results, while the details are in the Online Appendix.

When worker effort is imperfectly observed and enforced, wages are fixed, and punishment is limited (minimum wage constraints), the firm faces a moral hazard problem - workers have low incentives to put in high effort. On the one hand, the product market is highly competitive constraining the firm's ability to pay high wages to low ability workers. On the other hand, in our setting, workers who are paid close to minimum wages can get jobs easily at other factories at the same low wages, constraining the firm's ability to punish workers, even if low productivity workers can be observed. We exploit the asymmetry between high and low ability worker's incentives to work hard. While high ability workers have implicit incentives to generate goodwill from the supervisor due to higher probability of promotions, this is not the case for low ability workers.

As discussed previously, the supervisor's incentives are linked to line performance and they influence workers' chances of obtaining lucrative overtime work and promotions. High ability workers are more likely to get overtime and promotion in general, but particularly when *line* output is high, as it generates goodwill from the supervisor. Together with complementarity in production, this implies that high ability workers have incentives to push lower ability workers who are holding up line output to put in more effort by enforcing work norms or by mentoring them, and these incentives are higher the higher the promotion rank is. We assume it is costly to undertake monitoring/mentoring activities - therefore workers rationally monitor or mentor only when they have sufficiently low costs or sufficiently high benefits. Typically monitoring and mentoring is less costly for more experienced senior workers, while benefits are largest for the highest ability workers or more generally those who expect promotions to higher ranks. Commitment to the network is typically imposed through threats of exclusion from the network and/or social sanctions to deter deviations from cooperation or equivalently, rewards from cooperation (Munshi (2014)). If own-caste workers reside close to each other and depend on each other for information on jobs, referrals or financial help, these threats become credible.[36] The key assumption is that enforcing work norms/mentoring activities is more productive/less costly for workers of the same social network, given their social interactions/interdependence outside the workplace.

The description of job informant characteristics in Table 6 (Panel A), based on our worker survey data, suggests that job informants are residential neighbors and may also be co-workers in the production line. Table 6, Panel B shows that there is significant residential segregation by caste - the proportion of workers who belong to the same caste and town/cluster/colony/lane is high and increasing as the residential unit is defined more narrowly. 83.2% of workers who reside in the same lane in a colony also belong to the same caste category in our data. Consequently, the higher the own caste-proportion in the line on a day, the higher is the share of workers who co-reside, as shown in Panel C of Table 6, and the higher the chances of information on worker performance to network members and on jobs coming from co-workers/network members. Naturally, when there are more members of a worker's caste in a line, slacking can be both more observable and costly if it adversely affects the productivity of own-caste co-workers in the line which in turn reduces their financial payoffs as discussed in Section 2.2.[37]

According to the model, low performing workers who are holding up line output would increase their effort the most when share of own network in the line increases. Indeed, the effect of increasing own caste workers in the assembly line by 1 pp on a workers efficiency is larger for the lowest productivity worker (0.14–0.168 pp in column 4 of Table 4) as compared to the average worker (0.089–0.098 pp in column 4, Table 3). The lowest productivity workers are typically younger and have been

[36]Mentoring by high ability workers works by reducing the cost of effort for low ability workers - in this case, as long as the thresholds of the cost parameters remain the same, the low ability workers have to be induced to increase effort and therefore we still need network based rewards (mentoring can be interpreted as knowledge spillovers).

[37]Unfortunately, the managements denied access to overtime and earnings data due to which we are unable to directly test the effect of network strength on earnings.

in the garment industry for fewer years, according to our data - these are also the workers who depend the most on network rewards - e.g. 87.1% of workers with less than 1 year of experience obtained job information from network as opposed to 49.2% of those with 12.7 years of experience.

To further test for our proposed mechanism we interact a dummy for whether the job informant is still employed in the same factory or not with 'Network strength'.[38] If our mechanism is valid then we should see a significant positive coefficient on this interaction term. Our results suggest exactly that. In columns 1 - 4 in Table 7, we find that almost all of the effect of network strength can be explained by its interaction with informant presence in factory. In columns 5 and 9, for line level analysis, we find a negative coefficient on informant presence on the lines minimum (column 5) or average (column 9) efficiency, but a positive coefficient on the interaction terms. The total effect of informant presence is significantly positive in columns 5 and 9, but insignificant when we account for line level unobservables, which suggests that presence of an informant is a line-level characteristic (unobserved in our data).[39]

The model also suggests that that senior workers have more authority (and hence lower cost) to enforce co-operation (or mentor younger workers), so that low productivity workers should show higher performance when there are more senior workers of the same caste in the line. Indeed, we find a significant coefficient on CCI interacted with proportion of workers with higher than median years of work experience in the industry in the line (Table A.16 in the Online Appendix), suggesting that productivity of the least efficient worker increases when there are more own-caste, senior, experienced workers in the line.[40]

[38]Ideally, we should interact 'Network strength' with dummy for referee in the same line as worker but we don't have information on the line of the referee.

[39]We create a dummy variable that equals 1 if work days of a worker is greater than the median number of work days (22 days) and 0 otherwise. The coefficient on the interaction of this dummy with network strength is insignificant, as shown in Online Appendix Table A.15. Thus those attending work for fewer days did not respond significantly differently to the network strength from those who attend more often, suggesting that social networks impact workers irrespective of the number of days they interact with each other within the factory.

[40]Knowledge spillovers through (non-network) peer effects in the workplace is more likely when co-workers can observe each other's effort or output, are performing similar tasks and/or can communicate. However, as

A second explanation for our results might be altruism within the network. Our findings are consistent with the presence of pro-social preferences towards own-caste workers. Indeed, in a companion paper which is based on a lab-in-the-field experiment (Afridi et al. (2020)) simulating assembly lines, we show that directed altruism is the most likely mechanism, generating better coordination in a situation of multiple equilibria. Directed altruism, however, is not consistent with our findings on the effect of informant presence in the line (Table 7) or of the asymmetric effects of own caste senior workers in the line (Table A.16), as discussed above.

We do not find knowledge spillover effects attributable to more high ability co-workers, irrespective of caste network affiliation - the average efficiency of peers in a line l when any high ability worker shifts from her regular line to line l on a workday does not change. Further, while the O-Ring theory (Kremer (1993)) is a plausible mechanism given the complementarities in production, our empirical results are at odds with this theory's prediction that high productivity workers should have the biggest effects on lines with higher proportions of high ability workers.

Alternately, we might expect that conformism to an efficiency norm, that is network specific, can explain our results, even absent any enforcement considerations. The prediction then would be lower line level variance in individual output within the same caste group. But we do not find any effect on efficiency variance within own caste group in the line. Moreover, the more homogeneous the caste composition of the line the higher the within line variation in worker efficiency (Table A.17, Online Appendix).[41]

---

discussed previously, workers seated one behind the other in the line do not observe each other's output, and perform different operations in assembly lines. Hence spillovers are more likely to manifest in non-assembly lines. Since our results are driven by assembly lines (Tables 3-5) suggests that learning from peers (apart from network mediated learning) is unlikely to be driving our observed findings.

[41]Note that caste may be perceived as an identity rather than a network, making taste-based discrimination a possible explanation of our findings, even though we abstract from *jati*. However, we do not find a decline in the productivity of workers when network strength of *other* caste groups increases (i.e. the productivity of L workers does not depend on the fraction of H (or M) workers in the line) in a line on a workday. We also rule out the alternative explanation of side-payments between supervisor and line workers by analyzing the effect of having a supervisor of own caste on productivity, under the assumption that such side payments are likely within the same caste. We do not find any significant effects of line supervisor and worker matching on caste on productivity.

We conclude that economic interdependence within one's social network due to production externalities creates incentives for workers to put in greater effort. This can be facilitated when the network strength in the team is larger. Hence benefits of worker diversity related to information sharing and generating a variety of perspectives on a complex problem, might be less relevant in jobs involving manual labour and routine processes.

7 Conclusion

In this paper we show that the greater the strength of one's social network the higher the worker and line level productivity on a work day. Our findings suggest that when financial incentives are constrained, workers' social networks can be leveraged to improve efficiency, extending the literature on the role of social networks and job referrals, in general, and on productivity, in particular. Thus when production is team based, and tasks differ amongst the members of a team, even in the absence of explicit group based financial incentives social interdependence of group members can enforce good behavior due to the production externalities at work.

References

Adhvaryu, A., J.-F. Gauthier, A. Nyshadham, and J. Tamayo (2019): "Absenteeism, Productivity, and Relational Contracting Inside the Firm," *Working paper.*

Adukia, A., S. Asher, K. Jha, P. Novosad, and B. Tan (2022): "Residential Segregation and Unequal Access to Local Public Services in India: Evidence from 1.5m Neighborhoods," *Working Paper.*

Afridi, F., A. Dhillon, S. X. Li, and S. Sharma (2020): "Using Social Connections and Financial Incentives to Solve Coordination Failure: A Quasi-Field Experiment in India's Manufacturing Sector," *Journal of Development Economics*, 144, 1024–45.

Afridi, F., A. Dhillon, and S. Sharma (2015): "Social Networks and Labour Productivity: A Survey of Recent Theory and Evidence," *Indian Economic Review*, 50, 25–42.

Bandiera, O., I. Barankay, and I. Rasul (2010): "Social Incentives in the Workplace," *Review of Economic Studies*, 77, 417–458.

Baulch, B. and A. Quisumbing (2011): "Testing and adjusting for attrition in Household Panel Data," *CPRC Toolkit Note.*

Bharathi, N., D. V. Malghan, and A. Rahman (2019): "Village in the City: Residential Segregation in Urbanizing India," *IIM Bangalore Research Paper No. 588 (2019).*

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013): "Does Management Matter? Evidence from India," *Quarterly Journal of Economics*, 128, 1–51.

Bryan, G., D. Karlan, and J. Zinman (2015): "Referrals: Peer screening and enforcement in a consumer credit field experiment," *American Economic Journal: Microeconomics*, 7, 174–204.

Burnett-Hurst, A. (1925): *Labour and Housing in Bombay: A Study in the Economic Conditions of the Wage-Earning Classes of Bombay*, P.S. King Son, London.

Cameron, A., J. Gelbach, and D. Miller (2008): "Bootstrap-Based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, 90, 414–427.

Chandavarkar, R. (1994): *The Origins of Industrial Capitalism in India: Business Strategies and the Working Classes in Bombay 1900-1940*, Cambridge University Press.

Chang, J.-H., G. Rynhart, and P. Huynh (2016): "ASEAN in Transformation-Textiles, Clothing and Footwear:Refashioning the Future," *https://www.ilo.org/wcmsp5/groups/public/---ed_dialogue/---act_emp/documents/publication/wcms_579560.pdf, ILO.*

Che, Y.-K. and S.-W. Yoo (2001): "Optimal Incentives for Teams," *The American Economic Review*, 91, 525–541.

Dhillon, A., V. Iversen, and G. Torsvik (2019): "Employee Referral, Social Proximity and Worker," *Economic Development and Cultural Change.*

Eckel, C. and P. Grossman (2005): "Managing Diversity by Creating Team Identity," *Journal of Economic Behavior and Organization*, 58, 371–392.

Falk, A. and A. Ichino (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24, 39–58.

