I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13711

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

Farzana Afridi
Amrita Dhillon
Swati Sharma

SEPTEMBER 2020

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13711

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

Farzana Afridi
Indian Statistical Institute and IZA

Amrita Dhillon
King's College, London, CAGE and CEPR

Swati Sharma
Indian Statistical Institute

SEPTEMBER 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

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 but wages are paid at the individual level. Using exogenous variation in production line composition for almost 35,000 worker-days, we find that a 1 percentage point increase in the share of own caste workers in the line increases daily individual productivity by about 10 percentage points. The lowest performing worker increases her effort by more than 15 percentage points when the production line has a more homogeneous caste composition. Production externalities that impose financial costs due to worker's poor performance on co-workers within her social network can explain our findings. Our results suggest that even in the absence of explicit group-based financial incentives, social networks can be leveraged to improve both worker and group productivity.

JEL Classification: Y40, Z13, J15, J24

Keywords: caste, social networks, labor productivity, assembly lines, India

Corresponding author:
Farzana Afridi
Indian Statistical Institute
Economics and Planning Unit
7, S.J.S. Sansanwal Marg
New Delhi 110016
India
E-mail: fafridi@isid.ac.in

* This paper has benefited from suggestions by Ashwini Deshpande, James Fenske, Karla Hoff, Seema Jayachandran, Kaivan Munshi, Anant Nyshadham, Atonu Rabbani, Chris Woodruff and workshop/conference participants at the ISI, Ashoka University, IIM Kolkata, Delhi School of Economics, and the University of Edinburgh. Nikhil Bharadwaj and Karmini Sharma provided exceptional research assistance. 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.

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 can vary by almost 30 percentage points 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 90 percentage points 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. The determinants of coordination amongst workers in large assembly lines within firms has not been explored in the literature. We attempt to fill this gap by analysing the role of workers' caste-based 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 comple-

[1] In an ongoing project on garment productivity (`https://www.qeh.ox.ac.uk/content/readymade-garment-productivity-project`), Macchivello, Menzel, Rabbani and Woodruff find 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.

mentarities between workers generate externalities in the production process and the total output of the team is determined by the minimum individual output, the worker composition of these teams can play a significant role in determining both group and firm output. Using 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 identification strategy relies on exogenous variation in the daily worker composition of production lines due to unanticipated worker absenteeism to estimate the causal impact of the proportion of own-caste workers in a production line on individual and line productivity.

Our findings suggest that a 1 percentage point increase in the strength of the workers' social network - the proportion of workers belonging to own caste - in the line on a day, raises workers' own productivity by more than 10 percentage points. 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 over 15 percentage points while the average line performance improves by more than 23 percentage points when the caste composition of the line becomes more homogeneous. These results are driven by assembly lines as opposed to non-assembly production lines where workers are substitutes for each other. Our findings are robust to a host of sensitivity checks, including worker ability, line specific unobservables and seasonal trends in production in the industry and at the line level.

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 rate). Workers who are more productive have a higher probability of obtaining the limited overtime positions and also of being promoted to higher grades due to recommendations by the line supervisor. Since the line supervisor cares about the line output, there exist implicit individual financial incentives linked to higher team production. Thus higher productivity workers have strong incentives to monitor poor performance and enforce higher effort from workers who are holding up line output.

Our results suggest that this monitoring is more effective when workers belong to the same social networks. Hence 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 to ensure getting network benefits. Our findings can therefore be explained by the social incentives that workers face when their network strength is higher in their production line on a work day. 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, our worker level data suggest economic interdependence and benefits from one's caste-based networks as sources of information for job openings as well as for referrals. 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. Not surprisingly, our results are driven by workers whose job referee is still employed in the factory, validating the claim that possible exclusion from one's social network is a likely mechanism for improved efficiency of same caste workers.

Our accompanying theoretical analysis, therefore, underlines the role of social networks in improving worker productivity in highly competitive product markets, such as the garment industry, where profit maximizing firms are constrained in offering employees explicit monetary incentives.[2] Instead, in such industries firms can leverage social networks amongst workers to relax their constraints on worker compensation, as the insights from the microfinance literature and it's applications in labor economics have shown in different contexts (Hal (1990), Ghatak and Guinnane (1999), Bryan et al. (2015)), Heath (2018), Dhillon et al. (2019)).

Existing research 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)). Peer effects on productivity, mediated through social networks that create pressures to conform to a social norm, however, are ambiguous (Bandiera et al. (2010)).[3]

Identity motivations may also impact worker performance. A large literature on lab experiments suggests that team homogeneity leads to more efficient outcomes (Eckel and Grossman (2005), Goette et al. (2006), Charness et al. (2007), Chen and

[2] `https://www.mckinsey.com/business-functions/sustainability/our-insights/style-thats-sustainable-a-new-fast-fashion-formula`; Chang et al. (2016)

[3] 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.

Chen (2011)). 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)).[4]

While almost all of the above research focuses on workers as substitutes in the production process, workers' own productivity may not be influenced by co-worker performance either through a desire to conform to a social norm (e.g. peer pressure or local average network effect) or through strategic complementarities (e.g. knowledge spillovers or local aggregate network effects) when workers are complementary in the production process and observability of effort is imperfect as in the production lines in garment manufacturing. 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. Our research, thus, extends the broader literature on the role of social networks in job search to its impact on worker and firm productivity.

These findings speak to multiple strands of literature on worker incentives as well as to the existing research on management practices and firm behavior. We identify 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. 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.

[4]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.

It suggests that social connections amongst workers can incentivize them to be more productive even in the absence of monetary benefits for improving individual or group productivity. The results of our analysis indicate that identifying workers who are widely connected to co-workers through job referrals or residential location could carry implications for productivity through the optimal design of production schedules and composition of teams in the firm.

The remainder of the paper is organized as follows. Section (2) describes the background of our study, including the production process and worker incentives in garment factories. Section (3) summarizes the observed data regularities. Section (4) provides the theoretical framework. We discuss our empirical methodology, report the results of our analysis in Section (5) and conduct robustness checks in Section (6). We underscore the mechanism that explains our findings in Section (7) and conclude in Section (8).

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 highlights the salience of social networks based on caste and homophily in India (Munshi (2019)).[5] 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)).

[5]Caste, a unique feature of Indian society, is inherited at birth. The caste system classifies Hindu society into four hierarchical occupational groups or *varnas* - *Brahmins* (priests and scholars), *Kshatriyas* (warriors and rulers), *Vaishyas* (merchant class), and *Shudras* (cultivators). Those engaged in menial tasks, such as scavenging, are considered to be outside the varna system and untouchable.

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.[6]

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 workers obtained information about job openings through their network. Conditional on the informant being from the same factory as our survey respondent, 42% of workers were referred to the management by the informant and 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).

While our data suggest that the 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.[7] 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 residence information being available for a worker in our data). Thus, own-caste neighborhoods represent the social networks that workers derive economic benefits from.

[6]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)).

[7]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.

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 are placed one behind the other (see Figure A1 in Appendix A). Besides the machine operator who is responsible for stitching, the production line also consists of helpers (to fold, cut, match or iron different parts of garments) who assist operators. Henceforth, we will use the term 'worker' to denote operators and helpers who contribute to stitching of the garment. Each line is assigned a particular style of garment to be produced over a day or several days until the production target for that garment-style is met.

There are two types of production lines: assembly and non-assembly lines. 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 production process is, thus, simultaneous and complementary. The stitched garment is then assembled at the front of the line.[8] Hence there exist strong production externalities in the assembly line - the total number of finished garments produced by the line on a day would depend on the productivity of the least efficient worker.[9]

Observability of co-worker effort is imperfect due to differences in operations performed by workers in an assembly line. However, as can be seen from Figure A1, workers can see who is sitting in their line even though they cannot directly observe each other's output. Moreover, workers would be aware of where production bottle-

[8]Figure A2, Appendix A, illustrates the general production process for a shirt in an assembly line, for instance. While some workers perform different operations on collars (e.g. stitching, hemming), other workers may be responsible for operations on sleeves (e.g. attaching cuffs, stitching armholes) and so on.

[9]Our claim is validated by a significant, positive correlation between the line level output recorded by the factory management and the output of the least efficient worker in that line in our data.

necks exist. 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 financial incentives - bonus and promotions - are hence linked to his line's performance, as per our discussion with the factory management. Although workers receive a fixed, minimum wage paid as a monthly salary, there are different grades of workers classified according to skill measured through a performance test on entry and based on past experience and training they have received.[10] The wage differential between grades is about 10-12%. During the period of our study workers were not offered any performance linked bonuses.

Supervisors allot limited overtime positions to workers, which typically pay an hourly wage higher than minimum wages. Workers total earnings, therefore, depend on their fixed grade pay and overtime wages. Since overtime positions are few, more productive workers have a higher probability of receiving over time work. They also have a greater chance of being promoted to higher grades. The management maintains records of operational efficiency for each operation (but not worker), so the supervisor would know which operations are holding up the line output. Although workers are unlikely to be punished due to limited liability (minimum wage) constraints, the supervisor would likely know who is the weak link in the line. In essence, therefore, there exist implicit individual financial incentives linked to being a more productive worker. Given the production externalities in the assembly line, the performance of co-workers in an assembly line can impact the earnings of a worker.

[10] In our sampled factories, supervisors were also receiving a fixed monthly salary which was higher than the workers' salary. However, their promotions and salary increments within the factory were contingent on performance.

Our identification strategy, discussed in detail later, relies on unanticipated worker absenteeism leading to arguably exogenous changes in the daily composition of production lines. Given the constrained supply of skilled workers and the high proportion of migrant laborers in this industry, worker attrition and absenteeism is significant (GOI (2018)).[11] 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. Although most of this variation in manpower can be on account of changes in production targets, it does not account fully for daily variation. While supervisors may reassign workers within their lines, workers can also be moved across lines to address attrition and absenteeism to meet production targets. Any reassignment of workers across the lines is controlled by floor or line in-charge according to the supply and demand of workers, the relevant skill requirement and production deadlines.[12] Thus, the daily composition of a line can vary both due to worker absenteeism as well as any worker reallocation thereof. We discuss this in more detail in the following section.

