Religious Divisions and Production Technology: Experimental Evidence from India

Arkadev Ghosh *

July 17, 2023

Abstract

This paper implements a field experiment in India to understand whether the effects of religious diversity on productivity and attitudes depend on a firm's production technology. I randomly assigned Hindu and Muslim workers at a manufacturing plant in West Bengal to religiously mixed or homogeneous teams. Production tasks are categorized as high- or low-dependency based on the degree of continuous coordination required for production. I find that mixed teams are less productive than homogeneous teams in high-dependency tasks, but this effect attenuates completely in four months. In low-dependency tasks, diversity does not affect productivity. Despite lowering short-run productivity, mixing improves out-group attitudes for Hindu workers in high-dependency tasks, but there are little or no effects in low-dependency tasks. The improvements in production and attitudes in high-dependency tasks are consistent with the minority (Muslim) workers initiating and paying the cost of integration. Overall, this pattern of results suggests that technology that incentivizes individuals to learn to work together is important in overcoming existing intergroup differences – and leads to improved relations and team performance.

*Duke University. Email: arkadev.ghosh@duke.edu. I am indebted to Siwan Anderson, Munir Squires, Patrick Francois, Matt Lowe, and Jamie McCasland for their advice, support and guidance throughout the project. I would like to thank Anand Chopra, Claudio Ferraz, Ashok Kotwal, Gaurav Khanna, Elizabeth Lyons, Salma Mousa, Nathan Nunn, Daniel Rees and Thorsten Rogall for useful suggestions and comments. Thanks also to Aaheli Chakraborty, Mouiry Gope, Abhirup Majumder and Subhaman Supantha for excellent research assistance. This research was undertaken thanks to funding from the Centre for Innovative Data in Economics Research (CIDER) under the Canada Excellence Research Chair held by Prof. Erik Snowberg. This project received ethics approval from UBC BREB (Number: H19-00729), and was pre-registered in the AEA registry (ID #0004270).

1 Introduction

Evidence suggests that ethnic diversity can lower firm output due to weak social ties and taste-based discrimination among workers (Becker, 1957; Lazear, 1998; Hjort, 2014).[1] But how do these effects differ across production technologies that require different degrees of worker coordination? What are the dynamic effects of diversity in firms on productivity and intergroup relations? These are fundamental questions in understanding how firms respond to the costs of diversity. If managing a diverse workforce imposes large costs, firms may limit hiring to minimize inter-ethnic interactions, or segregate workers, perpetuating discrimination. But these market distortions could be avoided if the negative effects of diversity are mitigated in the long-run through repeated intergroup contact or through the adoption of appropriate production technology.

This paper contributes to our understanding of these issues by implementing a field experiment to estimate the short- and longer-run effects of religious diversity on team productivity and intergroup relations under different production technologies. To this end, I partnered with a processed food manufacturing plant in West Bengal, India that employs both Hindus and Muslims — the two main religious groups who have a long-standing history of conflict in India (Pillalamarri, 2019). Production tasks at the firm can be categorized into the following two types depending on the nature of contact between workers: High-Dependency (HD) and Low-Dependency (LD). This classification is based on the degree of coordination required amongst workers performing a task to ensure uninterrupted production, and the dependence on teammates for breaks. Worker effort choices have a higher degree of complementarity in HD tasks than in LD tasks, where workers are required to coordinate intermittently.[2]

There are two key features of my research design that are important for identification. The first is that I randomly assign 586 male workers to religiously mixed or Hindu-only production teams. The second is that the firm follows a quasi-random method of assignment of workers to

[1] There is a large literature on the negative effects of ethnic diversity in decision making in the public sphere as well (Easterly and Levine, 1997; Alesina and Spolaore, 1997; Miguel, 2004; Alesina and Ferrara, 2005). At the same time, diversity has been shown to have positive economic outcomes too – due to strategic complementarities in interacting with out-group individuals (Artiles, 2020; Montalvo and Reynal-Querol, 2017; Jha, 2013) or under certain specific requirements of ethnic interaction imposed by authority (Bhalotra et al., 2018; Marx et al., 2021).

[2] An example of a HD task is work on a fast moving conveyor belt where each worker is responsible for collecting every second or third piece of a product on the belt. Even if only one of them cannot keep up, the machine speed needs to be reduced affecting the productivity of all workers. An example of a LD task is work in a mixing room. Workers typically have well-defined individual duties: for example, one worker is responsible for ensuring that raw materials are weighed properly, another one is entrusted with arranging flour buckets while a third worker mixes the raw materials. The workers need to coordinate intermittently and the productivity of one worker does not directly or immediately influence other workers. A detailed description of HD and LD tasks follows in section 2.

production tasks.[3] Taken together, they allow me to attribute potentially different effects of religious mixing in HD and LD tasks to production function differences, as opposed to differences in worker types in these tasks.