Ghatak, M. and T. W. Guinnane (1999): "The economics of lending with joint liability: theory and practice," *Journal of Development Economics*, 60, 195–228.

GOI (2011): "2011 Census Data," *Ministry of Home Affairs, Government of India.*

——— (2018): "Study on "Garment Sector to understand their requirement for Capacity building"," *Ministry of Textiles, Government of India.*

Gokhale, R. G. (1957): *The Bombay Cotton Mill Worker*, Millowners' Association, Bombay.

Greif, A. (1993): "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *The American Economic Review*, 83, 525–548.

Heath, R. (2018): "Why Do Firms Hire Using Referrals? Evidence from Bangladeshi Garment Factories," *Journal of Political Economy*, 126, 1691–1746.

Hjort, J. (2014): "Ethnic Divisions and Production in Firms," *Quarterly Journal of Economics*, 129, 1899–1946.

Itoh, H. (1991): "Incentives to Help in Multi-Agent Situations," *Econometrica*, 59, 611–636.

Kato, T. and P. Shu (2016): "Competition and social identity in the workplace:Evidence from a Chinese textile firm." *Journal of Economic Behavior Organization*, 131, 37–50.

Kremer, M. (1993): "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, 108, 551–575.

Lindquist, M., J. Sauermann, and Y. Zenou (2015): "Network effects on worker productivity," *Working paper.*

Mas, A. and E. Moretti (2009): "Peers at Work," *American Economic Review*, 99, 112–145.

Moffitt, R., J. Fitzgerald, and P. Gottschalk (1999): "Sample Attrition in Panel Data: The Role of Selection on Observables," *Annals of Economics and Statistics*, 55/56, 129–152.

Munshi, K. (2014): "Community Networks and the Process of Development," *Journal of Economic Perspectives*, 28, 49–76.

——— (2019): "Caste and the Indian Economy," *Journal of Economic Literature*, 57, 781–834.

Risley, S. H. (1908): *The People of India*, Thacker, Spink & Co., Calcutta.

Singh, K. S. (1996): *People of India-National Series Volume VIII*, ASI-Oxford University Press.

Varian, H. (1990): "Monitoring Agents with Other Agents," *Journal of Institutional and Theoretical Economics*, 146, 153–174.

Vithayathil, T. and G. Singh (2012): "Spaces of discrimination: Residential segregation in Indian cities," *Economic and Political Weekly*, 47, 60–66.

Table 1: Worker characteristics

		Caste Category		
	All	**L**	**M**	**H**
Characteristics	N=1744	N=384	N=543	N=817
Age (years)	29.637	28.130	29.516	30.426
	(0.164)	(0.336)	(0.305)	(0.234)
Female	0.850	0.813	0.823	0.885
	(0.009)	(0.020)	(0.016)	(0.011)
Hindu	0.931	0.982	0.890	0.935
	(0.006)	(0.007)	(0.013)	(0.009)
Married	0.756	0.695	0.757	0.785
	(0.010)	(0.024)	(0.018)	(0.014)
Secondary or above education	0.170	0.151	0.158	0.186
	(0.009)	(0.018)	(0.016)	(0.014)
Migrant Status				
From U.P.	0.402	0.383	0.457	0.375
	(0.012)	(0.025)	(0.021)	(0.017)
From Bihar	0.264	0.156	0.322	0.277
	(0.011)	(0.019)	(0.020)	(0.016)
Workers' Network				
Experience in garment manufacturing (years)	3.574	3.090	3.497	3.854
	(0.092)	(0.178)	(0.170)	(0.137)
Received information on this job opening	0.745	0.794	0.753	0.717
	(0.010)	(0.021)	(0.019)	(0.016)
Obtained this job through referral#	0.421	0.347	0.451	0.435
	(0.024)	(0.049)	(0.042)	(0.036)
Number of friends in this factory	1.754	1.818	1.772	1.714
	(0.034)	(0.073)	(0.062)	(0.048)
Line supervisor of same caste category##	0.349	0.052	0.655	0.287
	(0.011)	(0.011)	(0.021)	(0.016)

Note: #conditional on job informant being still employed in the factory. ##conditional on reporting correct/non-missing regular line number (N=1735). Standard errors in parentheses.

Table 2: Worker, line level performance and line composition

	Efficiency				Network strength
Panel A	**Worker**		**Worker-days**		
	N	Mean	N	Mean	Mean
All	1744	0.312	34,641	0.317	0.395
		(0.005)		(0.001)	(0.001)
L	384	0.308	7,604	0.309	0.248
		(0.010)		(0.003)	(0.001)
M	543	0.300	10,923	0.308	0.347
		(0.009)		(0.003)	(0.001)
H	817	0.321	16,114	0.327	0.497
		(0.007)		(0.002)	(0.001)
Panel B	**Line**		**Line-days**		
Average worker efficiency	37	0.298	1043	0.301	0.402
		(0.011)		(0.003)	(0.003)
Minimum worker efficiency	37	0.051	1043	0.050	
		(0.006)		(0.001)	

Note: Efficiency is defined as the actual output/target output. Panel A shows the average worker efficiency (overall and by caste) at worker and worker-days level. Worker efficiency is the sum of efficiency over all work days/number of work days. The network strength is measured by 'Proportion Own Caste' which is the number of workers belonging to the caste category of the worker/ total number of workers in the line on a workday. The bottom panel shows the efficiency at the line and line-day level. Average worker efficiency is the mean efficiency of workers in the line; minimum worker efficiency is the lowest worker efficiency in the line. Average number of workers in a line is 33. The network strength in Panel B is measured by the 'Caste Concentration Index' which is the sum of square of the shares of each caste category in a line on a day. Standard errors in parentheses.

Table 3: Worker performance and line composition

	Worker efficiency					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A (all lines)						
Network strength (β)	0.137***	0.141***	0.141***	0.140***	0.089**	0.124***
Prop. own caste	(0.050)	(0.049)	(0.048)	(0.048)	(0.037)	(0.041)
Constant	0.110**	0.088**	0.228**	0.169**	0.011	0.080
	(0.044)	(0.039)	(0.106)	(0.083)	(0.081)	(0.086)
Number of observations	34641	34641	34641	34641	34641	34641
Number of workers	1744	1744	1744	1744	1744	1744
Number of lines	37	37	37	37	37	37
R-square	0.038	0.557	0.558	0.563	0.608	0.578
Panel B (assembly lines)						
Network strength (β)	0.195***	0.164***	0.163***	0.160***	0.098**	0.139***
Prop. own caste	(0.046)	(0.053)	(0.052)	(0.053)	(0.041)	(0.044)
Constant	0.107*	0.062	0.213*	0.154*	-0.007	0.059
	(0.053)	(0.037)	(0.114)	(0.088)	(0.088)	(0.091)
Number of observations	32176	32176	32176	32176	32176	32176
Number of workers	1633	1633	1633	1633	1633	1633
Number of lines	31	31	31	31	31	31
R-square	0.034	0.554	0.555	0.560	0.605	0.576
Fixed effects						
Individual	No	Yes	Yes	Yes	Yes	Yes
Floor	No	No	Yes	No	No	No
Line	No	No	No	Yes	Yes	Yes
Week	No	No	No	No	Yes	No
Line x week	No	No	No	No	Yes	No
Day	No	No	No	No	No	Yes

Note: The dependent variable is the efficiency of the worker on a work day. The network strength is measured by 'Proportion Own Caste' which is the number of workers belonging to the caste category of the worker/ total number of workers in the line on a workday. Individual level controls in column 1 include dummy for H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends. All regressions control for daily line strength. Robust standard errors clustered at the line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table 4: Line level performance and line composition

	Minimum worker efficiency in the line				
	(1)	(2)	(3)	(4)	(5)
Panel A (all lines)					
Network strength (β)	0.064	0.105***	0.157***	0.168***	0.129***
CCI	(0.039)	(0.028)	(0.042)	(0.036)	(0.044)
Constant	0.218**	0.256***	0.164*	0.036	0.150**
	(0.107)	(0.090)	(0.084)	(0.074)	(0.072)
Number of observations	1043	1043	1043	1043	1043
Number of lines	37	37	37	37	37
R-square	0.537	0.594	0.700	0.844	0.726
Panel B (assembly lines)					
Network strength (β)	0.046	0.114***	0.164***	0.140***	0.124***
CCI	(0.036)	(0.033)	(0.040)	(0.035)	(0.040)
Constant	0.392***	0.309***	0.321***	0.187***	0.277***
	(0.066)	(0.082)	(0.079)	(0.065)	(0.067)
Number of observations	868	868	868	868	868
Number of lines	31	31	31	31	31
R-square	0.551	0.641	0.697	0.870	0.725
Fixed effects					
Floor	No	Yes	No	No	No
Line	No	No	Yes	Yes	Yes
Week	No	No	No	Yes	No
Line x week	No	No	No	Yes	No
Day	No	No	No	No	Yes

Note: The dependent variable is the minimum efficiency of workers in a line on a work day. The network strength is measured by the 'Caste Concentration Index' (CCI) which is the sum of square of the shares of each caste category in a line on a day. Controls include average H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends on a line-day. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table 5: Line level performance and line composition

	Average worker efficiency in the line				
	(1)	(2)	(3)	(4)	(5)
Panel A (all lines)					
Network strength (β)	0.398***	0.320***	0.363***	0.265***	0.256***
CCI	(0.067)	(0.069)	(0.105)	(0.091)	(0.093)
Constant	0.387**	0.359*	0.276	-0.022	0.171
	(0.190)	(0.180)	(0.256)	(0.171)	(0.238)
Number of observations	1043	1043	1043	1043	1043
Number of lines	37	37	37	37	37
R-square	0.321	0.345	0.491	0.803	0.578
Panel B (assembly lines)					
Network strength (β)	0.385***	0.421***	0.528***	0.326**	0.366***
CCI	(0.083)	(0.075)	(0.118)	(0.128)	(0.108)
Constant	0.389	0.365	0.571	0.078	0.441
	(0.237)	(0.233)	(0.403)	(0.256)	(0.333)
Number of observations	868	868	868	868	868
Number of lines	31	31	31	31	31
R-square	0.299	0.318	0.455	0.797	0.570
Fixed effects					
Floor	No	Yes	No	No	No
Line	No	No	Yes	Yes	Yes
Week	No	No	No	Yes	No
Line x week	No	No	No	Yes	No
Day	No	No	No	No	Yes

Note: The dependent variable is the average efficiency of workers in a line on a work day. The network strength is measured by the 'Caste Concentration Index' (CCI) which is the sum of square of the shares of each caste category in a line on a day. Controls include average H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends on a line-day. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table 6: Job networks, residential location and caste category

Panel A: Job informant characteristic	**Number of workers**	**Proportion**
Obtained informal job information	1744	0.745
Informant was employed in this factory[@]	1300	0.648
Conditional on informant still employed in this factory:		
Informant referred worker	430	0.421
Informant was a line-worker	430	0.616
Informant employed in same line as worker[#]	203	0.192
Informant was a neighbour	430	0.521
Informant was a relative	430	0.272
Informant came from native village	430	0.051
Years informant known to worker	430	7.353
Panel B: Current residential location-caste		
Same caste if residing in same town	1720	0.535
Same caste if residing in same cluster	1707	0.632
Same caste if residing in same colony	1272	0.663
Same caste if residing in same lane	848	0.832
Panel C: Current residence-caste in a line	**Number of worker-days**	**Correlation**
Prop. residing in same cluster and prop. own caste in line on workday	33862	0.033***
Prop. residing in same colony and prop. own caste in line on workday	25313	0.032***
Prop. residing in same lane and prop. own caste in line on workday	16838	0.097***

Note: [@]conditional on informal flow of job opening information; [#]smaller number of observation due to non-response. In Panels B and C the sample is in worker-days, conditional on data on both caste and unit of residential location being available for a worker. Significant at *10%,**5% and ***1%.