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, the latter manufactures garments for the domestic market. 89% of our sample of workers belong to the exporting firm which was significantly larger. We construct our dataset from two

[11] 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.

[12] Adhvaryu et al. (2019) document the virtual absence of relational trading between supervisors inside garment factories to reallocate workers in order to address worker absenteeism.

main 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 in August - October 2015 (approximately 61 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.[13] The survey gathered information on individual characteristics, including native state of residence and caste, years of experience in the garment industry, and the process of obtaining the current job particularly referrals. We also conducted a shorter survey of supervisor characteristics.

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. Note that we view broad caste categories as suitable proxy for networks - relevant for residential decisions (e.g. areas are often classified as *harijan* or low caste) or in fostering shared experiences. Narrow caste categories, viz. *jati*, on the other hand, represent identity concerns, which is not the focus of this paper.

3.2 Worker productivity and attendance data

Since the factory managements were recording line level productivity only by operation, we designed a protocol for collecting hourly, worker level output, and line composition that mapped workers to an operation within each line. These data were

[13] Since worker attrition is high in this sector, we kept in touch with the Human Resource (HR) department to ensure that any new worker recruited during our study period was included in our survey.

obtained for a period of 31 working days between September-October 2015, a subset of the 61 days during which the worker census was conducted.[14] 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. This is calculated using the SAM (standard allowable minutes) based on a standardized global database that includes information on the universe of garment-styles.[15] 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.[16] Each worker's efficiency, therefore, is measured as follows:

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

We measure 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 since the lowest effort determines the total output (or units of complete garment) in the assembly line. Data on workers' and supervisors' daily attendance was obtained from the Human Resource (HR) departments of the two factories.[17] We match workers

[14]Every production line has a 'feeder' whose notes down productivity by operation in a line each hour. Using our data collection protocol, the 'feeder' recorded the name and unique ID of the worker at each operation in the line. This allowed us to obtain disaggregated worker level output, and also follow workers across lines over the 31 day period.

[15]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 can calculate the target output per worker per style operation. Note that worker productivity is *not* affected by downstream workers because the production process is simultaneous, not sequential.

[16]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.

[17]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.

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.[18]

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. In contrast to the pervasiveness of job network of workers, on average, a worker reports having less than 2 friends in the factory.[19] The same 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 categories(22%), in our sample. The characteristics of workers are largely similar across caste categories - in particular we find no evidence of systematic productivity differences between workers of different caste groups .[20]

Table 2, Panel A, shows the average efficiency of a worker and across worker-days on the stitching floor. Workers typically achieve only around 31% of their target

[18] 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 A1 in the Appendix for details.

[19] Majority of supervisors were from M category unlike workers who were more likely to belong to H category. Almost 23% of workers belong to the same caste category as their line supervisor. We do not find any impact of caste alignment of supervisor and worker on latter's productivity.

[20] 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.

output, on average. Worker efficiency is not statistically significantly different across caste categories. Panel B shows the performance of a line across the sampled period. The average efficiency of a line is about 30% and the average minimum efficiency of line is just over 5%, indicating that least performing worker is meeting only 5% of the target output. We find similar productivity statistics by line-days. Figure 1 exhibits the variation in the line performance cross-sectionally, averaged across work days, in terms of minimum efficiency (left panel) and average efficiency (right panel). While the mean minimum efficiency of a line varies from 2% to over 15%, the average efficiency, though higher, exhibits greater variance (16 - 44%).

The variation in performance across production lines is accompanied by wide variation in both the strength of a line (Figure 2a) and its performance across workdays (Figure 2b). Figure 2a shows the number of workers in a representative line and the day-to-day variation in its strength. The absolute deviation of the observed strength from average strength of the line is between 0 - 39% during our sample period. The average absolute deviation in line strength from the previous day is about 16%. Note that the daily changes in the number of workers in line underestimates changes in line composition since workers are also reallocated across lines.

Figure 2b traces the average efficiency of a line across workdays, which can be seen to vary by more than 25 percentage points. Thus average performance of a line may hide much higher variation in performance across workdays within the same line. The proportion of L, M and H category workers in the line as shown in Figure 2b varies along with changes in line strength and efficiency. The proportion of H caste workers in a line across work days can vary by up to 22 percentage points, 12 and 18 percentage points for the M and L caste categories, respectively.[21] As discussed in

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

the next section, neither worker productivity nor absenteeism rates differ significantly across caste groups in our sample.[22]

We correlate the caste composition of the assembly line, worker and line level productivity in Figure 3 to show that the higher the proportion of own caste workers in the line (Figure 3a) and the more homogeneous the caste composition of the line on a work day (Figure 3b), the higher the efficiency of the worker and the minimum efficiency of the line on that day. This suggests that social networks amongst co-workers, mediated through caste, may have a significant impact on individual and group productivity.

4 Theoretical Framework

The above discussions highlight the fact that when worker effort is imperfectly observed, 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. We build on the insights from the microfinance literature (Hal (1990), Ghatak and Guinnane (1999), Bryan et al. (2015)) and applications in labor economics (Heath (2018), Dhillon et al. (2019), Pallais and Sands (2016)) to theoretically demonstrate how social networks can solve moral hazard/adverse selection problems when formal institutions cannot, in a context where workers are complementary in the production process.

Simply put, when workers have to be paid minimum wages, it creates a limited liability constraint for firms, which in turn implies that to motivate workers the rewards for high effort have to be correspondingly higher. When there is a high degree of complementarity in the production function the firm gains more from inducing greater effort from all workers as this leads to disproportionately larger expected

[22] Since workers in our study come from approximately 300 districts across 16 states, the likelihood of workers of same *jati* sitting in a particular line on a day is negligible.

output than from inducing only a few workers to put in high effort. But since the minimum wage constraints push up the cost of performance based pay, the firm instead may decide to go in for lower powered incentives or no incentives at all.[23] In our context, by aligning the incentives of the high ability line supervisors to the line output, the management creates implicit team incentives for workers not only to put in more effort themselves but also to induce other co-workers to put in higher effort. Thus when a production team is large, workers' social networks can be leveraged to provide network based rewards and punishments to support the firm's own implicit incentives.

Formally, suppose there are two workers in the firm (the model is easily generalized to more workers) characterized by their observable ability types $\theta_i \in \{\bar{\theta}, \underline{\theta}\}$.[24] Output of worker i is increasing in θ and effort. For simplicity we assume the production function is given by $y_i = \theta + X$, where X is a random variable that takes one of the values $\{x_1, x_2\}$ with $x_1 > x_2$. Workers choose from two levels of effort $e_i \in \{h, l\}$ with $h > l$. Low effort has zero cost while high effort costs c. 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 $\pi_{h,h} = \alpha^{hh}x_1 + (1 - \alpha^{hh})x_2$. If only one worker chooses high effort the expected output is $\pi_{h,l} = \alpha^{hl}x_1 + (1 - \alpha^{hl})x_2$. 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. 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 $(\pi_{h,l}|\theta_i = \bar{\theta}, \theta_j = \underline{\theta}) > (\pi_{h,l}|\theta_i = \underline{\theta}, \theta_j = \bar{\theta})$. 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 $\pi_{h,h} > \pi_{h,l} > \pi_{l,l}$ and complementarity in effort levels

[23] Due to stiff product market competition in the garment industry there is also an upper bound on product prices (given by a zero profit condition) so that performance based wages cannot be recouped if worker ability is too low.

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

implies that $\pi_{h,h} - \pi_{h,l} > \pi_{h,l} - \pi_{l,l}$. Thus α^{e_i,e_j} must satisfy: $\alpha^{hh} > \alpha^{h,l} > \alpha^{ll}$ and $\alpha^{hh} - \alpha^{h,l} > \alpha^{h,l} - \alpha^{ll}$.

Since effort is imperfectly observed or, equivalently, is non-verifiable, the firm faces moral hazard. To induce workers to work harder the firm can offer incentive compatible contracts, such that wages are conditioned on individual output - w_1, w_2. Firms can commit to their wage contracts and there is a minimum wage of $\underline{w}$ in the industry. Workers are risk neutral.

4.1 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:

$$u_i(e_i, e_j) = E(w|e_i, e_j) - c \tag{1}$$

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. Details are in Appendix B. Below, we assume (w.l.o.g) that when the firm induces the same level of effort in each ability type of worker, it pays the same wages.

Case 1: The per worker expected profit of the firm 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)

$$\theta + E(\pi(e_h, e_h)) = \pi_{h,h} - \alpha^{hh} w_1 + (1 - \alpha^{hh}) w_2 \tag{2}$$

subject to the participation constraints (PC), the incentive compatibility (IC) constraints and a limited liability (LL) constraint.

(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:

$$\alpha^{hh}w_1 + (1-\alpha^{hh})w_2 - c \geq \underline{w} \tag{3}$$

(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 that high effort is a dominant strategy for worker i: IC(1) (given worker j puts in high effort):

$$\alpha^{hh}w_1 + (1-\alpha^{hh})w_2 - c \geq \alpha^{lh}w_1 + (1-\alpha^{lh})w_2 \tag{4}$$

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

$$\alpha^{hl}w_1 + (1-\alpha^{hl})w_2 - c \geq \alpha^{ll}w_1 + (1-\alpha^{ll})w_2 \tag{5}$$

and (3) the LL constraint: $w_1, w_2, w_3 \geq \underline{w}$. Denote average ability as $\mu = \frac{\theta+\bar{\theta}}{2}$. Using the solution to this problem (see Appendix B), expected profits per worker are: $E(\pi(e_h, e_h)) = \mu + \pi_{h,h} - \alpha^{hh}(\underline{w} + \frac{c}{\alpha^{hl}-\alpha^{ll}} - (1-\alpha^{hh})\underline{w}$.

Case 2: Alternately, the firm can induce high effort only from one worker. Since ability is assumed to be observable, the firm would find it profitable to pay higher wages to induce high effort from the high ability worker and induce low effort (and pay minimum wages) from the low ability worker (given our assumption that $\pi_{h,l}$ is higher when the high ability worker puts in high effort than when the low ability worker puts in high effort). The maximization problem has the same structure as (2).