Each production line at the factory comprises a series of (HD and LD) team tasks. I designed the experiment such that after randomization there were just two types of production lines: (1) those with mixed teams in HD tasks only (HD-Mixed lines) and (2) those with mixed teams in LD tasks only (LD-Mixed lines). I can therefore estimate the effects of religious mixing on line-level output (HD-Mixed line vs. LD-Mixed lines), as well as on team performance at the task-level (mixed teams in HD (LD) tasks vs. non-mixed teams in HD (LD) tasks). I kept the randomized teams intact for a period of four months in order to estimate *dynamic* effects.

The experiment uncovers three key findings. The first is that religious diversity negatively affects team output, but *only* in HD tasks. Overall, HD-Mixed lines produce 5% lower output than LD-Mixed lines during the experiment. An analysis of performance measures at the task-level reveals that this loss is entirely attributable to mixed teams in HD tasks. In LD tasks, the effect religious diversity on productivity is small and not statistically significant. The second key finding is that the difference in output between HD-Mixed and LD-Mixed lines attenuates significantly over the treatment period – from greater than 20% at the beginning of the experiment, the effect reduces to less than 1% by the end of the fourth month. This is driven entirely by output gains in mixed HD teams. The third key finding is that, at endline, there is a reduction in negative out-group attitudes for Hindu workers, which is substantially (21%-50%) larger from mixing in HD teams compared to LD teams. This is despite the fact that mixed HD teams suffered negative output shocks. In LD teams, mixing has little or no effects on attitudes of Hindus.

There are several plausible explanations for these core findings. Since there are no Muslim-only teams in this study, one might worry that these results are driven by productivity differences between Hindus and Muslims. In particular, if Muslims have lower productivity, the treatment effects could simply reflect differences in average productivity between mixed and Hindu-only teams. A number of results and additional tests suggest that this is unlikely. First, if Muslims were less productive overall (in both HD and LD tasks), we would expect mixing to reduce productivity in LD tasks too. Second, the rapid fall over time in the treatment effect of mixing in HD tasks is unlikely if Muslims were particularly unproductive at these tasks. Third, in the task-level analysis, I explicitly control for observable (and potentially confounding) dif-

[3]The HR manager keeps a pool of job applicants who are assigned to tasks on a first-come-first-served basis when vacancies become available — workers do not get to choose their task when they join or over their tenure. A detailed description of this process and tests to check its validity are presented in section 3.4 and Appendix A.3.

ferences between Hindus and Muslims (such as schooling and tenure), as well as between HD and LD tasks (such as team size). The results are not affected in any way. Finally, I show that at baseline (conditional on observables) Hindus and Muslims were equally likely to be promoted from being unskilled to semi-skilled or as a machine operator. This suggests that the firm does not perceive them to be differentially productive either. The null effect of religious diversity on productivity in LD sections[4] further rules out other explanations based on social reputation concerns around in-group members (Afridi et al., 2020; Bursztyn and Jensen, 2017) or strong distaste for out-group members (Hjort, 2014). Even though worker efforts have a lower degree of complementarity in LD tasks, teams are still required to coordinate on many aspects.

I develop a conceptual framework and instead argue that the most plausible explanation for the findings here is that Hindus have lower priors regarding how hardworking their Muslim co-workers are, relative to in-group Hindu co-workers. But Muslim workers do not make this distinction. This is because of the asymmetry between Hindus and Muslims in their exposure to non-coreligionists at baseline. Consistent with majority-minority relations, Muslims are always in mixed teams with Hindus, while a large section of Hindu workers in the firm do not work with Muslims.[5] This leads to Muslims having accurate priors about Hindus, but Hindus (depending on past exposure) not necessarily having accurate priors about Muslims. In HD tasks with complementary worker efforts, Hindus optimally choose low effort based on the low initial prior about their Muslim co-workers, leading to low team output.[6] Hindu workers do update their beliefs about Muslims and forward-looking Muslim workers internalize this behaviour of Hindu workers. Given a long enough interaction period, Muslims exert high effort despite the fact that Hindus initially exert low effort. This follows because Muslims can persuade Hindus to eventually exert high effort as the latter begin to observe *greater* realizations of high output days

[4]During a period of religious tensions in West Bengal following the passing of the Citizenship Amendment Act (CAA) and subsequent riots in New Delhi, I find religious diversity to have negative effects in LD tasks too. This rules out that mixing in LD tasks is simply a placebo treatment where there are no interaction among workers. Instead, the production technology is such that output is less sensitive to frictions amongst workers. However, consistent with Hjort (2014), I find that extreme events can create strong distaste for outgroup members affecting team productivity (see Table B.20.)