Table 7: Worker, line level performance and job referee presence

	Worker efficiency				Line level efficiency: Minimum worker efficiency				Line level efficiency: Average worker efficiency			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Proportion own caste	0.086*	0.086*	0.038	0.066								
	(0.050)	(0.049)	(0.037)	(0.042)								
(2) Proportion own caste x referee employed in factory	0.212***	0.215***	0.207***	0.226***								
	(0.066)	(0.061)	(0.064)	(0.063)								
(3) Caste concentration index					0.015	0.112	0.116*	0.107	0.206**	0.324**	0.221	0.299**
					(0.064)	(0.066)	(0.061)	(0.065)	(0.101)	(0.157)	(0.136)	(0.138)
(4) Proportion with referee employed in factory					-0.133*	-0.053	-0.105	-0.012	-0.220	0.004	-0.027	0.163
					(0.075)	(0.064)	(0.086)	(0.075)	(0.155)	(0.206)	(0.211)	(0.216)
(5) Caste concentration index x proportion with referee employed in factory					0.370*	0.191	0.212	0.108	0.449**	0.211	0.218	-0.094
					(0.184)	(0.134)	(0.146)	(0.148)	(0.192)	(0.355)	(0.354)	(0.363)
Constant	0.229**	0.175**	0.018	0.086	0.337***	0.231***	0.114**	0.186***	0.488**	0.355	0.087	0.142
	(0.106)	(0.084)	(0.082)	(0.087)	(0.066)	(0.067)	(0.054)	(0.061)	(0.193)	(0.317)	(0.195)	(0.296)
Effect of referee employed in factory: (4) + (5)					0.236**	0.138	0.106	0.096	0.229**	0.216	0.191	0.069
					[0.047]	[0.124]	[0.178]	[0.288]	[0.018]	[0.269]	[0.260]	[0.702]
Number of observations	34641	34641	34641	34641	1043	1043	1043	1043	1043	1043	1043	1043
Number of workers	1744	1744	1744	1744	1744	1744	1744	1744	1744	1744	1744	1744
Number of lines	37	37	37	37	37	37	37	37	37	37	37	37
R-square	0.558	0.563	0.608	0.578	0.608	0.704	0.845	0.728	0.349	0.494	0.804	0.581
Fixed effects												
Individual	Yes	Yes	Yes	Yes								
Floor	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No
Line	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Week	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Line x week	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Day	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes

Note: In columns 1-4 the dependent variable is the efficiency of the worker on a work day. In columns 5-8 the dependent variable in the minimum efficiency of the line. In columns 9-12 the dependent variable is the average efficiency of the line. Referee employed in the factory is a dummy variable that takes value 1 if the worker's job informant (conditional on job information receipt from network) is still employed in the factory. Proportion with referee employed in factory is the proportion of workers in the line whose referee is employed in the factory (conditional on job information receipt from network). Controls include average H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends on a line-day in columns 5 and 9. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. p-values reported in square brackets. Significant at *10%, **5% and ***1%.

8 ONLINE APPENDIX

APPENDIX A: Additional Results

Table A.1: Worker characteristics

Characteristics	Original sample N=1916	Analysis sample N=1744
Age (years)	29.44	29.64
	(0.157)	(0.164)
Female	0.848	0.850
	(0.008)	(0.009)
Hindu	0.928	0.931
	(0.006)	(0.006)
Married	0.749	0.756
	(0.010)	(0.010)
Secondary or above education	0.169	0.170
	(0.009)	(0.009)
H	0.470	0.468
	(0.012)	(0.012)
M	0.308	0.311
	(0.011)	(0.011)
L	0.222	0.220
	(0.010)	(0.010)
Migrant Status		
From U.P.	0.404	0.402
	(0.011)	(0.012)
From Bihar	0.259	0.264
	(0.010)	(0.011)
Workers' network		
Experience in garment manufacturing (years)	3.498	3.574
	(0.087)	(0.092)
Received information on this job opening	0.743	0.745
	(0.010)	(0.010)
Obtained this job through referral#	0.422	0.421
	(0.023)	(0.024)
Number of friends in this factory	1.735	1.754
	(0.032)	(0.034)
Line supervisor of same caste category	0.347	0.349
	(0.011)	(0.011)

Note:# conditional on referee being still employed in the factory. Caste data for 1857 workers in column 1. Standard errors in parentheses.

Table A.2: Chi-square test of exogeneity of caste assignment to line (export factory)

Line Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**Total**
Caste Category																											
L	13	7	12	15	11	9	13	11	15	11	8	13	12	9	10	9	2	5	3	6	6	2	5	8	5	7	**227**
	10	8	10	10.2	10	10.6	9.6	8.9	10.4	11.3	13.7	9.6	12.4	10	9.8	10.9	3.5	9.8	6.7	6.5	3.7	4.8	8.7	6.3	6.7	5	**227**
	0.9	0.1	0.4	2.2	0.1	0.3	1.2	0.5	2	0	2.4	1.2	0	0.1	0	0.3	0.6	2.3	2.1	0	1.4	1.6	1.6	0.5	0.4	0.8	**23.2**
M	16	12	14	14	7	16	16	15	10	14	20	15	18	12	16	13	6	15	9	7	3	3	12	6	11	8	**308**
	13.6	10.9	13.6	13.9	13.6	14.4	13	12.1	14.1	15.3	18.6	13	16.8	13.6	13.3	14.7	4.7	13.3	9.1	8.8	5	6.5	11.8	8.5	9.1	6.8	**308**
	0.4	0.1	0	0	3.2	0.2	0.7	0.7	1.2	0.1	0.1	0.3	0.1	0.2	0.6	0.2	0.3	0.2	0	0.4	0.8	1.9	0	0.8	0.4	0.2	**13.1**
H	17	18	20	18	28	24	15	15	23	27	35	16	27	25	19	28	8	25	19	17	8	17	23	15	15	8	**510**
	22.4	18.1	22.4	22.9	22.4	23.9	21.5	20	23.4	25.4	30.7	21.5	27.8	22.4	22	24.4	7.8	22	15.1	14.6	8.3	10.7	19.5	14.2	15.1	11.2	**510**
	1.3	0	0.3	1.1	1.4	0	2	1.3	0	0.1	0.6	1.4	0	0.3	0.4	0.5	0	0.4	1	0.4	0	3.7	0.6	0.1	0	0.9	**17.6**
Total	**46**	**37**	**46**	**47**	**46**	**49**	**44**	**41**	**48**	**52**	**63**	**44**	**57**	**46**	**45**	**50**	**16**	**45**	**31**	**30**	**17**	**22**	**40**	**29**	**31**	**23**	**1045**
	46	**37**	**46**	**47**	**46**	**49**	**44**	**41**	**48**	**52**	**63**	**44**	**57**	**46**	**45**	**50**	**16**	**45**	**31**	**30**	**17**	**22**	**40**	**29**	**31**	**23**	**1045**
	2.7	**0.2**	**0.7**	**3.3**	**4.6**	**0.4**	**3.9**	**2.4**	**3.2**	**0.2**	**3.1**	**3**	**0.1**	**0.6**	**1**	**1.1**	**1**	**3**	**3.1**	**0.8**	**2.3**	**7.1**	**2.2**	**1.3**	**0.8**	**2**	**54**

Note: Data for the larger factory with 26 lines working on a randomly selected workday. There are three corresponding rows for each caste group. The first row shows the actual proportion of L/M/H in each line. The second row shows the expected proportion under the null hypothesis of independence of probability of caste and line. The third row shows the contribution of Pearsons χ^2. Pearsons χ^2 statistics is 53.975 with 50 degrees of freedom and p value =0.325. We cant reject the null hypothesis of independence of caste distribution and line composition. Similar results for all 31 workdays. p value ranges from 0.629 to 0.026 with two working days having$p$ value <0.05.

Table A.3: Chi-square test of exogeneity of caste assignment to line (domestic factory)

Line Number / Caste Category	1	2	3	4	5	6	7	8	9	10	Total
L	4	2	1	4	4	6	4	2	4	3	**34**
	3.3	3	3.8	4.1	2.5	6.6	2.5	1	2.5	4.6	**34**
	0.1	0.4	2.1	0	0.8	0.1	0.8	1	0.8	0.5	**6.7**
M	4	5	14	9	4	12	4	1	4	9	**66**
	6.4	5.9	7.4	7.9	4.9	12.8	4.9	2	4.9	8.9	**66**
	0.9	0.1	5.9	0.2	0.2	0.1	0.2	0.5	0.2	0	**8.2**
H	5	5	0	3	2	8	2	1	2	6	**34**
	3.3	3	3.8	4.1	2.5	6.6	2.5	1	2.5	4.6	**34**
	0.9	1.3	3.8	0.3	0.1	0.3	0.1	0	0.1	0.4	7.3
Total	**0.9**	**1.3**	**3.8**	**0.3**	**0.1**	**0.3**	**0.1**	**0**	**0.1**	**0.4**	**7.3**
	13	**12**	**15**	**16**	**10**	**26**	**10**	**4**	**10**	**18**	**134**
	1.9	**1.8**	**11.8**	**0.4**	**1.1**	**0.4**	**1.1**	**1.4**	**1.1**	**1**	**22.1**

Note: Data for the smaller factory with 10 lines working on a randomly selected workday. There are three corresponding rows for each caste group. The first row shows the actual proportion of L/M/H in each line. The second row shows the expected proportion under the null hypothesis of independence of probability of caste and line. The third row shows the contribution of Pearsons χ^2. Pearsons χ^2 statistics is 22.13 with 18 degrees of freedom and p value =0.226. We cant reject the null hypothesis of independence of caste distribution and line composition. Similar results for all 31 workdays. p value ranges from 0.802 to 0.017 with three working days having p value$<$0.05.