Expected profits per worker are now $\mu + \pi_{h,l} - \frac{\alpha^{hl}}{2}(\frac{c}{\alpha^{hl}-\alpha^{ll}}) - \underline{w}$ (see Appendix B for details).

Case 3: 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 per worker are $\mu + \pi_{ll} - \underline{w}$.

What effort profile will the firm induce out of cases (1)-(3)? Let $T_1 \equiv \frac{2\alpha^{hh}-\alpha^{hl}}{2(\alpha^{hh}-\alpha^{hl})}\frac{c}{\alpha^{hl}-\alpha^{ll}}$ and $T_2 \equiv \frac{\alpha^{hh}}{\alpha^{hh}-\alpha^{ll}}\frac{c}{\alpha^{hl}-\alpha^{ll}}$. The firm induces high effort from both workers iff expected profits are higher in case (1) as compared to both cases (2) and (3). Expected profits in case (1) are higher than expected profits in cases (2) and (3) iff $x_1 - x_2 \geq max(T_1, T_2)$. Intuitively, the firm will induce high effort in both workers only if the marginal gains from doing so for each worker, $x_1 - x_2$, are higher than the marginal cost or higher expected wages that have to be paid, which is $\max(T_1, T_2)$, depending on which of the other options is more profitable.

The key point is that, in the absence of benefits from social networks, both types of workers get higher expected wages when the firm induces high effort than when the firm induces high effort in only one worker or does not induce high effort at all.

4.2 With social networks

Social networks can be leveraged to provide monitoring or social collateral when team incentives are involved. Thus, networks can help to reduce the wages that must be paid by the firm to workers to reward them for higher effort, increasing the profitability of inducing high effort.[25]

Assume that the per worker costs of enforcing contracts using collective rewards and punishments by the network are sufficiently small. There is an exogenous probability of separation from the firm $1 - \gamma$. Separated workers rely on their social

[25]Note that mentoring with rewards for cooperation is equivalent to monitoring with punishment for non-cooperation in the model.

networks 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 of coworkers in the social network of worker i of caste k. V can be conditioned on effort of worker i (in our setting, low output workers who are holding up line output are often called out by the supervisor- this observability is all that is needed for the model). 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 the better is information transmission on worker i to others in the network but outside the team. Suppose the firm wishes to induce high effort in both workers. The utility function with networks is:

$$u_i(e_i, e_i)_i^k = \gamma(E(w|e_h, e_h) - c(\theta)) + (1-\gamma)V(f_i^k|e_i) \tag{6}$$

Note that $V(f_i^k|l) = \underline{V} < V(f_i^k|h)$. We can re-write the constraints for the maximization problem of the firm, (2) as follows:

(1) the PCs:

$$\gamma(E(w|e_h, e_h) - c) + (1-\gamma)V(f_i^k|h) \geq \gamma\underline{w} + (1-\gamma)\underline{V} \tag{7}$$

(2) The ICs:

$$\gamma(E(w|h, h) - c) + (1-\gamma)V(f_i^k|h) \geq \gamma(E(w|l, h)) + (1-\gamma)\underline{V} \tag{8}$$

and

$$\gamma(E(w|h,l)-c)+(1-\gamma)V(f_i^k|h)\geq\gamma(E(w|l,l))+(1-\gamma)\underline{V} \tag{9}$$

and (3) the LL constraint: $w_1, w_2, w_3 \geq \underline{w}$

Denote $\frac{1-\gamma}{\gamma}(V(f_i^k|h)-\underline{V})=K$. Suppose the firm wants to induce low effort by both workers. There are no incentive constraints. Since $V(f_i^k|l)=\underline{V}$ the wages that satisfy the participation constraint are $w_1=w_2=\underline{w}$. Below we assume $c>K$ to ensure that the bonus for high effort is positive.

Let $\tilde{T}_1 \equiv \frac{2\alpha^{hh}-\alpha^{hl}}{2(\alpha^{hh}-\alpha^{hl})}\frac{c-K}{\alpha^{hl}-\alpha^{ll}}$ and $\tilde{T}_2 \equiv \frac{\alpha^{hh}}{\alpha^{hh}-\alpha^{ll}}\frac{c-K}{\alpha^{hl}-\alpha^{ll}}$. In the analysis without social networks, we saw that if $x_1-x_2<\max(T_1,T_2)$ then the firm would not induce high effort in both workers (Proposition (1) in the Appendix B). Proposition (2) in Appendix B shows, however, that it may be possible to induce high effort in both workers when social networks can ensure that K, the network rewards for high effort, are sufficiently high. For simplicity, suppose that the degree of complementarity is high then the binding constraint is T_1 without networks and $\tilde{T}_1$ with networks. The firm cannot induce high effort in both workers without the power of social networks, e.g. if $\tilde{T}_1 \leq x_1-x_2<T_1$. Similarly, if the binding constraint is T_2 without networks and $\tilde{T}_2$ with networks, then the firm cannot induce high effort in both workers without networks but can do so with networks under the condition $\tilde{T}_2 \leq x_1-x_2<T_2$. Moreover, as f_i^k increases, the wages needed to reward worker i for high effort will decrease, therefore for any given expected monetary incentives (such as overtime bonus or promotions), worker i puts in higher effort.

Overall, our theoretical analysis suggests that less able workers are more likely to be holding up wages of the high ability workers due to low assembly line output. However, when the social network size in the line increases it leads to higher effort by low ability workers for the same fixed wages, but coupled with greater chances of getting overtime or promotions. High ability workers will then increase effort in response to the rise in potential expected wages they can get from the supervisor. The key part of our theory is that due to complementarities in production, high ability workers have strong incentives to enforce greater effort from low ability workers using social network rewards or punishments. By themselves, high ability workers cannot increase line level output and therefore the probability of getting higher expected wages from the firm.[26] Thus, the effort level of high ability workers responds less to an increase in monitoring by the network while it responds more for precisely those workers who might be holding up line output.[27] As the number of such potential enforcers/monitors/informants (to other network members outside the line) in the line increases, low ability workers increase their effort correspondingly.

5 Methodology and Results

5.1 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. As discussed previously, the management allocates workers to lines when they join the factory. We observe a significant difference in the allocated and observed line strength across work days. Daily changes in line strength leads to changes in the worker composition of the line due to unanticipated worker absenteeism and attrition, which is higher than the average in the manufacturing sector. In addition the floor manager has to re-allocate workers across lines due to worker absence so as to meet production targets. Given the high pressure to meet production targets (due to high competition in the product

[26] Note that assuming $c > K$, expected wages are higher when both workers put in high effort than when only the high ability worker puts in high effort.

[27] When complementarities are sufficiently strong, i.e. $T_1 > T_2$, high ability workers start from a higher wage and higher productivity level than low ability workers, so as a percentage of initial output, responsiveness is higher for the low ability workers. But within line variance is unaffected.

market), the scope for being able to selectively choose workers is limited.[28]

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 Pearsons 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 A2 in Appendix A gives a snapshot of the caste distribution of workers in production lines on a randomly selected work day for the export factory and Table A3 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.

Moreover, we find that worker absenteeism is not systematically correlated with workers' caste category (see Table A4 in Appendix A). 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 both worker absenteeism and reallocation are independent of own caste. In our empirical analysis, therefore, we use worker absenteeism as a source of exogenous variation in the caste composition of workers in a line across days.

[28]We deliberately emphasise the use of caste as a proxy for networks. Given the politically sensitive nature of such classifications and the possibilities of conflict among workers, it is unlikely that the factory would group workers according to caste. In our sample the management did not collect information on workers' caste at the time of recruitment.

5.2 Estimation methodology

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

$$\mathrm{Y}_{ilt} = \alpha + \beta network\_strength_{ilt} + \gamma \mathrm{X}_i + \epsilon_{ilt} \tag{10}$$

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. $\mathbf{X_i}$ 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. 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 workers productivity.

Equation (10) 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 in subsequent specifications, besides factory floor and line fixed effects to account for floor and line level unobservables (e.g. floor managers' and line supervisors' characteristics).[29]

To analyze line level productivity we estimate equation (10) at the line level and measure social connections amongst workers in the line by the caste concentration

[29] Suppose worker motivation to work on date t is affected by caste composition in line l on day t, then it may be argued that absenteeism (and hence caste composition) in line l on day $t+1$ is affected by caste composition on day t. But we have already shown that assignment of workers is independent of caste and absenteeism does not vary systematically by caste. If motivation of workers is indeed affected by caste composition, then note that since on average the largest worker group is H type, we would expect minority caste groups, M and L, to be disproportionately more affected by caste composition of their line. However, despite the asymmetry in the share of castes of H vs. M and L in the workforce, we do not find a significant difference in the absenteeism rates for the three castes.

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 (10), as controls. In subsequent, stricter specifications, we include floor and line fixed effects to control for time invariant, line level unobservables.[30] The standard errors are clustered at factory-line level, as in the individual level analysis.

5.3 Results

5.3.1 Line composition and worker performance

The results of the analysis using equation (10) are presented in Table 3. In columns 1-4 we conduct the analysis for all production lines - assembly and non-assembly. Column (1) shows estimates of equation (10), where 'Network strength' is as defined in equation (10). The coefficient β is positive, suggesting that a one percentage point increase in the proportion of workers of one's own caste increases, albeit insignificantly, an individual worker's efficiency by 6.7 percentage points. In column 2 we include individual fixed effects. The coefficient of interest is now not only significant at the 5% level, it is also larger in magnitude. A percentage point increase in the proportion of workers who are own caste in the line raises individual productivity by more than 10 percentage points. In subsequent columns we include floor (column 3) and line (column 4) fixed effects. The magnitude and significance of the estimate is robust.

To elaborate on what this estimate implies, 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

[30]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.

pieces while her daily target is 640 pieces (8 hours x 80 pieces). Given the average efficiency of 31%, assume she manages to complete only 192 pieces. An increase of 10 percentage points in her daily efficiency implies that her daily output increases by 64 pieces or, on average, 8 additional stitched pieces per hour when the number of own caste workers increases by about one-half (i.e. about 1 percentage point in an average line of 33 workers with equally distributed H, M and L caste.). Since the mean worker efficiency is 31% the estimates in columns (2) - (3) suggest that worker efficiency can rise by approximately 30.6 - 33.2% when a worker is more socially connected within her line. While these effects may seem large, note that the the base is very low (average worker productivity is 0.3) implying large increases in percentage point terms.