[5]In factories and other formal workplaces across India, Muslims are generally used to working alongside Hindus, while a large share of Hindus are not used to working with Muslims. In this firm, roughly 50% of the Hindu workers worked in homogeneous teams at baseline, while all Muslim workers worked alongside Hindus. Similarly, 43% of Hindus reported to have no contact with Muslims outside of work, whereas only 9% of Muslims reported the same about Hindus. Based on this, together with evidence on discrimination against Muslims in access to education and labor markets in India (Kalpagam et al., 2010; Basant, 2007), I assume Hindus on average (mistakenly) have lower priors regarding how hardworking their Muslim co-workers are, relative to in-group Hindu co-workers. Of course, I show evidence that Hindus and Muslims are not differentially productive in section 6.

[6]In LD tasks, worker efforts are assumed to be non-complements whereby the effort levels of Hindu workers are not dependent on their priors about Muslims. As a result, team output is not affected by diversity.

than expected under low effort from their Muslim teammates, and as a result gradually update their beliefs. By bearing this short-run cost, Muslim workers benefit from a high-output equilibrium in the long-run.[7] Consistent with this mechanism, I first show that relative to Hindus, Muslims in general are more likely sacrifice their relief/break time (an important marker of individual effort) for teammates. Second, I find that they disproportionately allocate such higher effort towards their Hindu teammates. This is despite the fact that (at least initially) Hindus are more likely to blame low output on Muslims (consistent with their initial low prior about Muslim capabilities). I also test for heterogeneity in the effects by baseline experience of working with outgroup workers: I find that teams in which Hindus have had greater past contact with Muslims, suffer little to no losses. The effects are driven by teams in which Hindus have little past experience with Muslims. At the same time, teams that include Muslims who have experienced "similar" religious diversity in the past as in the experiment (i.e. less than 80% Hindus in a team), suffer smaller loses and experience more rapid adjustment. But, teams where Muslims have had exceptionally high exposure to Hindus in the past perform worse. This is likely because the priors of Hindus only matter when the share of Muslims in a team is sufficiently high and past experience of Muslims workers in such teams allows them to adjust more quickly during the experiment. These dynamics are also consistent with the explanation above and *inconsistent* with Muslims being less productive.

The policy implications of my findings hinge crucially on whether firms are aware of the costs of religious diversity, and how they depend on the production technology. To explore this, I surveyed more than one hundred production supervisors across five different firms that produce similar products. I asked them to predict the results of my experiment and about ways to mitigate possible negative effects of religious divisions. They correctly predicted that religious mixing would be more costly in HD tasks than in LD. But despite the possibility of losses, the majority of supervisors reported to be averse to segregation of workers by religion.[8] About a quarter of the supervisors correctly cited negative effects of diversity dissipating with repeated intergroup contact, but the first-order concern was about such segregation potentially causing tensions. These findings suggest that effective policy design in this context must look beyond just the direct effects of diversity on production and also trade-off potential short-run costs for

[7] Note that if the interaction period is not sufficiently long, then the minority group (Muslims) does not invest in the majority group. This is because there will not be enough periods of high-output payoff to recover the loss that the minority group suffers initially by exerting high effort, even as the majority group exerts low effort.

[8] Note that having religiously mixed and Hindu-only teams at the individual task-level (as in the experiment) is natural in this context – because Hindus comprise 80% of the population and each task requires five to six workers on average (see Figure 1). Supervisors showed concerns about complete segregation of workers by religion on the production floor i.e., having only all-Hindu and all-Muslim teams.

long-run benefits of integration.

This paper contributes to work on ethnic diversity and firm production. A number of papers document negative effects of diversity on productivity (Hjort, 2014; Afridi et al., 2020; Parrotta et al., 2012; Hamilton et al., 2012; Churchill et al., 2017). Hjort (2014) exploits quasi-random variation in the ethnic composition of teams in a Kenyan flower plant and finds that ethnically mixed teams have lower productivity due to taste-based discrimination. Afridi et al. (2020) use variation in the caste composition of teams in Indian garment factories and show that caste homogeneity boosts productivity. The identification strategy in these studies mainly relies on frequent team switching (due to HR policies) leading to short-term variation in the ethnic composition of teams.

To the best of my knowledge, this is the first paper to randomly assign firm workers across a wide *range of tasks* to long-term teams,[9] and estimate the *dynamic* effects of repeated intergroup contact on team production and social preferences in the *same* setting. Secondly, I show how differences in the incentives to interact with co-workers due to *production function differences*, affect team productivity. My results suggest a potential tension between the goals of maximizing short-run productivity and improving inter-group relations. But they also emphasize that intergroup contact over time can bring about improvements to productivity and attitudes, as the minority group invests in shifting priors of the majority group.[10] The disincentive to invest in out-group members in short-term interactions may explain why previous studies do not find that a history of being in mixed teams reduces discrimination and improves productivity.