Table A.4: Worker attendance, production targets and caste composition

	Worker level		*Line level (SAM)*	
	Present rate	**Work-days**	**Target**	**Lag-target**
	(1)	(2)	(3)	(4)
Age (years)	0.001***	0.058	1.350	1.050
	(0.000)	(0.035)	(0.885)	(0.827)
Married	-0.013*	-1.566***	1.855	15.012
	(0.007)	(0.515)	(10.119)	(9.178)
Female	-0.006	1.788***	-27.293**	-24.738**
	(0.008)	(0.562)	(11.968)	(11.495)
Native state Bihar	0.010**	0.465	-9.064	-6.586
	(0.005)	(0.295)	(10.345)	(7.879)
Hindu	0.033***	2.076***	9.164	2.066
	(0.010)	(0.617)	(12.376)	(11.272)
Secondary education or more	0.003	0.172	9.926	9.577
	(0.005)	(0.410)	(12.498)	(12.929)
Obtained job information informally	0.000	0.874*	-10.704	-8.598
	(0.006)	(0.464)	(8.105)	(7.689)
Experience (years)	-0.001***	0.233***	-2.140**	-1.942**
	(0.000)	(0.056)	(0.951)	(0.899)
Number of reported friends	0.000	0.205	-1.336	-2.379
	(0.002)	(0.124)	(1.962)	(2.287)
Line strength	-0.000***	-0.204***	0.128*	0.143**
	(0.000)	(0.009)	(0.068)	(0.061)
H	0.003	-0.371		
	(0.006)	(0.285)		
M	0.006	0.177		
	(0.007)	(0.438)		
Line supervisor same caste	0.003	0.197		
	(0.005)	(0.275)		
Caste concentration index			4.500	1.073
			(10.581)	(11.056)
Constant	0.882***	16.231***	6.878	8.264
	(0.013)	(0.768)	(32.140)	(31.634)
Number of observations	1731	1731	751	681
Number of workers	1731	1731	1548	1548
Number of lines	37	37	27	27
R-square	0.052	0.197	0.462	0.485
Line Fixed Effects	Yes	Yes	Yes	Yes

Note: Col (1) uses factory attendance data. 'Attendance rate' is the number of days present/number of on-roll days for each worker (excluding half days, forming 0.45 of the attendance person days). The mean attendance rate is 0.923. Col (2) is based on the production data. 'Work-days' is the count of days a worker appears in the productivity data (excluding half days, 0.30% of the worker days). Attendance data missing for 4 workers; reported line information missing for 9 workers. Individual level controls as elucidated in Table 3. Sample in cols (3)-(4) conditional on availability of line-daily target (SAM for 1 finished product). Dependent variable in col (3) (col(4)) is line-daily target on day t (t-1). Some lines did not operate the day before, hence missing observations in col (4). Line-day level controls as elucidated in Table 4. Robust standard errors, clustered at the line level, in parentheses. Significant at *10%, **5% and ***1%.

Table A.5: Worker performance and line composition (by caste type)

	L	**M**	**H**
	(1)	(2)	(3)
Network strength (β)	-0.031	-0.003	0.151*
Prop. own caste	(0.093)	(0.094)	(0.075)
Constant	0.256**	0.175	-0.028
	(0.119)	(0.135)	(0.076)
Number of observations	7604	10923	16114
Number of lines	37	37	37
R-square	0.602	0.639	0.614
Fixed effects			
Individual	Yes	Yes	Yes
Floor	No	No	No
Line	Yes	Yes	Yes
Week	Yes	Yes	Yes
Line x week	Yes	Yes	Yes
Day	No	No	No

Note: Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.6: Worker performance and line composition *(jati level)*

	Worker efficiency					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A (all lines)						
Network strength (β)	0.078	0.249***	0.248***	0.299***	0.125*	0.251***
Prop. own jati	(0.067)	(0.087)	(0.086)	(0.072)	(0.068)	(0.070)
Constant	0.142***	0.112***	0.132	0.178**	0.230***	0.091
	(0.042)	(0.034)	(0.086)	(0.083)	(0.082)	(0.087)
Number of observations	34641	34641	34641	34641	34641	34641
Number of workers	1744	1744	1744	1744	1744	1744
Number of lines	37	37	37	37	37	37
R-square	0.035	0.557	0.557	0.563	0.607	0.578
Panel B (assembly lines)						
Network strength (β)	0.087	0.263**	0.261**	0.313***	0.114	0.249***
Prop. own jati	(0.072)	(0.098)	(0.097)	(0.081)	(0.077)	(0.081)
Constant	0.157***	0.095***	0.121	0.169*	0.234**	0.078
	(0.049)	(0.033)	(0.093)	(0.089)	(0.089)	(0.092)
Number of observations	32176	32176	32176	32176	32176	32176
Number of workers	1633	1633	1633	1633	1633	1633
Number of lines	31	31	31	31	31	31
R-square	0.030	0.554	0.554	0.560	0.604	0.576
Fixed effects						
Individual	No	Yes	Yes	Yes	Yes	Yes
Floor	No	No	Yes	No	No	No
Line	No	No	No	Yes	Yes	Yes
Week	No	No	No	No	Yes	No
Line x week	No	No	No	No	Yes	No
Day	No	No	No	No	No	Yes

Note: The dependent variable is the efficiency of the worker on a work day. The network strength is measured by 'Proportion Own *Jati*' which is the number of workers belonging to the *jati* category of the worker/ total number of workers in the line on a workday. Individual level controls in column 1 include dummy for H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends. All regressions control for daily line strength. Robust standard errors clustered at the line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.7: Line level performance and line composition (*jati* level)

	Minimum worker efficiency in the line				
	(1)	(2)	(3)	(4)	(5)
Panel A (all lines)					
Network strength (β)	0.260***	0.286***	0.345***	0.291***	0.323***
JCI	(0.056)	(0.036)	(0.050)	(0.061)	(0.047)
Constant	0.264***	0.178**	0.254***	0.104	0.246***
	(0.084)	(0.078)	(0.065)	(0.064)	(0.053)
Number of observations	1043	1043	1043	1043	1043
Number of lines	37	37	37	37	37
R-square	0.578	0.627	0.728	0.855	0.754
Panel B (assembly lines)					
Network strength (β)	0.196***	0.251***	0.337***	0.204***	0.303***
JCI	(0.060)	(0.035)	(0.057)	(0.051)	(0.055)
Constant	0.400***	0.251***	0.282***	0.174**	0.256***
	(0.069)	(0.070)	(0.079)	(0.073)	(0.066)
Number of observations	868	868	868	868	868
Number of lines	31	31	31	31	31
R-square	0.578	0.666	0.722	0.874	0.749
Fixed effects					
Floor	No	Yes	No	No	No
Line	No	No	Yes	Yes	Yes
Week	No	No	No	Yes	No
Line x week	No	No	No	Yes	No
Day	No	No	No	No	Yes

Note: The dependent variable is the minimum efficiency of workers in a line on a work day. The network strength is measured by the '*Jati* Concentration Index' (JCI) which is the sum of square of the shares of each *jati* category in a line on a day. Controls include average H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends on a line-day. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.8: Line level performance and line composition (*jati* level)

	Average worker efficiency in the line				
	(1)	(2)	(3)	(4)	(5)
Panel A (all lines)					
Network strength (β)	0.472***	0.539***	0.663***	0.396***	0.568***
JCI	(0.119)	(0.112)	(0.132)	(0.120)	(0.106)
Constant	0.373**	0.313*	0.440*	0.065	0.336
	(0.168)	(0.167)	(0.249)	(0.173)	(0.232)
Number of observations	1043	1043	1043	1043	1043
Number of lines	37	37	37	37	37
R-square	0.299	0.351	0.507	0.806	0.594
Panel B (assembly lines)					
Network strength (β)	0.494**	0.611***	0.813***	0.335**	0.624***
JCI	(0.189)	(0.164)	(0.174)	(0.164)	(0.158)
Constant	0.287	0.171	0.449	0.033	0.369
	(0.212)	(0.234)	(0.412)	(0.278)	(0.341)
Number of observations	868	868	868	868	868
Number of lines	31	31	31	31	31
R-square	0.285	0.316	0.463	0.795	0.578
Fixed effects					
Floor	No	Yes	No	No	No
Line	No	No	Yes	Yes	Yes
Week	No	No	No	Yes	No
Line x week	No	No	No	Yes	No
Day	No	No	No	No	Yes

Note: The dependent variable is the average efficiency of workers in a line on a work day. The network strength is measured by the '*Jati* Concentration Index' (JCI) which is the sum of square of the shares of each *jati* category in a line on a day. Controls include average H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends on a line-day. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.9: Worker performance and line composition (inverse probability weights)

	Worker efficiency									
	Original estimates					*With inverse probability weights*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Network strength (β)	0.141***	0.141***	0.140***	0.089**	0.124***	0.141***	0.140***	0.140***	0.089**	0.123***
Prop. own caste	(0.049)	(0.048)	(0.048)	(0.037)	(0.041)	(0.049)	(0.049)	(0.048)	(0.037)	(0.041)
Constant	0.088**	0.228**	0.169**	0.011	0.080	0.087**	0.228**	0.169*	0.010	0.079
	(0.039)	(0.106)	(0.083)	(0.081)	(0.086)	(0.039)	(0.106)	(0.083)	(0.082)	(0.086)
Number of observations	34641	34641	34641	34641	34641	34623	34623	34623	34623	34623
R-square	0.557	0.558	0.563	0.608	0.578	0.557	0.557	0.562	0.607	0.577
Fixed Effects										
Individual	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Floor	No	Yes	No	No	No	No	Yes	No	No	No
Line	No	No	Yes	Yes	No	No	No	Yes	Yes	No
Week	No	No	No	Yes	No	No	No	No	Yes	No
Line x week	No	No	No	Yes	No	No	No	No	Yes	No
Day	No	No	No	No	Yes	No	No	No	No	Yes

Note: The dependent variable is the efficiency of the worker on a work day. The network strength is measured by 'Proportion Own Caste' which is the number of workers belonging to the caste category of the worker/ total number of workers in the line on a workday. The sample consist of all lines. Original estimates from Table 3 in columns 1-5. Regressions weighted by inverse of the probability (IPW) of worker being present on a workday in columns 6-10. We first regress worker attendance (=1 if present and 0 otherwise) on worker characteristics (age, secondary level or higher education, married, female, Hindu, native state (Bihar), caste dummies, experience in garment manufacturing (in years), used social ties for obtaining job information for the current job, reported number of friends in the current factory and reported line strength) to predict the probability of the worker being present on a workday. We then use the inverse of the predicted probability of each workers attendance from this probit model to weight our main estimating equation. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.10: Worker, line level performance and line composition (dropping outlier workers and lines)

	Worker efficiency		*Line level efficiency*			
			Minimum worker		*Average worker*	
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion own caste	0.130***	0.076**				
	(0.044)	(0.032)				
Caste concentration index			0.104***	0.083	0.358***	0.262***
			(0.038)	(0.057)	(0.093)	(0.086)
Constant	0.097	-0.067	0.238***	0.113*	0.298	0.040
	(0.085)	(0.087)	(0.066)	(0.057)	(0.252)	(0.156)
Equality of coeff between specifications	[0.102]		[0.501]		[0.354]	
Number of observations	33605	33605	1021	1021	1023	1023
R-square	0.556	0.604	0.698	0.833	0.483	0.798
Fixed effects						
Individual	Yes	Yes				
Floor	No	No	No	No	No	No
Line	Yes	Yes	Yes	Yes	Yes	Yes
Week	No	Yes	No	Yes	No	Yes
Line x week	No	Yes	No	Yes	No	Yes
Day	No	No	No	No	No	No

Note: The top and bottom percentile of the distribution of the reported outcome is dropped from the sample. p-values reported in square brackets. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.11: Worker, line level performance and line composition (dropping outlier workdays)