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 columns (5) - (8). The coefficient β is somewhat stronger, suggesting 34.2 to 37.7% higher worker efficiency when the proportion of own caste workers in the line rises by 1 percentage points. This also suggests that the overall effects we observe in columns 1-4 are driven by assembly lines.

5.3.2 Line composition and line performance

In Table 4 and Table 5 we estimate the minimum and average line efficiency, respectively, using equation (10) for all lines and assembly lines, as in the worker level analysis in Table 3. In Table 4, column 1 we include only line level characteristics as controls. A one percentage point increase in the network strength as measured by the CCI causes a 11.3 percentage point increase in the line's minimum efficiency. Augmenting the specification with floor fixed effects increases the point estimate to 12.1

percentage points and to 15.8 percentage points when we address line level heterogeneity. Restricting the sample to assembly lines alone does not change our estimates much. Given that the average minimum line efficiency is 5%, the estimates of the impact of network strength are very large. In the strictest specification with line fixed effects, the results suggest that the minimum efficiency of the line or the least productive workers performance increases by 316% when more workers in the line belong to the same caste-based social network.

In Table 5 we show the results of the same analysis but when the dependent variable is the average efficiency of the line. Columns 1- 3 indicate a 22 to 24.7 percentage point improvement in a lines average efficiency when the caste composition of the line is more homogeneous. We restrict the sample to only assembly lines and redo the analysis in columns 4-6. The sample size falls from 1043 to 868 but the point estimates are similar to the ones obtained from the entire sample in columns 1-3. Our preferred specification with line fixed effects suggests 78.3 - 122% higher average efficiency when the lines network strength increases by 1%.

Overall, and in line with the theoretical model, our results suggest that the higher the proportion of co-workers from the same caste in a line on a day the higher is the performance of the worker and the line. We see the largest impact of network strength on the least productive workers in a line, given the near zero output of the most lax worker.[31]

6 Robustness

6.1 Sample selection

A simple *t*-test for those workers who have lower vis-a-vis higher than median attendance shows that the former have significantly lower efficiency. Even though we

[31] We do not find any non-linear impacts of network strength on either individual or line level performance.

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 data is systematically correlated with worker productivity or ability. Using the daily attendance data from the HR records for 61 working days (1^{st} August to 14^{th} October 2015) and worker production days data from the stitching department for 31 days (8^{th} September to 14^{th} October 2015), we analyze the characteristics of workers who are observed more regularly. As shown in Table A4, 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, and this is systematically correlated with the caste composition of co-workers in a line. We adopt a non-parametric method to check the robustness of our results in Table 6 to this potential selection bias - inverse probability weights (IPW) suggested by Moffitt et al. (1999) and Baulch and Quisumbing (2011). Intuitively, IPW method gives greater weightage to workers who are more likely to be absent (and of lower productivity) on a given work day. Using the inverse of predicted probability of being present, we re-run the worker level analysis in Table 6. Columns 1-3 report the original, unweighted estimates while columns 4-6 show the IPW estimates for corresponding specifications. We do not find any significant difference either in the magnitude or significance of the estimates, suggesting that selection on worker characteristics is not driving our results.

[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.

6.2 Trends

As we mentioned previously, demand can vary over time (due to seasonal changes, festivals etc.) both within and across lines in a garment factory. This can influence individual and line productivity, as well as the composition of workers in a line. Supervisors and managers may reallocate workers across or within lines purposively to meet production targets which may be correlated with caste categories of workers. In Table A5 we report the results of the analysis with month of production and line specific month of production fixed effects. Our results are robust to secular and line specific trends except in column 6 when the outcome is the average efficiency of the line. The impact of network strength on average efficiency is, however, marginally significant ($p<0.10$) when we restrict the sample to only assembly lines. Note that our measure of efficiency accounts for any changes in production style. Nevertheless, we check the robustness of our estimates to trends at the production week level as well as production style fixed effects. The results are unchanged.

6.3 Number of clusters

Another concern with our estimates is that high intra-cluster correlation, coupled with the small number of clusters (production lines) in our study, would lead to incorrect standard errors. Although we have addressed the possibility of high intra-cluster correlation by clustering our standard errors at the line level, the presumption that these standard errors are correct is based on having a large number of clusters. Even though the number of clusters (or lines) do not fall below the acceptable standard of 30, we may be falsely inferring the significance of the coefficients. We, therefore, report our results with bootstrapped standard errors in Table 7. Columns 1-2 report pair-wise bootstrapped standard errors, with and without line fixed effects, respectively. In columns 3 and 5 we report the pair-wise bootstrap standard errors and use the cluster-bootstrap procedure proposed by Cameron et al. (2008) in columns 4 and 5. Our standard errors are marginally higher but the main coefficient of interest remains significant, consistent with results reported in Tables 3-5.[33]

7 Mechanism

Our theoretical framework relies on the ability of social networks to provide reciprocal benefits when workers help their peers to get overtime or promotions. 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. The description of job informant characteristics in Table 8 (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 8, 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 in the line, as shown in Panel C of Table 8, 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 more costly if it adversely affects the productivity of own-caste co-workers in

[33] We also drop outlier observations, i.e. those line-days (not the entire day) whose worker strength falls in the lowest one percentile of the distribution of strength and those days on which number of factory lines is less than 30. From 1043 line-days we end up with 972 line-days. We then wild-cluster bootstrap our standard errors, which gives the same conclusions as in Tables 5 and 6.

the line which in turn reduces their financial payoffs as discussed in Section (2.2).[34] Since co-workers are aware of where the bottlenecks in the line are, a worker who slacks can potentially lose the benefits she derives from her network through network retribution. This threat of social sanctions or loss of reputation would be higher for the low performing worker, who is holding up line output. Indeed our results show that the effect of more own caste workers in the assembly line on a workers efficiency is larger for least performing worker (16 percentage points) as compared to the average productivity worker (10 percentage points). The lowest efficiency workers are typically younger and have been in the garment industry for fewer years, according to our data. Hence workers may want to maintain their reputation with fellow caste members so as to ensure future access to jobs and referrals.[35]

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'. If the reputation mechanism is valid then we should see a significant positive coefficient on this interaction term. Our results suggest exactly that. In columns 1 and 2 in Table 9, we find that almost all of the effect of network strength can be explained by its interaction with informant presence in factory. In columns 3 and 4, for line level analysis, we find a negative albeit insignificant effect of informant presence on the lines average (column 3) or minimum (column 4) efficiency, but a positive (insignificant) effect of the interaction term. The total effect of informant presence is significant in column 3 ($p<0.10$) and only marginally insignificant in column 4.[36]

[34] 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 payoffs.

[35] 87.1% of workers with less than 1 year of experience obtained job information from network as opposed to 49.2% of those with almost 13 years of experience.

[36] 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 Table A6. 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.

There are, however, other candidate mechanisms that can explain our findings. First, knowledge spillovers through peer effects is likely when co-workers can observe each other's effort or output, are performing similar tasks and/or can communicate. However, as 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. But when we restrict our sample to only assembly lines in Tables 3-5, the coefficient on network strength is more robust, suggesting that learning from peers is unlikely to be driving our observed findings. Further, suppose spillovers amongst socially connected peers are present. We might expect that knowledge spillovers from high productivity to low productivity workers among the socially connected may lower the line level variance in individual output. But we do not find any significant impact of network strength on within line variation in efficiency, using equation (10). Hence, while knowledge spillovers may exist, it is unlikely that they alone can explain our results.[37]

A second possible mechanism is peer effects through conformism to a group norm, since greater caste homogeneity in the line may make norms more salient. Thus across all caste groups, a worker would respond with decreased (increased) effort when her own caste proportion in the line increases, if she has a higher (lower) effort than the caste norm. Hence we should observe decreased variance in the efficiency of the same caste workers with an increase in own caste proportion. First, recall that we do not find evidence of productivity differences by caste groups. This implies that (if effort norms are the same across castes) then there should be a fall in variance in productivity on the line as well. However we do not observe any impact on variance in worker productivity in the line with increase in caste-based network strength, as

[37] In addition, note that co-workers in adjoining lines perform different style-operations which may have limited knowledge spill-over effects, if any, on the performance of a given worker.

would be expected if there was norm conformity (Table A7, Appendix A). [38]

Can altruism among the same caste affect productivity? Social preferences affect the cost of effort in our framework - if the marginal cost decreases with higher proportion of same caste due to better motivation or higher competitiveness, then we expect individual effort to increase for all workers, irrespective of experience, tenure and productivity. While altruism predicts that lower productivity workers will work harder to improve the chances of higher productivity workers of the same caste having greater access to overtime or promotions, it is also consistent with high productivity workers increasing their effort and within line dispersion in productivity declining with more homogeneous caste composition. Our regression analysis, however, shows that the impact of the caste concentration index on average efficiency is less robust than on the least efficient worker in the line. Also, we do not find a decline in productivity variance with increase in network strength or homogeneity.[39]

We conclude that economic interdependence within one's social network is a likely mechanism through which workers put in greater effort when the presence of co-workers within the network in the team is larger.

8 Conclusion

Using caste as the defining characteristic of social networks amongst workers along with exogenous variation in the caste composition of production lines across work days in garment factories in India, we show that the greater the strength of one's caste-based social network the higher the worker and line level productivity on a work day.

[38]Note that our analysis captures social network rather than caste identity motivations. We do not find a decline in the productivity of workers whose network strength falls in a line on a workday. We, therefore, rule out taste-based preferences as an explanation of our findings.

[39]Furthermore, 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 A8, Appendix A), suggesting that productivity of the least efficient worker increases more when there are more own-caste, senior workers in the line. This negates altruism as a possible mechanism.

Our findings suggest that in competitive product markets, workers' social networks can be leveraged to improve efficiency in the absence of high-powered performance based incentives.