	Worker efficiency			*Line level efficiency*					
				Minimum worker efficiency			*Average worker efficiency*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Proportion own caste	0.142**	0.086**	0.119**						
	(0.057)	(0.042)	(0.050)						
Caste concentration index				0.149***	0.166***	0.119**	0.354***	0.211*	0.238**
				(0.043)	(0.047)	(0.045)	(0.123)	(0.110)	(0.109)
Constant	0.229*	0.389***	0.176*	0.111	0.022	0.110	0.218	0.002	0.138
	(0.123)	(0.112)	(0.104)	(0.079)	(0.074)	(0.068)	(0.245)	(0.162)	(0.226)
Number of observations	26475	26475	26475	797	797	797	797	797	797
Number of workers	1741	1741	1741	1741	1741	1741	1741	1741	1741
Number of lines	37	37	37	37	37	37	37	37	37
R-square	0.568	0.616	0.585	0.723	0.861	0.749	0.484	0.822	0.587
Fixed effects									
Individual	Yes	Yes	Yes						
Line	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week	No	Yes	No	No	Yes	No	No	Yes	No
Line x week	No	Yes	No	No	Yes	No	No	Yes	No
Day	No	No	Yes	No	No	Yes	No	No	Yes

Note: We drop outlier workdays 18 to 24 with high average efficiency and/or high variation in line strength, relative to the average. The dependent variable is the efficiency of the worker on a work day in columns 1-3; minimum efficiency in columns 4-6; average efficiency in columns 7-9. The network strength is measured by 'Proportion Own Caste', which is the number of workers belonging to the caste category of the worker/ total number of workers in the line on a workday in columns 1-3 and 'Caste Concentration Index', which is the sum of the square of the shares of each caste category in a line on a day in columns 4-9. All regressions control for daily line strength. Robust standard errors, clustered at line-day level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.12: Worker performance and composition (with style fixed effects)

	Worker level				
	(1)	(2)	(3)	(4)	(5)
Proportion own caste	0.146***	0.137***	0.108**	0.109**	0.111**
	(0.050)	(0.049)	(0.044)	(0.044)	(0.043)
Constant	0.099*	-0.307***	-0.281***	-0.283***	-0.280***
	(0.058)	(0.079)	(0.075)	(0.078)	(0.080)
Number of observations	30621	30621	30621	30621	30621
Number of workers	1548	1548	1548	1548	1548
Number of lines	27	27	27	27	27
Number of styles	45	45	45	45	45
R-square	0.581	0.584	0.591	0.595	0.597
Fixed effects					
Individual	Yes	Yes	Yes	Yes	Yes
Style	Yes	Yes	Yes	Yes	Yes
Line	No	Yes	Yes	Yes	Yes
Line x style	No	No	Yes	Yes	Yes
Week	No	No	No	Yes	No
Day	No	No	No	No	Yes

Note: The sample consists of assembly lines (exporting factory) for which we have daily-style information. On average, 2.8 unique styles ran per line during our sample period. The dependent variable is the efficiency of the worker on a work day. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.13: Line level performance and composition (with style fixed effects)

	Minimum worker efficiency				*Average worker efficiency*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Caste concentration index	0.203***	0.163***	0.158***	0.158***	0.526***	0.428***	0.377***	0.404***
	(0.046)	(0.044)	(0.044)	(0.043)	(0.127)	(0.116)	(0.120)	(0.118)
Constant	0.106	0.113*	0.092	0.101*	0.253	0.387	0.209	0.378*
	(0.065)	(0.062)	(0.060)	(0.057)	(0.219)	(0.228)	(0.212)	(0.204)
Number of observations	765	765	765	765	765	765	765	765
Number of lines	27	27	27	27	27	27	27	27
Number of styles	45	45	45	45	45	45	45	45
R-square	0.603	0.648	0.657	0.663	0.693	0.725	0.751	0.764
Fixed effects								
Style	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Line	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Line x style	No	No	Yes	Yes	No	No	Yes	Yes
Week	No	No	Yes	No	No	No	Yes	No
Day	No	No	No	Yes	No	No	No	Yes

Note: The sample consists of assembly lines (exporting factory) for which we have daily-style information. On average, 2.8 unique styles ran per line during our sample period. The dependent variable is minimum efficiency in columns 1-4; average efficiency in columns 5-8. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.14: Worker, line level performance and composition (bootstrapped standard errors)

	Worker efficiency				Line level efficiency: Minimum worker efficiency				Line level efficiency: Average worker efficiency			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Proportion own caste	0.141***	0.140***	0.089**	0.124***								
	[0.006]	[0.001]	[0.014]	[0.002]								
Caste concentration index					0.157***	0.157**	0.168***	0.129**	0.363***	0.363***	0.265***	0.256***
					[0.004]	[0.011]	[0.000]	[0.036]	[0.006]	[0.004]	[0.010]	[0.006]
Constant	0.088**	0.169	0.011	0.080	0.066	0.164	0.036	0.150*	0.248	0.276	-0.022	0.171
	[0.016]	[0.193]	[0.937]	[0.555]	[0.571]	[0.122]	[0.664]	[0.088]	[0.411]	[0.302]	[0.925]	[0.492]
Number of observations	34641	34641	34641	34641	1043	1043	1043	1043	1043	1043	1043	1043
Number of workers	1744	1744	1744	1744	1744	1744	1744	1744	1744	1744	1744	1744
Number of lines	37	37	37	37	37	37	37	37	37	37	37	37
R-square	0.557	0.030	0.129	0.063	0.246	0.700	0.844	0.726	0.123	0.491	0.803	0.578
Fixed effects												
Individual	Yes	Yes	Yes	Yes								
Line	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Line x week	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Day	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes

Note: The sample consist of all lines. p-values in parentheses. The network strength is measured by 'Proportion Own Caste' which is the number of workers belonging to the caste category of the worker/ total number of workers in the line on a workday in columns 1-4, and by the 'Caste Concentration Index' which is the sum of square of the shares of each caste category in a line on a day in columns 5-12. Regressions results with pairwise bootstrapped standard errors clustered at line level in column 1; pairwise bootstrapped standard errors in columns 2-4, 5, 9. One of the limitation of available bootstrap procedures is that they do not run if the units of observation shifts across clusters with fixed effects at the clustering unit. Since individuals move across lines in our data, we are not able to run worker level specifications with line fixed effects and bootstrapped standard errors clustered at line level. The line-level data does not pose such challenge and we present wild cluster bootstrapped standard errors (at line level) in col (6)-(8) and (10)-(12). All regressions control for daily line strength. 2000 replications across all regressions. Significant at *10%, **5% and ***1%.

Table A.15: Worker performance and work days

	Worker level					
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion own caste	0.158***	0.078	0.077	0.080	0.018	0.065
	(0.057)	(0.062)	(0.060)	(0.057)	(0.057)	(0.055)
Proportion own caste x	-0.034	0.099	0.100	0.094	0.113	0.092
Above median attendance	(0.064)	(0.070)	(0.069)	(0.070)	(0.071)	(0.068)
Constant	0.095*	0.086**	0.226**	0.173**	0.017	0.084
	(0.049)	(0.039)	(0.105)	(0.082)	(0.081)	(0.085)
Number of observations	34641	34641	34641	34641	34641	34641
Number of workers	1744	1744	1744	1744	1744	1744
Number of lines	37	37	37	37	37	37
R-square	0.039	0.557	0.558	0.563	0.608	0.578
Fixed effects						
Individual	No	Yes	Yes	Yes	Yes	Yes
Floor	No	No	Yes	No	No	No
Line	No	No	No	Yes	Yes	Yes
Week	No	No	No	No	Yes	No
Line x week	No	No	No	No	Yes	No
Day	No	No	No	No	No	Yes

Note: The dependent variable is the efficiency of the worker on a work day. Individual level controls in column 1 include dummy for H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends. Above median attendance is a dummy variable that takes value 1 if the number of work days of a worker $\geq$ median work days; 0 otherwise. Median working days = 22. Individual controls in column 1 include dummy for H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience, and number of reported co-workers who are friends. All regressions control for daily line strength. Robust standard errors, clustered at the line level, in parentheses. Significant at *10%,**5% and ***1%.

Table A.16: Worker performance, experience and network strength

	Minimum worker efficiency				
	(1)	(2)	(3)	(4)	(5)
Caste concentration index (CCI)	-0.158*	-0.113	-0.028	-0.026	-0.095
	(0.086)	(0.084)	(0.087)	(0.123)	(0.081)
Proportion high experience	-0.262***	-0.238***	-0.170***	-0.159***	-0.181***
	(0.077)	(0.073)	(0.054)	(0.058)	(0.048)
Proportion high experience x CCI	0.512***	0.501***	0.399**	0.360	0.479***
	(0.171)	(0.164)	(0.149)	(0.216)	(0.145)
Constant	0.342***	0.341***	0.265***	0.118	0.266***
	(0.078)	(0.063)	(0.079)	(0.078)	(0.071)
Number of observations	1043	1043	1043	1043	1043
Number of lines	37	37	37	37	37
R-square	0.570	0.618	0.709	0.848	0.737
Fixed effects					
Floor	No	Yes	No	No	No
Line	No	No	Yes	Yes	Yes
Week	No	No	No	Yes	No
Line x week	No	No	No	Yes	No
Day	No	No	No	No	Yes

Note: The dependent variable is the minimum efficiency of workers in a line on a work day. 'Proportion high experience' is the number of workers with above or equal to median years of experience in the garment industry sitting in line l on day d /strength in line l on day d. Median experience in garment industry for 1744 workers is 2.129 years. Controls include average H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends on a line-day. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A.17: Dispersion in worker performance and network strength