These findings extend the literature on the role of social networks and job referrals, in general, and on productivity, in particular. They suggest that when production is team based, and tasks differ amongst the members of a team, even in the absence of group based financial incentives social interdependence of group members can enforce good behavior due to the interdependence of financial payoffs emanating from production externalities at work. Although our analysis is based on garment factory production lines, the results are applicable to contexts where workers are complementary in the production process but financial compensation is fixed and at the individual level.

References

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

Charness, G., L. Rigotti, and A. Rustichini (2007): "Individual Behavior and Group Membership," *American Economic Review*, 97, 1340–1352.

Chen, Y. and R. Chen (2011): "The Potential of Social Identity for Equilibrium Selection," *American Economic Review*, 101, 2562–2589.

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.

Goette, L., D. Huffman, and S. Meier (2006): "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups," *American Economic Review*, 96, 212–216.

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.

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

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.

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.

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.

Pallais, A. and E. G. Sands (2016): "Why the Referential Treatment? Evidence from Field Experiments on Referrals," *Journal of Political Economy*, 124, 1793–1828.

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		
Characteristics	**All** N=1744	**L** N=384	**M** N=543	**H** 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. Standard errors in parentheses.

Table 2: Worker and line performance

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

Note: Efficiency is defined as the actual output/target output. The top panel 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 bottom panel shows the efficiency at the line and line-day level. Average line efficiency is the mean efficiency of workers in the line; minimum line efficiency is the lowest worker efficiency in the line. Average number of workers in a line is 33. Standard errors in parentheses.

Table 3: Worker performance and line composition

	Worker efficiency							
	All lines				*Assembly lines*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Network strength (β)	0.067	0.103**	0.103**	0.095**	0.105**	0.117**	0.116**	0.106**
	(0.045)	(0.047)	(0.046)	(0.045)	(0.046)	(0.052)	(0.051)	(0.050)
Constant	0.254***	0.276***	0.259***	0.328***	0.278***	0.279***	0.262***	0.333***
	(0.031)	(0.019)	(0.075)	(0.071)	(0.031)	(0.020)	(0.080)	(0.076)
Individual fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Floor fixed effects	No	No	Yes	No	No	No	Yes	No
Line fixed effects	No	No	No	Yes	No	No	No	Yes
Number of observations	34,641	34,641	34,641	34,641	32,176	32,176	32,176	32,176
Number of workers	1744	1744	1744	1744	1633	1633	1633	1633
Number of lines	37	37	37	37	31	31	31	31
R-square	0.010	0.550	0.550	0.555	0.011	0.546	0.546	0.550

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

Table 4: Line performance and composition

	Minimum Worker efficiency					
	All lines			*Assembly lines*		
	(1)	(2)	(3)	(4)	(5)	(6)
Network strength (β)	0.113**	0.121***	0.158***	0.067*	0.110***	0.159***
	(0.045)	(0.028)	(0.042)	(0.037)	(0.034)	(0.038)
Constant	0.214*	0.232**	0.163*	0.402***	0.309***	0.328***
	(0.123)	(0.103)	(0.085)	(0.074)	(0.081)	(0.077)
Floor fixed effects	No	Yes	No	No	Yes	No
Line fixed effects	No	No	Yes	No	No	Yes
Number of observations	1043	1043	1043	868	868	868
Number of lines	37	37	37	31	31	31
R-square	0.484	0.588	0.700	0.537	0.641	0.697

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

Table 5: Average line performance and composition

	Average efficiency of line					
	All lines			*Assembly lines*		
	(1)	(2)	(3)	(4)	(5)	(6)
Network strength (β)	0.247***	0.220***	0.235**	0.221**	0.241***	0.359**
	(0.075)	(0.065)	(0.111)	(0.090)	(0.085)	(0.137)
Constant	0.398**	0.461**	0.457*	0.311	0.395*	0.853**
	(0.196)	(0.171)	(0.246)	(0.215)	(0.222)	(0.396)
Floor fixed effects	No	Yes	No	No	Yes	No
Line fixed effects	No	No	Yes	No	No	Yes
Number of observations	1043	1043	1043	868	868	868
Number of lines	37	37	37	31	31	31
R-square	0.214	0.296	0.449	0.179	0.213	0.395

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

Table 6: Worker performance and line composition (inverse probability weights)

	Worker efficiency					
	(1)	(2)	(3)	(4)	(5)	(6)
Network strength (β)	0.103**	0.103**	0.095**	0.103**	0.102**	0.094**
	(0.047)	(0.046)	(0.046)	(0.047)	(0.046)	(0.046)
Constant	0.276***	0.259***	0.328***	0.276***	0.258***	0.329***
	(0.019)	(0.075)	(0.071)	(0.019)	(0.075)	(0.071)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Floor fixed effects	No	Yes	No	No	Yes	No
Line fixed effects	No	No	Yes	No	No	Yes
Number of observations	34,641	34,641	34,641	34,623	34,623	34,623
Number of workers	1744	1744	1744	1740	1740	1740
Number of lines	37	37	37	37	37	37
R-square	0.550	0.550	0.555	0.549	0.550	0.554

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 1-3. Regressions weighted by inverse of the probability of worker being present on a workday in columns 4-6. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table 7: Worker, line performance and composition (bootstrap standard errors)

	Worker efficiency		Line level: Minimum efficiency		Line level: Average efficiency	
	(1)	(2)	(3)	(4)	(5)	(6)
Network strength (β)	0.103**	0.095**	0.158***	0.158**	0.235*	0.235*
	(0.032)	(0.019)	(0.004)	(0.015)	(0.084)	(0.088)
Constant	0.276***	0.328**	0.064	0.163	0.511*	0.456*
	(0.000)	(0.009)	(0.564)	(0.126)	(0.086)	(0.08)
Individual fixed effects	Yes	Yes				
Line fixed effects	No	Yes	Yes	Yes	Yes	Yes
Number of observations	34,641	34,641	1043	1043	1043	1043
Number of workers	1744	1744				
Number of lines	37	37	37	37	37	37
R-square	0.550	0.013	0.273	0.700	0.001	0.449

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 1-2, 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 3-6. Regressions results with pairwise bootstrapped standard errors clustered at line level in columns 1, 3 and 5; pairwise bootstrapped standard errors in column 2; wild-cluster (at line level) bootstrapped standard errors in columns 4 and 6. 2000 replications across all regressions. Significant at *10%, **5% and ***1%.

Table 8: Job networks, residential location and caste

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: 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: 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***

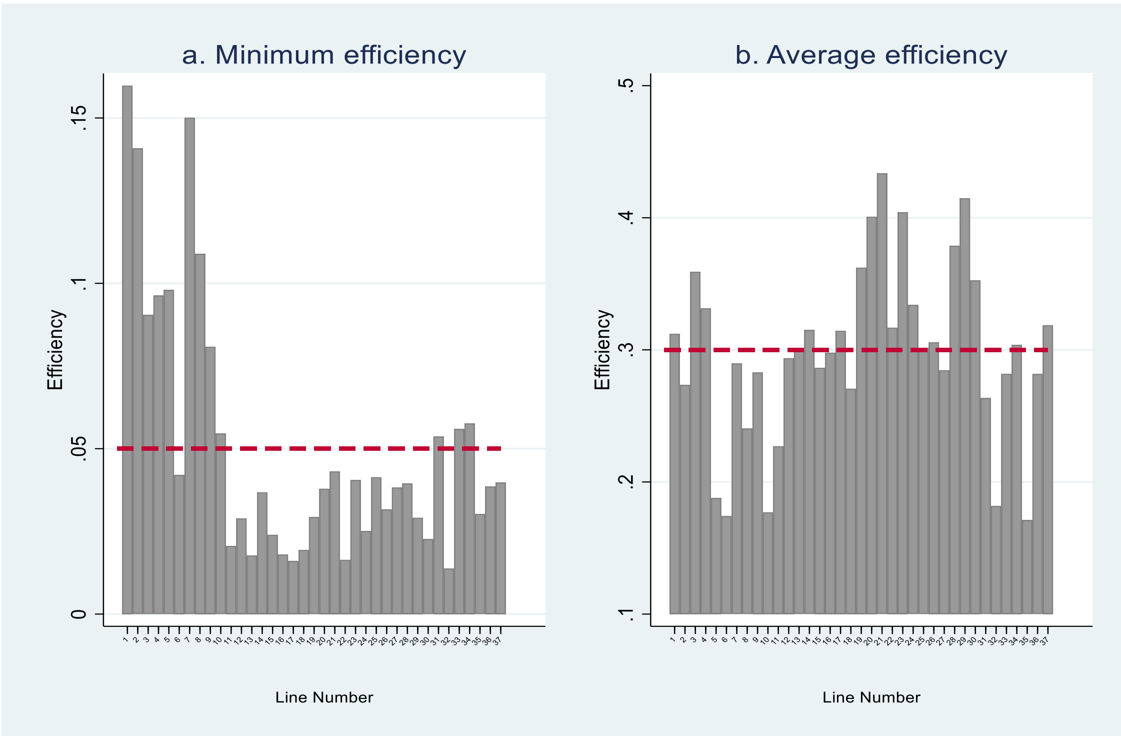
Note: [@]conditional on informal flow of job opening information; [#]smaller number of observation due non-response. In Panels B and C the sample is in worker-days, conditional on data on both caste and unit of residential location being available for a worker. Significant at *10%,**5% and ***1%.

Table 9: Worker, line performance and job referee presence

	Worker efficiency		*Line Efficiency*	
	(1)	(2)	(3)	(4)
(1) Proportion own caste	0.044	0.038		
	(0.047)	(0.046)		
(2) Proportion own caste x referee employed in factory	0.227***	0.225***		
	(0.062)	(0.059)		
(3) Caste concentration index			0.137	0.117*
			(0.146)	(0.064)
(4) Proportion with referee employed in factory			-0.107	-0.050
			(0.204)	(0.063)
(5) Caste concentration index x proportion with referee employed in factory			0.449	0.185
			(0.354)	(0.133)
Constant	0.266***	0.334***	0.609*	0.225***
	(0.075)	(0.071)	(0.301)	(0.069)
Effect of referee employed in factory:				
(4) + (5)			0.343*	0.135
			(0.189)	(0.087)
Individual fixed effects	Yes	Yes	No	No
Floor fixed effects	Yes	No	No	No
Line fixed effects	No	Yes	Yes	Yes
Number of observations	34,641	34,641	1043	1043
Number of workers	1744	1744		
Number of lines	37	37	37	37
R-square	0.551	0.555	0.454	0.704

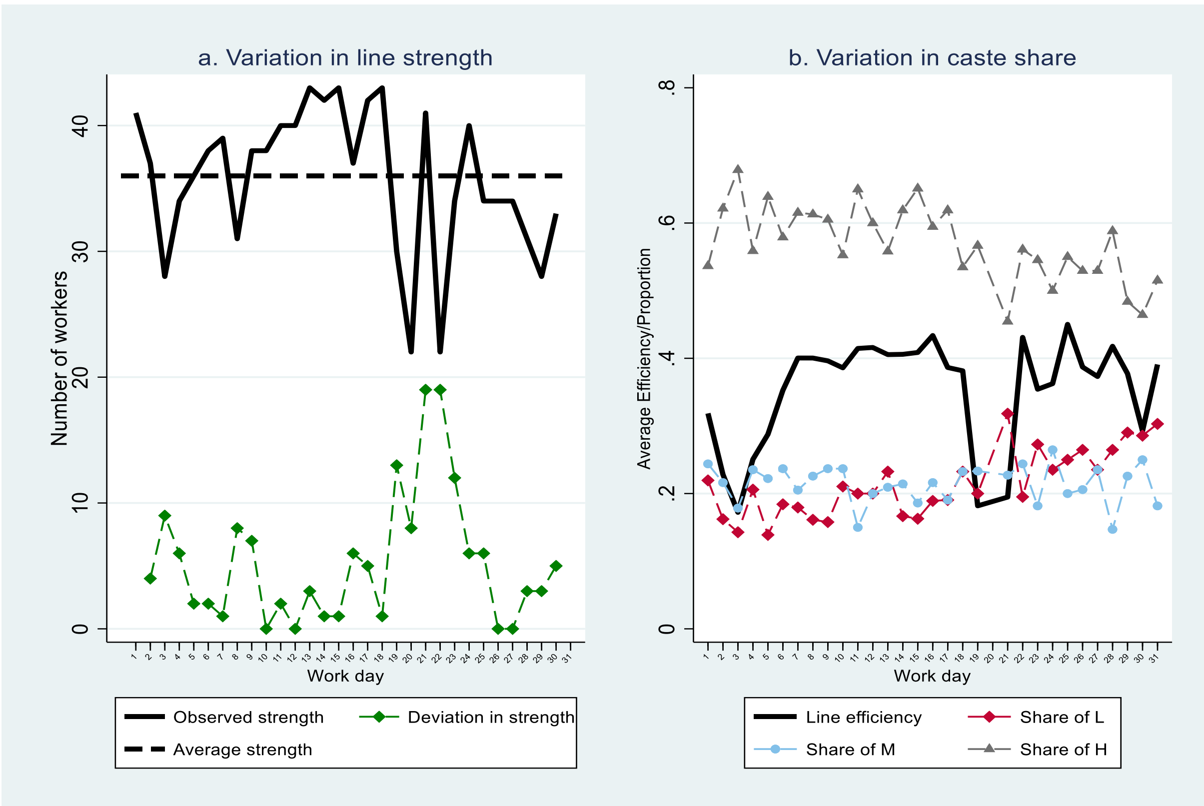
Note: In columns 1 and 2 the dependent variable is the efficiency of the worker on a work day. In column 3 the dependent variable in the average efficiency of the line. In column 4 the dependent variable is the minimum efficiency of the line. Referee employed in the factory is a dummy variable that takes value 1 if the workers 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). Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Figure 1: Line performance



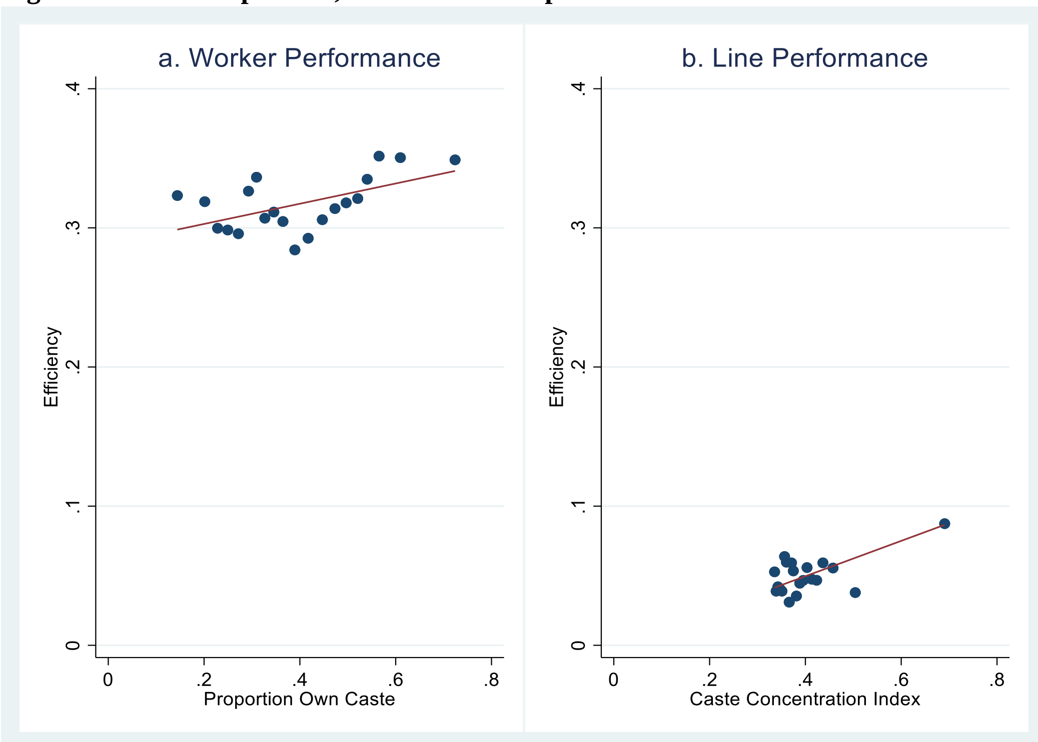
Note: Fig. 1(a) shows the mean daily minimum efficiency of each production line over workdays. Average minimum efficiency over the sample period is 0.05 (given by dashed red line). Fig. 1(b) shows the mean daily average worker efficiency of each line over workdays. Average line efficiency over the sample 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 2: Daily variation in line composition and performance (representative line)



Note: Fig. 2(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. Fig. 2 (b) shows the corresponding changes in each caste share and the daily average efficiency of the same line. Data obtained for September-October 2015 from factory records and worker level primary survey.

Figure 3: Caste composition, worker and line performance



Note: Fig. 1(a) 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; Fig. 1(b) shows the minimum worker efficiency in an assembly line on a production day for 1043 line days. Average minimum efficiency per line is 0.05. Caste concentration index=$\Sigma c^2 i$, i.e. the sum of squared share of each caste group (L, M, or H) among the workers in an assembly line on a day. Linear fit depicted in both figures using the 'binscatter' command in STATA dividing the data into 20 bins, plotting the mean X and Y values for each bin. 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.

FOR ONLINE PUBLICATION

APPENDIX A. Additional Results

Table A1: 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 A2: Chi-square test of exogeneity of caste assignment to line (export factory)

Line Number / Caste Category	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
L	13	7	12	15	11	9	13	11	15	11	8	13	12	9	10
	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
	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
M	16	12	14	14	7	16	16	15	10	14	20	15	18	12	16
	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
	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
H	17	18	20	18	28	24	15	15	23	27	35	16	27	25	19
	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
	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
Total	**46**	**37**	**46**	**47**	**46**	**49**	**44**	**41**	**48**	**52**	**63**	**44**	**57**	**46**	**45**
	46	**37**	**46**	**47**	**46**	**49**	**44**	**41**	**48**	**52**	**63**	**44**	**57**	**46**	**45**
	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**

Table A2: Chi-square test of exogeneity of caste assignment to line *(continued)*

Line Number / Caste Category	16	17	18	19	20	21	22	23	24	25	26	TOTAL
L	9	2	5	3	6	6	2	5	8	5	7	**227**
	10.9	3.5	9.8	6.7	6.5	3.7	4.8	8.7	6.3	6.7	5	**227**
	0.3	0.6	2.3	2.1	0	1.4	1.6	1.6	0.5	0.4	0.8	**23.2**
M	13	6	15	9	7	3	3	12	6	11	8	**308**
	14.7	4.7	13.3	9.1	8.8	5	6.5	11.8	8.5	9.1	6.8	**308**
	0.2	0.3	0.2	0	0.4	0.8	1.9	0	0.8	0.4	0.2	**13.1**
H	28	8	25	19	17	8	17	23	15	15	8	**510**
	24.4	7.8	22	15.1	14.6	8.3	10.7	19.5	14.2	15.1	11.2	**510**
	0.5	0	0.4	1	0.4	0	3.7	0.6	0.1	0	0.9	**17.6**
Total	**50**	**16**	**45**	**31**	**30**	**17**	**22**	**40**	**29**	**31**	**23**	**1045**
	50	**16**	**45**	**31**	**30**	**17**	**22**	**40**	**29**	**31**	**23**	**1045**
	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 Pearson's Chi-square. Pearson's Chi-square statistics is 53.975 with 50 degrees of freedom and *p* value =0.325. We can't 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 A3: 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	**13**	**12**	**15**	**16**	**10**	**26**	**10**	**4**	**10**	**18**	**134**
	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 Pearson's Chi-square. Pearson's Chi-square statistics is 22.13 with 18 degrees of freedom and p value =0.226. We can't 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 A4: Worker attendance