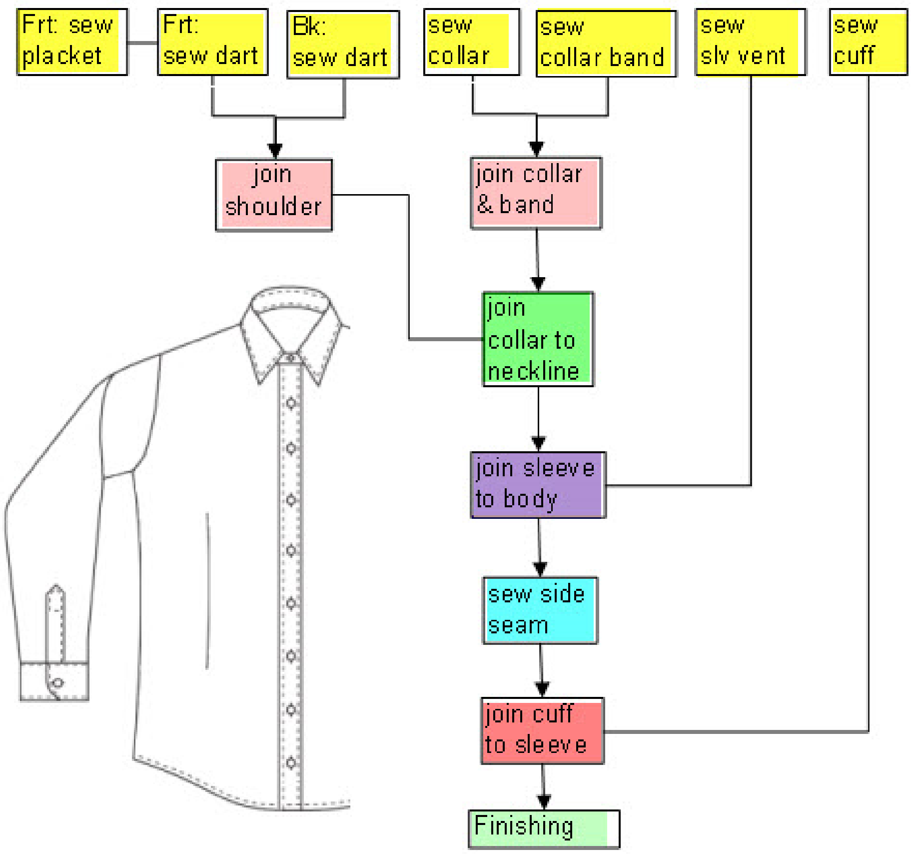
	Dispersion in worker productivity				
	(1)	(2)	(3)	(4)	(5)
Network strength (β)	0.204***	0.123***	0.150**	0.090	0.142**
CCI	(0.051)	(0.040)	(0.070)	(0.057)	(0.070)
Constant	0.162	0.142	0.021	0.008	-0.025
	(0.125)	(0.106)	(0.153)	(0.109)	(0.161)
Number of observations	1041	1041	1041	1041	1041
Number of lines	37	37	37	37	37
R-square	0.465	0.543	0.619	0.823	0.631
Fixed effects					
Floor	No	Yes	No	No	No
Line	No	No	Yes	Yes	Yes
Week	No	No	No	Yes	No
Line x week	No	No	No	Yes	No
Day	No	No	No	No	Yes

Note: The dependent variable is the standard deviation of efficiency of all workers sitting in line l on day d. We lose 2 line-days with line strength of 1 worker out of 1043 line-days while calculating standard deviation. Controls include average H, M, age, married, woman, Hindu, migrant from Bihar, received information on job opening through network, secondary or higher level of education, years of experience and number of reported co-workers who are friends on a line-day. All regressions control for daily line strength. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Figure A.1: Factory floor and line organisation

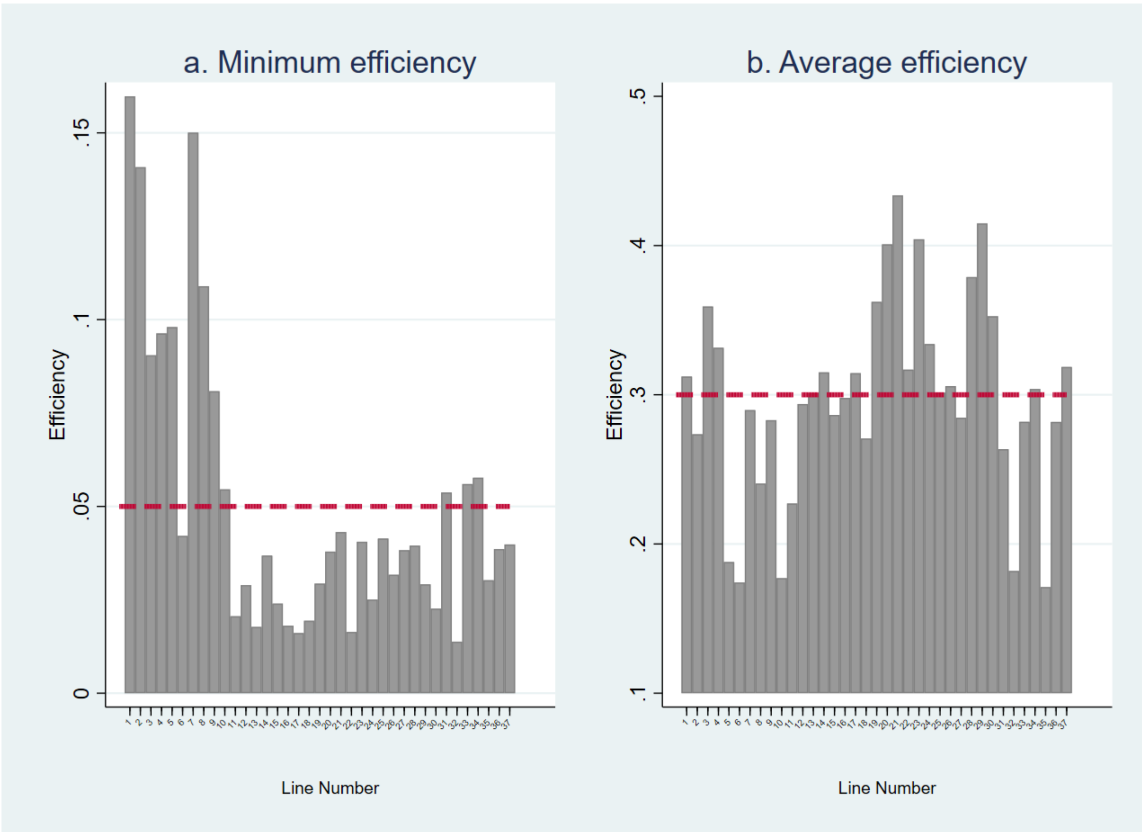
Location: Faridabad
Source: icrw.org

Figure A.2: Manufacturing process of a shirt



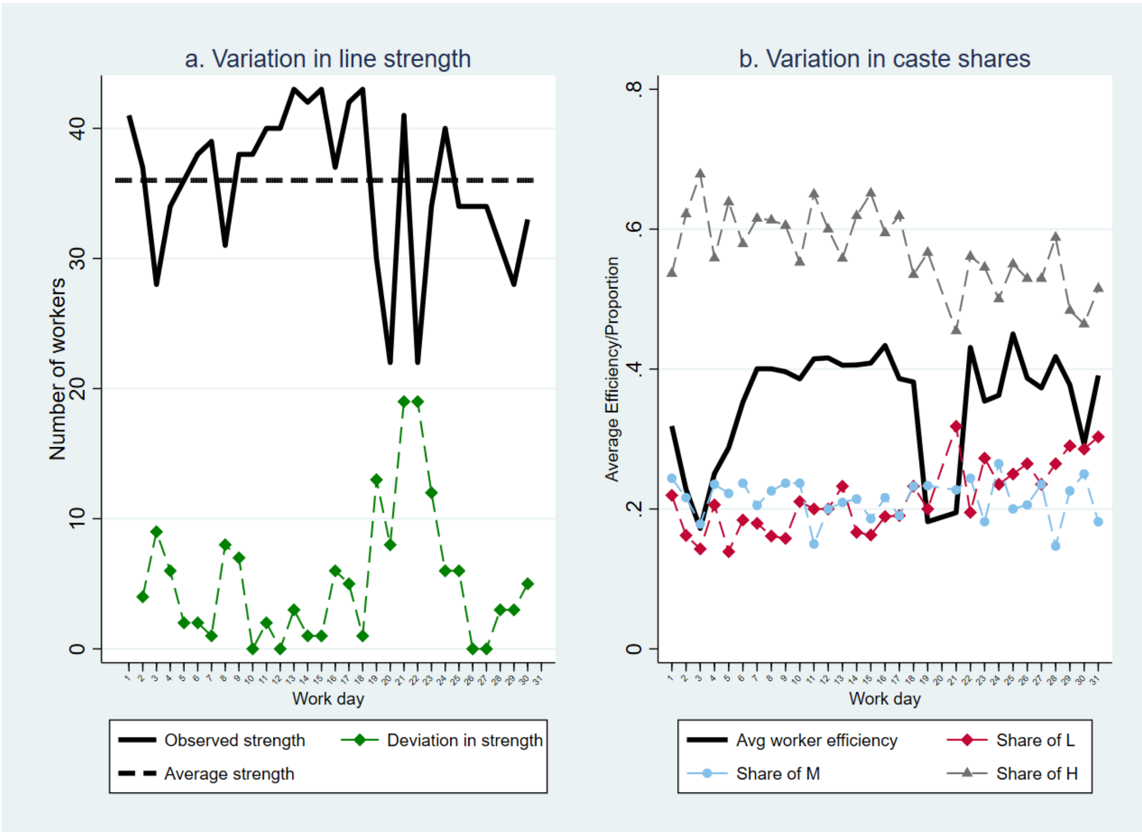
Source: `https://www.pinterest.co.uk/neelamparveen78/garment-production-manufacturing`

Figure A.3: Line level performance



Note: Panel A shows the mean daily minimum worker efficiency in each production line over workdays. Average minimum worker efficiency in a line over the sample period is 0.05 (given by dashed red line). Panel B shows the mean daily average worker efficiency in each line over workdays. Average worker efficiency in a line over the sample period is 0.30 (given by dashed red line).The number of working days for 37 production lines vary from 18 to 31 days. Production data obtained for September-October 2015 from factory records.

Figure A.4: Daily variation in line composition and performance (representative line)



Note: Panel A shows the observed line strength, average line strength (36 workers) and the absolute deviation of the line strength from the previous work day for a representative line. The allocated strength of this line is 54 workers the number of workers who report this line to be their allotted line. Panel B shows the corresponding changes in each caste share and the daily average worker efficiency in the same line. Data obtained for September-October 2015 from factory records and worker level primary survey.

Figure A.5: Caste composition and worker performance

Note: This figure shows worker level efficiency for 34,641 worker days. Worker efficiency = Daily output / Daily target output for each worker. Average efficiency per worker is 0.312. Proportion own Caste = Number of workers belonging to own caste category / Total number of workers in the line on a day; Linear fit depicted using the binscatter command in STATA dividing the data into 20 bins, plotting the mean X and Y values for each bin; 95% confidence interval in grey. The sample consists of 1744 workers in 37 assembly lines in two garment factories. Worker level production data obtained for September-October 2015 from factory records and caste data collected through a census survey of workers during August-October 2015.

APPENDIX B: Theoretical Framework

In a setting where worker effort is imperfectly observed, or, equivalently, is non-verifiable, firms face the usual moral hazard problem. In an assembly line if some workers are expected to put in low levels of effort then the whole line may be stuck in a bad equilibrium with low output. Since there are complementarities in production, team incentives seem intuitively the right solution though there are still free riding issues (see e.g. Itoh (1991), Che and Yoo (2001) for team incentives using peer sanctioning to encourage cooperation when there is moral hazard). In the factory, we did not observe any explicit team pay, however, individual incentives such as overtime pay did exist. Moreover promotions between different grades also act as individual incentives. Although incentives are individual, note that the supervisor is the one who decides on overtime and also to some extent on promotions. Since the supervisor cares about team output, the incentives are implicitly team incentives. Indeed, as long as the production function has complementarities, it is impossible for an individual worker to increase her own productivity if others do not cooperate. We therefore model (implicit) wage contracts based on joint output. We assume workers are risk neutral, and there is a minimum wage of $\underline{w}$ in the industry.

Formally, suppose there are two workers in the firm (the model is easily generalized to more workers) characterized by their observable ability types $\theta_i \in \{\bar{\theta}, \underline{\theta}\}$.[42] Output of worker i is increasing in θ and effort. For simplicity we assume the production function for worker i is given by $y_i = \theta_i + X$, where X is a random variable that takes one of the values $\{x_1, x_2\}$ with $x_1 > x_2$. The production function therefore has an individual component θ_i and a joint component, X which depends on the profile of efforts by the two workers. Average line output is the minimum of the y_i.

A line can have workers of different productivity: in our model we can have either both high ability, both low ability or one low and one high ability worker. Line output is highest when workers are high ability and put in high effort.

[42] Usually workers in an assembly line are of different grades, based on their efficiency levels.

Workers choose from two levels of effort $e_i \in \{h, l\}$ with $h > l$. The cost of effort is given by $c(e) = c$, if $e = h$ and $c(e) = 0$ if $e = l$. Below we focus only on the joint components part of the production function, since θ_i is fixed and does not change with incentives. We also assume first that $\theta_i \neq \theta_j$.

The probability of obtaining output level x_1 is denoted by α^{e_i,e_j}. If both workers choose $e_i = h$ the expected output is $\pi_{h,h} = \alpha^{hh}x_1 + (1-\alpha^{hh})x_2$. When effort levels are not equal then it is likely that expected output in this case depends on whether the high ability or the low ability worker is putting in high effort. This captures situations where the supervisor might ask a high ability worker to help a low productivity worker. Thus we assume that when $i \neq j$ then π_{e_i,e_j} depends also on the ability levels of workers i, j. In particular $\bar{\pi}_{h,l} = (\pi_{h,l}|\theta_i = \bar{\theta}, \theta_j = \underline{\theta}) > \underline{\pi}_{h,l} = (\pi_{h,l}|\theta_i = \underline{\theta}, \theta_j = \bar{\theta})$. Denote $\bar{\alpha}^{hl}$ ($\underline{\alpha}^{hl}$) as the probability of high output when the high (low) ability worker puts in high effort and the low (high) ability worker puts in low effort. Therefore, $\bar{\pi}_{h,l} = \bar{\alpha}^{hl}x_1 + (1 - \bar{\alpha}^{hl})x_2$ and $\underline{\pi}_{h,l} = \underline{\alpha}^{hl}x_1 + (1 - \underline{\alpha}^{hl})x_2$. Finally, if both workers choose low effort then expected output is $\pi_{l,l} = \alpha^{ll}x_1 + (1-\alpha^{ll})x_2$. Higher effort always increases output so $\pi_{h,h} > \bar{\pi}_{h,l} > \underline{\pi}_{h,l} > \pi_{l,l}$ which implies $\alpha^{hh} > \bar{\alpha}^{hl} > \underline{\alpha}^{hl} > \alpha^{ll}$ and complementarity in effort levels implies that $\pi_{h,h} - \bar{\pi}_{h,l} > \bar{\pi}_{h,l} - \pi_{l,l}$, and $\pi_{h,h} - \underline{\pi}_{h,l} > \underline{\pi}_{h,l} - \pi_{l,l}$. This implies $\alpha^{hh} - \bar{\alpha}^{h,l} > \bar{\alpha}^{h,l} - \alpha^{ll}$, and $\alpha^{hh} - \underline{\alpha}^{h,l} > \underline{\alpha}^{h,l} - \alpha^{ll}$.

First we show that if minimum worker's ability on the line ($\underline{\theta}$) is very low, then it may not be possible to induce high effort from the low ability worker in the absence of social networks. This is because, in order to induce the lowest ability workers to put in high effort, the wages have to be higher than the worker's contribution to output. Put another way the expected overtime and promotions needed to incentivize the worker is too costly relative to the gain in line output. Therefore the solution might be either that only high ability workers put in high effort or none of the types do.

B.1 Benchmark case without social networks

In this section we show the conditions under which the firm can induce high effort by workers when social networks are not present.

Let worker's utility function be:

$$u_i(e_i, e_j) = E(w|e_i, e_j) - c(e_i) \tag{B.1.1}$$

where $E(w|e_i, e_j)$ is the expected wage given the effort profile e_i, e_j. We can compute expected profits under three cases: (1) when the firm induces high effort from both workers, (2) when the firm induces high effort from only one worker and (3) when the firm does not induce high effort from any worker.

Case 1: The *per worker* expected profit of the firm (for a worker with ability θ), if it wants to induce high effort from both workers is, therefore, given by: $E(\pi|e_h, e_h) = \theta + \pi_{h,h} - (\alpha^{hh} w_1 + (1 - \alpha^{hh})w_2)$ The optimization problem is to choose w_1, w_2 to maximize (per worker expected profit)

$$\theta + E(\pi(e_h, e_h)) = \theta + \pi_{h,h} - \alpha^{hh} w_1 + (1 - \alpha^{hh})w_2 \tag{B.1.2}$$

subject to the participation constraints (PC), the incentive compatibility (IC) constraints and a limited liability (LL) constraint. Notice that θ appears on both sides of the inequality and therefore drops out in the conditions. Let $\bar{w}_1$ and $\bar{w}_2$ denote the wages for a high ability worker and $\underline{w}_1$ and $\underline{w}_2$ denote the wages for a low ability worker

(1) The PC is that a worker will only accept the implicit contract offering expected wages $E(w|h, h)$ if the cost of effort is low enough so that utility is higher than the outside option of minimum wages in another firm:

$$\alpha^{hh} \bar{w}_1 + (1 - \alpha^{hh})\bar{w}_2 - c \geq \underline{w} \tag{B.1.3}$$

and

$$\alpha^{hh} \underline{w}_1 + (1 - \alpha^{hh})\underline{w}_2 - c \geq \underline{w} \tag{B.1.4}$$

(2) The ICs are that, given complementarity, the firm must take account of the other

worker's effort in designing the incentive wages. Below we have conditions IC(1) and IC(2) that ensure (a) that high effort is a dominant strategy for a high ability worker i: IC(1) (given worker j puts in high effort):

$$\alpha^{hh}\bar{w}_1 + (1-\alpha^{hh})\bar{w}_2 - c \geq \bar{\alpha}^{lh}\bar{w}_1 + (1-\bar{\alpha}^{lh})\bar{w}_2 \tag{B.1.5}$$

and IC(2) (given worker j puts in low effort):

$$\bar{\alpha}^{hl}\bar{w}_1 + (1-\bar{\alpha}^{hl})\bar{w}_2 - c \geq \alpha^{ll}\bar{w}_1 + (1-\alpha^{ll})\bar{w}_2 \tag{B.1.6}$$

(b) Conditions IC'(1) and IC'(2) that ensure (a) that high effort is a dominant strategy for a low ability worker i: IC'(1) (given worker j puts in high effort):

$$\alpha^{hh}\underline{w}_1 + (1-\alpha^{hh})\underline{w}_2 - c \geq \underline{\alpha}^{lh}\underline{w}_1 + (1-\underline{\alpha}^{lh})\underline{w}_2 \tag{B.1.7}$$

and IC'(2) (given worker j puts in low effort):

$$\underline{\alpha}^{hl}\underline{w}_1 + (1-\underline{\alpha}^{hl})\underline{w}_2 - c \geq \alpha^{ll}\underline{w}_1 + (1-\alpha^{ll})\underline{w}_2 \tag{B.1.8}$$

and (3) the LL constraint: $\bar{w}_1, \bar{w}_2, \underline{w}_1, \underline{w}_2 \geq \underline{w}$.

Lemma 1 *The solution to the maximization problem (B.1.2) for the high ability worker is* $\bar{w}_1 = \underline{w} + \frac{c}{\bar{\alpha}^{hl}-\alpha^{ll}}$ *and* $\bar{w}_2 = \underline{w}$ *and for the low ability worker is* $\underline{w}_1 = \underline{w} + \frac{c}{\underline{\alpha}^{hl}-\alpha^{ll}}$ *and* $\underline{w}_2 = \underline{w}$

Proof

We show the result for the high ability worker: The IC constraints for the high ability worker can be re-written as:

$$(\alpha^{hh} - \bar{\alpha}^{lh})(\bar{w}_1 - \bar{w}_2) \geq c \tag{B.1.9}$$

and

$$(\bar{\alpha}^{hl} - \alpha^{ll})(\bar{w}_1 - \bar{w}_2) \geq c \tag{B.1.10}$$

Since $(\alpha^{hh} - \bar{\alpha}^{hl}) > (\bar{\alpha}^{lh} - \alpha^{ll})$, IC (B.1.10) $\implies$ IC(B.1.9). Moreover IC (B.1.10) $\implies$ $\bar{w}_1 > \bar{w}_2$. Let $\bar{w}_2 = w$ be the base wage and $\bar{w}_1 - \bar{w}_2 = b$, the bonus. Then we have the following solution $\bar{w}_1 = w + b = \underline{w} + \frac{c}{\bar{\alpha}^{hl} - \alpha^{ll}}$ and $\bar{w}_2 = \underline{w}$. This solution satisfies the PC.

The same logic implies the solution for $\underline{w}_1$ and $\underline{w}_2$. ■

Expected profits for the high ability worker are $= \bar{\theta} + \pi_{h,h} - \alpha^{hh}(\frac{c}{\bar{\alpha}^{hl} - \alpha^{ll}}) - \underline{w}$, and for the low ability worker are $= \underline{\theta} + \pi_{h,h} - \alpha^{hh}(\frac{c}{\underline{\alpha}^{hl} - \alpha^{ll}}) - \underline{w}$. It is easy to see that profits are lower for the low ability worker both because θ is lower but also because the compensation needed to induce high effort is higher. Thus if e.g. $\underline{\theta} + \pi_{h,h} < \alpha^{hh}(\frac{c}{\underline{\alpha}^{hl} - \alpha^{ll}}) + \underline{w}$, then a scheme to induce high effort in both workers is not profitable for the firm. It can however cap wages at the expected productivity of the low ability worker $\underline{\theta} + \pi_{h,h}$. At this wage rate the low ability worker strictly prefers to put in low effort. Therefore the contractual wages shift to a different regime as Lemma 2 below shows.

Lemma 2 *If the firm induces high effort from the high ability worker and low effort from the low ability worker then wages of the high ability worker are given by:* $\tilde{w}_2 = \underline{w}, \tilde{w}_1 = \underline{w} + \frac{c}{(\bar{\alpha}^{hl} - \alpha^{ll})}$. *The low ability worker gets* $\underline{w}$.

Proof:

The problem for the high ability worker is to choose w_1, w_2 to maximize:

$$E(\pi(e_h, e_l)) = \bar{\theta} + \bar{\pi}_{h,l} - \bar{\alpha}^{hl}\tilde{w}_1 + (1 - \bar{\alpha}^{hl})\tilde{w}_2 \tag{B.1.11}$$

subject to:

(1) the PC:

$$\bar{\alpha}^{hl}\tilde{w}_1 + (1 - \bar{\alpha}^{hl})\tilde{w}_2 - c \geq \underline{w} \tag{B.1.12}$$

which can be re-written as:

$$\bar{\alpha}^{hl}(\tilde{w}_1\tilde{w}_2) + \tilde{w}_2 - c \geq \underline{w} \tag{B.1.13}$$

(2) The IC which can be re-written as:

$$(\bar{\alpha}^{hl} - \alpha^{ll})(\tilde{w}_1 - \tilde{w}_2) \geq c \tag{B.1.14}$$

and (3) the LL constraint: $\tilde{w}_1, \tilde{w}_2 \geq \underline{w}$

The proof follows the same logic as the proof of Lemma (1). By the same logic, $\tilde{w}_2 = \underline{w}, \tilde{w}_1 = \underline{w} + \frac{c}{(\bar{\alpha}^{hl} - \alpha^{ll})}$. ■

Expected profits are positive iff $\bar{\theta} + \bar{\pi}_{h,l} - \bar{\alpha}^{hl}(\frac{c}{\bar{\alpha}^{hl} - \alpha^{ll}}) - \underline{w} \geq 0$.