	Attendance rate		Working days	
Characteristics	(1)	(2)	(3)	(4)
Age (years)	0.001***	0.001***	0.051	0.060*
	(0.000)	(0.000)	(0.036)	(0.035)
Married	-0.013*	-0.013*	-1.798***	-1.583***
	(0.006)	(0.007)	(0.527)	(0.512)
Female	-0.010	-0.006	1.463**	1.757***
	(0.008)	(0.008)	(0.548)	(0.556)
Native state Bihar	0.014***	0.010**	0.636*	0.509*
	(0.004)	(0.005)	(0.352)	(0.298)
Hindu	0.032***	0.033***	2.534***	2.155***
	(0.010)	(0.010)	(0.632)	(0.609)
Secondary education or more	0.005	0.003	0.014	0.203
	(0.005)	(0.005)	(0.477)	(0.410)
Obtained job information informally	0.00004	0.0002	0.380	0.899*
	(0.005)	(0.006)	(0.570)	(0.460)
Experience (years)	-0.001***	-0.001***	0.322***	0.238***
	(0.0004)	(0.0005)	(0.062)	(0.055)
H	0.001	0.003	-0.356	-0.440
	(0.006)	(0.006)	(0.430)	(0.283)
M	0.008	0.006	0.280	0.064
	(0.007)	(0.007)	(0.503)	(0.453)
Number of reported friends	-0.0002	0.0004	0.089	0.227*
	(0.002)	(0.002)	(0.169)	(0.125)
Line supervisor same caste	-0.001	0.003	0.316	0.293
	(0.006)	(0.005)	(0.291)	(0.297)
Constant	0.865***	0.876***	14.36***	13.17***
	(0.014)	(0.013)	(1.204)	(0.869)
Line Fixed Effects	No	Yes	No	Yes
Number of workers	1731	1731	1735	1735
Psuedo-R2	0.023	0.052	0.041	0.197

Note: The first column uses factory attendance data. Attendance rate is the number of present days/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. The second column is based on the production data. Working days is the count of days a worker appears in the productivity data (excluding half days, 0.30% of the worker days). Robust standard errors, clustered at the line level, in parentheses. Attendance data missing for 4 workers; line information missing for 9 workers. Significant at *10%, **5% and ***1%.

Table A5: Worker, line performance and caste composition

	Worker efficiency		Line level: Minimum efficiency		Line level: Average efficiency	
	(1)	(2)	(3)	(4)	(5)	(6)
Proportion own caste	0.079*	0.087**				
	(0.041)	(0.039)				
Caste concentration index			0.108***	0.139**	0.165**	0.172
			(0.031)	(0.046)	(0.067)	(0.112)
Constant	0.262***	0.240***	0.209**	0.111	0.366**	0.367
	(0.076)	(0.076)	(0.102)	(0.077)	(0.176)	(0.227)
Individual FE	Yes	Yes				
Floor FE	No	No	Yes	No	Yes	No
Line FE	Yes	Yes	No	Yes	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Month x Line FE	No	Yes	No	Yes	No	Yes
Number of	34,641	34,641	1043	1043	1043	1043
Number of workers	1744	1744				
Number of lines	37	37	37	37	37	37
R-square	0.565	0.576	0.607	0.752	0.362	0.586

Note: The dependent variable is worker efficiency in columns 1-2; minimum efficiency of line in columns 3-4 and average efficiency of line in columns 5-6. 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. Robust standard errors, clustered at line level, reported in parentheses. Significant at *10%, **5% and ***1%.

Table A6: Worker performance and attendance rate

	Worker efficiency			
	(1)	(2)	(3)	(4)
Proportion own caste	0.098	0.049	0.048	0.038
	(0.059)	(0.059)	(0.057)	(0.055)
Proportion own caste x Above median attendance	-0.046	0.086	0.087	0.089
	(0.066)	(0.069)	(0.068)	(0.070)
Constant	0.228***	0.275***	0.260***	0.332***
	(0.036)	(0.019)	(0.074)	(0.070)
Individual fixed effects	No	Yes	Yes	Yes
Floor fixed effects	No	No	Yes	No
Line fixed effects	No	No	No	Yes
Number of observations	34,641	34,641	34,641	34,641
Number of workers	1744	1744	1744	1744
Number of lines	37	37	37	37
R-square	0.013	0.550	0.550	0.555

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 worker attendance >= median work days; 0 otherwise. Median working days = 22. Robust standard errors, clustered at the line level, in parentheses. Significant at *10%,**5% and ***1%.

Table A7: Dispersion in worker performance and network strength

	Dispersion in individual worker productivity			
	(1)	(2)	(3)	(4)
Caste concentration index	0.093*	0.080**	0.066	0.051
	(0.055)	(0.031)	(0.064)	(0.060)
Constant	0.165	0.238**	0.176	0.156
	(0.159)	(0.110)	(0.153)	(0.156)
Floor fixed effects	No	Yes	No	No
Line fixed effects	No	No	Yes	Yes
Month fixed effects	No	No	No	Yes
Number of observations	1041	1041	1041	1041
Number of lines	37	37	37	37
R-square	0.314	0.512	0.584	0.586

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

Table A8: Worker performance, experience and network strength

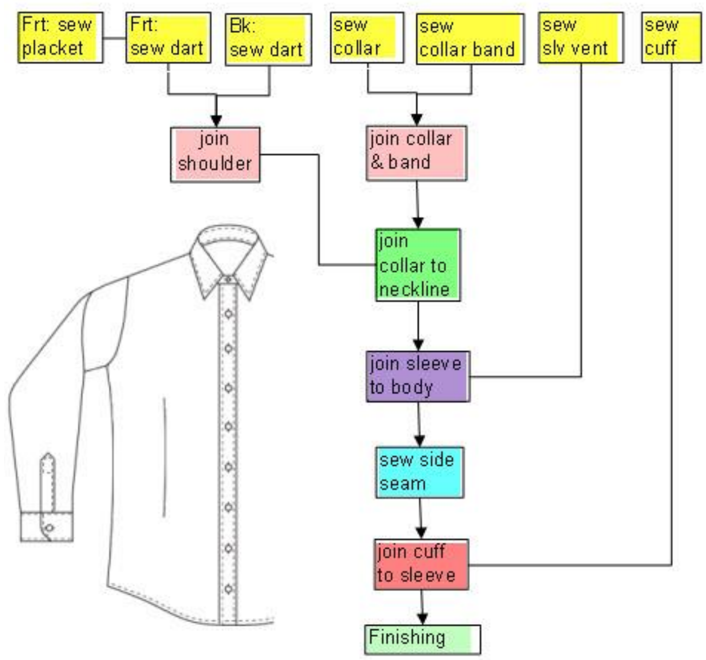
	Minimum efficiency			
	(1)	(2)	(3)	(4)
Caste concentration index (CCI)	-0.159*	-0.120	-0.028	-0.078
	(0.091)	(0.086)	(0.087)	(0.082)
Proportion high experience	-0.326***	-0.250***	-0.170***	-0.175***
	(0.076)	(0.075)	(0.054)	(0.048)
Proportion high experience x CCI	0.598***	0.538***	0.398**	0.445***
	(0.173)	(0.174)	(0.149)	(0.147)
Constant	0.360***	0.304***	0.266***	0.244***
	(0.094)	(0.073)	(0.081)	(0.075)
Floor fixed effects	No	Yes	No	No
Line fixed effects	No	No	Yes	Yes
Month fixed effects	No	No	No	Yes
Number of observations	1043	1043	1043	1043
Number of lines	37	37	37	37
R-square	0.537	0.616	0.709	0.728

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

Figure A1: Factory floor and line organisation

Location: Faridabad
Source: icrw.org

Figure A2: Manufacturing process of a shirt



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

APPENDIX B: Theoretical Framework

B.1 Benchmark model without social networks

The optimization problem is to choose w_1, w_2 to maximize (per worker expected profit):

$$E(\pi(e_h, e_h)) = \theta + \pi_{h,h} - \alpha^{hh} w_1 + (1 - \alpha^{hh}) w_2 \tag{B.1.1}$$

subject to the incentive compatibility (IC) constraints, the participation constraints (PC) and a limited liability (LL) constraint.

(1) the PC says that a worker will only accept the implicit contract offering expected wages $\alpha^{hh} w_1 + (1 - \alpha^{hh}) w_2$ if the cost of effort is low enough that utility is higher than the outside option of minimum wages in another firm:

$$\alpha^{hh} w_1 + (1 - \alpha^{hh}) w_2 - c \geq \underline{w} \tag{B.1.2}$$

which can be re-written as

$$\alpha^{hh} (w_1 - w_2) + w_2 - c \geq \underline{w} \tag{B.1.3}$$

(2) The ICs: 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 that high effort is a dominant strategy for worker i: IC(1)(worker j puts in high effort)

$$\alpha^{hh} w_1 + (1 - \alpha^{hh}) w_2 - c \geq \alpha^{lh} w_1 + (1 - \alpha^{lh}) w_2 \tag{B.1.4}$$

which can be re-written as:

$$(\alpha^{hh} - \alpha^{lh})(w_1 - w_2) \geq c \tag{B.1.5}$$

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

$$\alpha^{hl} w_1 + (1 - \alpha^{hl}) w_2 - c \geq \alpha^{ll} w_1 + (1 - \alpha^{ll}) w_2 \tag{B.1.6}$$

which can be re-written as:

$$(\alpha^{hl} - \alpha^{ll})(w_1 - w_2) \geq c \tag{B.1.7}$$

and (3) the LL constraint: $w_1, w_2, w_3 \geq \underline{w}$

Lemma 1 *The solution to the maximization problem (B.1.1) is* $w_1 = \underline{w} + \frac{c}{\alpha^{hl} - \alpha^{ll}}$ *and* $w_2 = \underline{w}$.

Since $(\alpha^{hh} - \alpha^{lh}) > (\alpha^{hl} - \alpha^{ll})$, IC (B.1.7) $\implies$ IC(B.1.5). Moreover IC (B.1.7) $\implies$ $w_1 > w_2$. Let $w_2 = w$ be the base wage and $w_1 - w_2 = b$, the bonus. Then we have the following solution $w_1 = w + b = \underline{w} + \frac{c}{\alpha^{hl} - \alpha^{ll}}$ and $w_2 = \underline{w}$. This solution satisfies the PC.