A third option for the firm is to simply not induce high effort in both workers and pay minimum wages to both workers. In this case profits are positive iff $\underline{\theta} + \pi_{ll} - \underline{w} \geq 0$.

Proposition 1 *Suppose $\theta_i \neq \theta_j$. Assume that the minimum ability in the line satisfies $\underline{\theta} \geq T_1 \equiv \alpha^{hh}(\frac{c}{\underline{\alpha}^{hl} - \alpha^{ll}}) + \underline{w} - \pi_{h,h}$, then both workers put in high effort and (average) line output is $\underline{\theta} + \pi_{h,h}$. If $\alpha^{hh}(\frac{c}{\bar{\alpha}^{hl} - \alpha^{ll}}) + \underline{w} - \pi_{h,h} \equiv T_2 \leq \underline{\theta} < T_1$ then only the high ability worker puts in high effort, line output is $\underline{\theta} + \bar{\pi}_{h,l}$. Finally, if $\underline{\theta} < T_2$, then both workers put in low effort and line output is $\underline{\theta} + \pi_{l,l}$.*

Proposition 2 *Assume $\theta_i = \theta_j = \bar{\theta}$, then both high ability workers put in high effort iff $\bar{\theta} \geq T_2$. Assume $\theta_i = \theta_j = \underline{\theta}$, then both low ability workers put in high effort iff $\underline{\theta} \geq T_1$*

The proof is obvious and follows from lemmas above.

Suppose that $\underline{\theta} \geq T_1$, while $\bar{\theta} \geq T_2$ then in any symmetric equilibrium, the line output is $\bar{\theta} + \pi_{h,h}$ when $\theta_i = \theta_j = \bar{\theta}$ and is $\underline{\theta} + \pi_{l,l}$ when $\theta_i = \theta_j = \underline{\theta}$.

B.2 With social networks

There is an exogenous probability of separation from the firm $1-\gamma(\theta)$, which is higher for low ability workers, $\gamma(\underline{\theta}) < \gamma(\bar{\theta})$, as chances of being fired are higher even for the same effort levels. W.l.o.g we assume that $\gamma(\bar{\theta})$ is normalised to 1, so that $\gamma(\underline{\theta}) = \gamma \in (0,1)$. Separated workers rely on their social networks, in particular on more experienced workers for getting other jobs via referrals or for helping over a financially difficult period. We denote the utility from the network as $V(f_i^k|e_i)$ where f_i^k is the number (or fraction) of co-workers in the social network of worker i of caste k in the line. When monitoring[43]/mentoring is feasible then V can be conditioned on effort of worker i. The higher the number of co-workers from one's social network, the higher is V, because co-workers of the same network are likely to observe worker i if called out for holding up the line by supervisor, live close to worker i and have links with other network members who can help/ostracize the worker, and may themselves not provide referrals to the worker in future. The larger the strength of the network on the line or in the factory, the better is information on worker i and its transmission to others in the network both inside and outside the line/factory. We assume that the share of other out group or out of network peers in the line does not affect workers. The key point in the formal model is to introduce an extra term in the incentive and participation constraints for low ability workers that depends on γ and $V(\cdot)$. For ease of exposition we focus on the low ability worker only, as that is the binding constraint on line output. Let wages for the low ability worker be denoted as $\hat{w}_1, \hat{w}_2$.

The utility function for worker i in network k is:

$$u_i(e_i, e_i)_i^k = \gamma(E(w|e_h, e_h) - c(e_i)) + (1-\gamma)V(f_i^k|e_i) \tag{B.2.1}$$

Assume that monitoring or mentoring by high ability workers is profitable, i.e. $\alpha^{hh}\bar{w}_1 - \underline{m} > \bar{\alpha}^{hl}\tilde{w}_1$, high ability workers benefit from monitoring/ helping/mentoring low ability workers as they get a higher expected wage when line output is higher. In this

[43] We use "monitoring" loosely to refer to enforcement. Supervisors also observe low performing workers but are not able to enforce high effort in some cases using financial incentives.

case, $V(f_i^k|e)$ depends on the effort level of worker i and $V(f_i^k|e_l) = \underline{V} < V(f_i^k|e_h)$. We can re-write the constraints for low ability worker as follows: (1) the PC:

$$\alpha^{hh}\hat{w}_1 + (1-\alpha^{hh})\hat{w}_2 \geq c + \underline{w} - \frac{(1-\gamma)}{\gamma}(V(f_i^k|e_h) - \underline{V}) \tag{B.2.2}$$

(2) The ICs

$$\gamma(\alpha^{hh}\hat{w}_1 + (1-\alpha^{hh})\hat{w}_2 - c) + (1-\gamma)V(f_i^k|e_h) \geq \gamma(\underline{\alpha}^{lh}\hat{w}_1 + (1-\underline{\alpha}^{lh})\hat{w}_2) + (1-\gamma)\underline{V} \tag{B.2.3}$$

which can be re-written as:

$$(\alpha^{hh} - \underline{\alpha}^{lh})(\hat{w}_1 - \hat{w}_2) \geq c - \frac{1-\gamma}{\gamma}(V(f_i^k|e_h) - \underline{V}) \tag{B.2.4}$$

and

$$\gamma(\underline{\alpha}^{hl}\hat{w}_1 + (1-\underline{\alpha}^{hl})\hat{w}_2 - c) + (1-\gamma)V(f_i^k|e_h) \geq \gamma(\alpha^{ll}\hat{w}_1 + (1-\alpha^{ll})\hat{w}_2) + (1-\gamma)\underline{V} \tag{B.2.5}$$

which can be re-written as:

$$(\underline{\alpha}^{hl} - \alpha^{ll})(\hat{w}_1 - \hat{w}_2) \geq c - \frac{1-\gamma}{\gamma}(V(f_i^k|e_h) - \underline{V}) \tag{B.2.6}$$

and (3) the LL constraint: $\hat{w}_1, \hat{w}_2 \geq \underline{w}$

Denote $\frac{1-\gamma}{\gamma}(V(f_i^k|e_h) - \underline{V}) = K$. Using the same proof as in Lemma 1 and 2, the wages to induce high effort from worker 2 satisfy: $\hat{w}_2 = \underline{w}, \hat{w}_1 \geq \underline{w} + \frac{c-K}{(\underline{\alpha}^{hl} - \alpha^{ll})}$. The high ability worker gets as above, $\bar{w}_1, \bar{w}_2$.

The implication of monitoring or mentoring of low ability workers and being able to condition network benefits on effort is that low ability workers will have a higher chance of putting in high effort even with a lower monetary payoff from the firm. High ability workers gain from monitoring/mentoring when the costs of doing so are low compared to the higher probability of getting overtime or being promoted. Low ability workers also get a higher monetary payoff if the firm finds it profitable i.e. $\underline{\theta} \geq \alpha^{hh}\hat{w}_1 + \underline{w} - \pi_{h,h}$.

Let $\alpha^{hh}(\frac{K}{\underline{\alpha}^{hl}-\alpha^{ll}}) \equiv k_1$ and $\alpha^{hh}(\frac{K}{\bar{\alpha}^{hl}-\alpha^{ll}}) \equiv k_2$. Then $\alpha^{hh}(\frac{c-K}{\underline{\alpha}^{hl}-\alpha^{ll}}) + \underline{w} - \pi_{h,h} = T_1 - \alpha^{hh}(\frac{K}{\underline{\alpha}^{hl}-\alpha^{ll}}) = T_1 - k_1$ and $\alpha^{hh}(\frac{c-K}{\bar{\alpha}^{hl}-\alpha^{ll}}) + \underline{w} - \pi_{h,h} = T_2 - \alpha^{hh}(\frac{K}{\bar{\alpha}^{hl}-\alpha^{ll}}) = T_2 - k_2$.

Proposition 3 *Assume that $\theta_i \neq \theta_j$ and the minimum ability in the line satisfies $\underline{\theta} \geq T_1 - k_1$, then both workers put in high effort and line output is $\underline{\theta} + \pi_{h,h}$. If $T_2 - k_2 \leq \underline{\theta} < T_1 - k_1$, then only the high ability worker puts in high effort and line output is $\underline{\theta} + \bar{\pi}_{h,l}$. Finally, if $\underline{\theta} < T_2 - k_2$, then both workers put in low effort and line output is $\underline{\theta} + \pi_{l,l}$*

Let $\underline{m}$ be the cost of monitoring or mentoring within a network. The high ability worker will be willing to monitor the low ability worker within the same network if $\alpha^{hh}\bar{w}_1 - \underline{m} > \bar{\alpha}^{hl}\tilde{w}_1$. In the presence of social networks, the low ability worker has lower expected wages than the high ability worker though they both put in high effort. This corresponds to higher probability of overtime and or promotions for higher ability workers in these lines. The bigger the difference $\alpha^{hh} - \bar{\alpha}^{hl}$ the stronger are the effects of monitoring or mentoring.

Note that within a network in a line the composition may be one of three types: Both high ability, both low ability or one low and one high ability worker. Assuming that $\bar{\theta} \geq T_1$. Increasing the share of own caste in the first type will have no effect on average output- social networks have no effect when the ability level is high enough that individual wage incentives are profitable for the firm. When both workers have ability below the threshold $\underline{\theta} < T_2$ then again, there are no incentives to monitor the low ability workers and there is no change in individual output. This is because for the mechanism to work, there must be some workers whose productivity is above the threshold where the firm finds it profitable to increase expected wages. Our predictions are therefore: Assume that there is sufficient heterogeneity in worker productivity within a caste group in a line and minimum productivity in the caste group is low, $\underline{\theta} < T_1$. When the share of own network workers in a line increases, the average output of the own caste group increases. This average increase is driven by a larger increase in the productivity of low ability workers.[44] This kind of monitoring and reward/punishment schemes depends on the presence of long term repeated relationships which arise when workers are in the same networks.

Summarizing, our main results are: (1) as the proportion of own network workers increases, the average productivity within the same network increases but this is driven mainly by an increase in the effort of low ability workers within the network, (2) these effects are stronger when there are either high ability workers in the line or there are more experienced workers with lower costs of monitoring or mentoring, who potentially stand to gain a lot in terms of expected wages when joint output increases. Therefore social networks can help improve line output and individual output in lines where the minimum efficiency is very low to begin with, and where the number of same caste-residence network is higher. Moreover it requires the presence of different levels of productivity in the line.

[44] When $T_1 > \underline{\theta} \geq T_2$ then only low ability workers increase effort. If $\underline{\theta} < T_2$, either only low ability worker increases effort (due to our assumption that high ability workers gain more from monitoring/mentoring) or both do. Thus overall the low ability worker increases effort for a larger range of parameters.