Expected profits, assuming all workers get the same wages are: $E(\pi(e_h, e_h)) = \underline{\theta} + \bar{\theta} + 2(\pi_{h,h} - \alpha^{hh}(\underline{w} + \frac{c}{\alpha^{hl} - \alpha^{ll}}) + (1 - \alpha^{hh})\underline{w})$. Denote average ability as $\mu = \frac{\underline{\theta} + \bar{\theta}}{2}$. Then expected profits per worker are: $E(\pi(e_h, e_h)) = \mu + \pi_{h,h} - \alpha^{hh}(\underline{w} + \frac{c}{\alpha^{hl} - \alpha^{ll}} - (1 - \alpha^{hh})\underline{w}$.

Alternately, the firm can induce high effort only from one worker. Since ability is observable, w.l.o.g the firm would find it profitable to pay higher wages to induce high effort from the high ability worker and induce low effort (and pay minimum wages) from the low ability worker (or vice versa as long as only one worker is induced to put in high effort). Then the problem for the high ability worker is to choose w_1, w_2 to maximize:

$$E(\pi(e_h, e_l)) = \bar{\theta} + \pi_{h,l} - \alpha^{hl} w_1 + (1 - \alpha^{hl}) w_2 \tag{B.1.8}$$

subject to:

(1) the PC:

$$\alpha^{hl} w_1 + (1 - \alpha^{hl}) w_2 - c \geq \underline{w} \tag{B.1.9}$$

which can be re-written as:

$$\alpha^{hl}(w_1 - w_2) + w_2 - c \geq \underline{w} \tag{B.1.10}$$

(2) The IC

$$\alpha^{hl} w_1 + (1 - \alpha^{hl}) w_2 - c \geq \alpha^{ll} w_1 + (1 - \alpha^{ll}) w_2 \tag{B.1.11}$$

which can be re-written as:

$$(\alpha^{hl} - \alpha^{ll})(w_1 - w_2) \geq c \tag{B.1.12}$$

and (3) the LL constraint: $w_1, w_2, w_3 \geq \underline{w}$

Lemma 2 *The solution to the maximization problem (B.1.8) is* $w_2 = \underline{w}, w_1 = \underline{w} + \frac{c}{(\alpha^{hl} - \alpha^{ll})}$.

The proof follows the same logic as the proof of Lemma (1). By the same logic, $w_2 = \underline{w}, w_1 = \underline{w} + \frac{c}{(\alpha^{hl} - \alpha^{ll})}$. Total costs are now $\alpha^{hl} \frac{c}{(\alpha^{hl} - \alpha^{ll})} + \underline{w}$ and expected profits are positive iff $\mu + \pi_{h,l} - \alpha^{hl}(\frac{c}{\alpha^{hl} - \alpha^{ll}}) - \underline{w} \geq 0$.

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 $\mu + \pi_{ll} - \underline{w} \geq 0$.

What effort profile will the firm induce? Observe that (1) Expected profits with high effort for both workers are higher than expected profits when only one worker is induced to put in high effort if $\bar{\theta} + \underline{\theta} + 2\pi_{h,h} - 2\alpha^{hh}(\frac{c}{\alpha^{hl}-\alpha^{ll}}) - 2\underline{w} \geq \bar{\theta} + \underline{\theta} + 2\pi_{h,l} - \alpha^{hl}(\frac{c}{\alpha^{hl}-\alpha^{ll}}) - 2\underline{w}$, i.e. iff $\pi_{h,h} - \pi_{h,l} \geq (\alpha^{hh} - \frac{\alpha^{hl}}{2})(\frac{c}{\alpha^{hl}-\alpha^{ll}})$. (2) Expected profits with high effort for both workers are higher than expected profits when no worker is induced to put in high effort iff $\mu + \pi_{h,h} - \alpha^{hh}(\frac{c}{\alpha^{hl}-\alpha^{ll}}) - \underline{w} \geq \mu + \pi_{ll} - \underline{w}$. Thus high effort is induced for both workers when both (1) and (2) hold, or

$$x_1 - x_2 \geq \frac{2\alpha^{hh} - \alpha^{hl}}{2(\alpha^{hh} - \alpha^{hl})} \frac{c}{\alpha^{hl} - \alpha^{ll}} \tag{B.1.13}$$

and

$$x_1 - x_2 \geq \frac{\alpha^{hh}}{\alpha^{hh} - \alpha^{ll}} \frac{c}{\alpha^{hl} - \alpha^{ll}} \tag{B.1.14}$$

Let $T_1 \equiv \frac{2\alpha^{hh}-\alpha^{hl}}{2(\alpha^{hh}-\alpha^{hl})}\frac{c}{\alpha^{hl}-\alpha^{ll}}$ and $T_2 \equiv \frac{\alpha^{hh}}{\alpha^{hh}-\alpha^{ll}}\frac{c}{\alpha^{hl}-\alpha^{ll}}$. The firm induces high effort from both workers iff $x_1 - x_2 \geq max(T_1, T_2)$.

Inequality (B.1.13) $\implies$ inequality (B.1.14) iff $\frac{2\alpha^{hh}-\alpha^{hl}}{2(\alpha^{hh}-\alpha^{hl})} \geq \frac{\alpha^{hh}}{\alpha^{hh}-\alpha^{ll}}$. A necessary and sufficient condition for this is that the degree of complementarity is sufficiently high, i.e $\alpha^{hh} - \alpha^{ll} > A(\alpha^{hh} - \alpha^{hl})$, where $A = \frac{2\alpha^{hh}-\alpha^{hl}}{2\alpha^{hh}}$. This leads to our first Proposition (1):

Proposition 1 *Assume that the firm makes positive profits when low effort is induced for both workers, i.e.* $\mu \geq \underline{w} - \pi_{l,l}$. *The firm induces high effort in both workers iff* $x_1 - x_2 \geq \max(T_1, T_2)$. *Expected wages are* $\alpha^{hh}\frac{c}{(\alpha^{hl}-\alpha^{ll})} + \underline{w}$ *for each worker. If* $T_1 > T_2$, *(the degree of complementarity in the production function is sufficiently high) and* $x_1 - x_2 < T_1$, *then the firm induces high effort in the high ability worker and low effort in the low ability worker. The corresponding expected wages are* $\alpha^{hl}\frac{c}{(\alpha^{hl}-\alpha^{ll})} + \underline{w}$

to the high ability worker and $\underline{w}$ to the low ability worker. If $T_2 > T_1$ and $x_1 - x_2 < T_2$ then the firm induces low effort in both types of workers. The corresponding wages are $\underline{w}$ for each worker.

The proof is obvious.

B.2 With social networks

Recall the utility function, (6), with social networks. $V(f_i^k|e)$ depends only on the effort level of worker i and $V(f_i^k|e_l) = \underline{V} < V(f_i^k|e_h)$. Suppose the firm wants to induce high effort in both workers. We can re-write the constraints for the maximization problem of the firm, (2) as follows:

(1) the PCs:

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

which can be re-written as:

$$\alpha^{hh}w_1 + (1-\alpha^{hh})w_2 \geq c + \underline{w} - \frac{(1-\gamma)}{\gamma}(V(f_i^k|e_h) - \underline{V}) \tag{B.2.2}$$

(2) The ICs

$$\gamma(\alpha^{hh}w_1 + (1-\alpha^{hh})w_2 - c) + (1-\gamma)V(f_i^k|e_h) \geq \gamma(\alpha^{lh}w_1 + (1-\alpha^{lh})w_2) + (1-\gamma)\underline{V} \tag{B.2.3}$$

which can be re-written as:

$$(\alpha^{hh} - \alpha^{lh})(w_1 - w_2) \geq c - \frac{1-\gamma}{\gamma}(V(f_i^k|e_h) - \underline{V}) \tag{B.2.4}$$

and

$$\gamma(\alpha^{hl}w_1+(1-\alpha^{hl})w_2-c)+(1-\gamma)V(f_i^k|e_h) \geq \gamma(\alpha^{ll}w_1+(1-\alpha^{ll})w_2)+(1-\gamma)\underline{V} \quad \text{(B.2.5)}$$

which can be re-written as:

$$(\alpha^{hl}-\alpha^{ll})(w_1-w_2) \geq c - \frac{1-\gamma}{\gamma}(V(f_i^k|e_h)-\underline{V}) \quad \text{(B.2.6)}$$

and (3) the LL constraint: $w_1, w_2, w_3 \geq \underline{w}$

Denote $\frac{1-\gamma}{\gamma}(V(f_i^k|e_h)-\underline{V}) = K$. Then the inequalities (B.2.1) to (B.2.6) are the same as inequalities (B.1.2) to (B.1.7) except for the RHS which is now lower at $c-K$. Suppose the firm wants to induce low effort by both workers. There are no incentive constraints. Since $V(f_i^k|e_l) = \underline{V}$ the wages that satisfy the participation constraint are $w_1 = w_2 = \underline{w}$. Below we assume $c > K$ to ensure that the bonus for high effort is positive.

Let $\tilde{T}_1 \equiv \frac{2\alpha^{hh}-\alpha^{hl}}{2(\alpha^{hh}-\alpha^{hl})}\frac{c-K}{\alpha^{hl}-\alpha^{ll}}$ and $\tilde{T}_2 \equiv \frac{\alpha^{hh}}{\alpha^{hh}-\alpha^{ll}}\frac{c-K}{\alpha^{hl}-\alpha^{ll}}$. This proves Proposition (2), below:

Proposition 2 *Assume that the firm makes positive profits when low effort is induced for both workers, i.e. $\mu \geq \underline{w} - \pi_{l,l}$ and $c > K$. The firm induces high effort in both workers iff $x_1 - x_2 \geq \max(\tilde{T}_1, \tilde{T}_2)$. Expected wages are $\alpha^{hh}\frac{c-K}{(\alpha^{hl}-\alpha^{ll})} + \underline{w}$ for each worker. If $\tilde{T}_1 > \tilde{T}_2$, (the degree of complementarity in the production function is sufficiently high) and $x_1 - x_2 < \tilde{T}_1$, then the firm induces high effort in the high ability worker and low effort in the low ability worker. The corresponding expected wages are $\alpha^{hl}\frac{c-K}{(\alpha^{hl}-\alpha^{ll})} + \underline{w}$ to the high ability worker and $\underline{w}$ to the low ability worker. If $T_2 > T_1$ and $x_1 - x_2 < T_2$ then the firm induces low effort in both types of workers. The corresponding wages are $\underline{w}$ for each worker.*