Evidence from lab experiments suggest that under group incentives, team homogeneity (in social identity) leads to more cooperation and efficient outcomes in public goods (Eckel and Grossman, 2005) as well as minimum effort games (Chen and Chen, 2011), and also causes better norm enforcement (Goette et al., 2006). I contribute to this literature by showing that team incentives shaped by a firm's production technology can induce cooperation between religious groups in a real-world team production setting.

My paper also adds to work on social preferences at the workplace (Bandiera et al., 2010, 2013; Mas and Moretti, 2009; Carpenter and Seki, 2011; Hjort, 2014; Ashraf and Bandiera, 2018). I show that in the Indian context, factory workers discriminate against non-coreligionists leading to output losses. A particular focus of this literature has been on studying team performance under different incentive structures. The plant I study offers a flat monthly wage to its employ-

[9] Ideally, I would have liked to observe outcomes for a longer period of time since the firm did not want to change the teams again. Unfortunately, the experiment could not continue because of COVID-19 related restrictions.

[10] Interestingly, Baggio and Cosgel (2023) observe that the adverse effects of racial diversity on team performance in the American whaling industry diminish over time, indicating a crew's gradual adaptation to diversity.

ees based on seniority and experience at the firm. I find that even without explicit daily pay incentives, social relations at the workplace can have large effects on team productivity.

This paper also relates closely to the literature on how social preferences are formed through inter-group contact (Fershtman and Gneezy, 2001; Boisjoly et al., 2006; Jakiela et al., 2011; Kato and Shu, 2016; Mousa, 2018; Rao, 2019; Corno et al., 2022[11]) and how its effects depend upon the *type and nature* of contact (Allport et al., 1954; Pettigrew et al., 2011; Bazzi et al., 2017; Paluck et al., 2019). Lowe (2021) shows that intergroup contact has different effects by creating collaborative versus adversarial in a sport setting. Bursztyn et al. (2021) demonstrate that long-term, natural interactions with Arab-Muslim immigrants lead to improved attitudes and behavior towards the group among U.S. natives. I use naturally occurring variation in *types of collaborative contact* (due to production function differences), and by estimating dynamic effects over a four-month period, show that there is a trade-off between maximizing short-run productivity and improving inter-group relations.

Finally, the literature on employer learning (Farber and Gibbons, 1996; Altonji and Pierret, 1998; Altonji and Pierret, 2001; Lange, 2007; Macchiavello et al., 2020) argues that if firms discriminate among workers based on easily observable characteristics (such as race, gender), then as employers begin to observe (noisy) indicators of workers' performances, the initial information should gradually become redundant. I show that such discrimination exists amongst co-workers in production teams, and while stable in the short-run it gradually unravels over time.

The rest of the paper is organized in the following manner. Section 2 describes the context: Hindu-Muslim relations in India (in brief) and the study firm: its workers, as well as high- and low-dependency tasks. I discuss the research design and data, and present balance checks in Section 3. Section 4 presents the econometric specifications used. The results and robustness checks are presented in Section 5. In Section 6, I discuss plausible mechanisms behind the core findings, and describe an outline of a conceptual framework (the model is presented in Appendix D) for the favored mechanism, and provide some subsequent empirical support. Section 7 discusses some policy implications. Finally, section 8 concludes.

[11] Corno et al. (2022) study the effects of interracial interaction not only on stereotypes but also on academic performance.

2 Context

2.1 Hindu-Muslim relations in India

Hindus form the majority of the Indian population (79.8%), while Muslims are the largest minority (14.2%) group (Census, 2011). Hindu-Muslim conflict has plagued India for centuries and has been a recurring phenomenon since partition and independence in 1947 when the country was divided on religious lines – an episode which itself was marked by large scale religious violence (Talbot and Singh, 2009). Muslims have since suffered greater discrimination and violence against them, as well as borne larger economic losses due to such tensions (Mitra and Ray, 2014). Across the country, Muslims continue to lag behind Hindus on various economic indicators including income and education (Asher et al., 2018; Bhattacharjee and Chaudhuri, 2022), face social exclusion (Alam, 2010) as well as discrimination in the labor market (Kalpagam et al., 2010; Khan, 2019) due to their minority status. Hindu-Muslim relations have especially deteriorated in West Bengal recently as local state politics has seen significant polarization on religious lines (Nath and Chowdhury, 2019).

The share of Muslim population varies greatly across states and districts in India. Muslims constitute roughly 25% of the population in the district where my partner factory is located: this is close to the share of Muslims in the factory itself, as well as in other manufacturing plants in the area. Therefore, in terms of representation of Muslims, the factory resembles the average manufacturing plant in the area.

2.2 The Factory: Production lines and worker characteristics

In this section, I describe the factory: the structure of production lines and sections, HD and LD tasks, as well as the operation of shifts. I also discuss the pay structure of workers and report characteristics of the workers by religion.

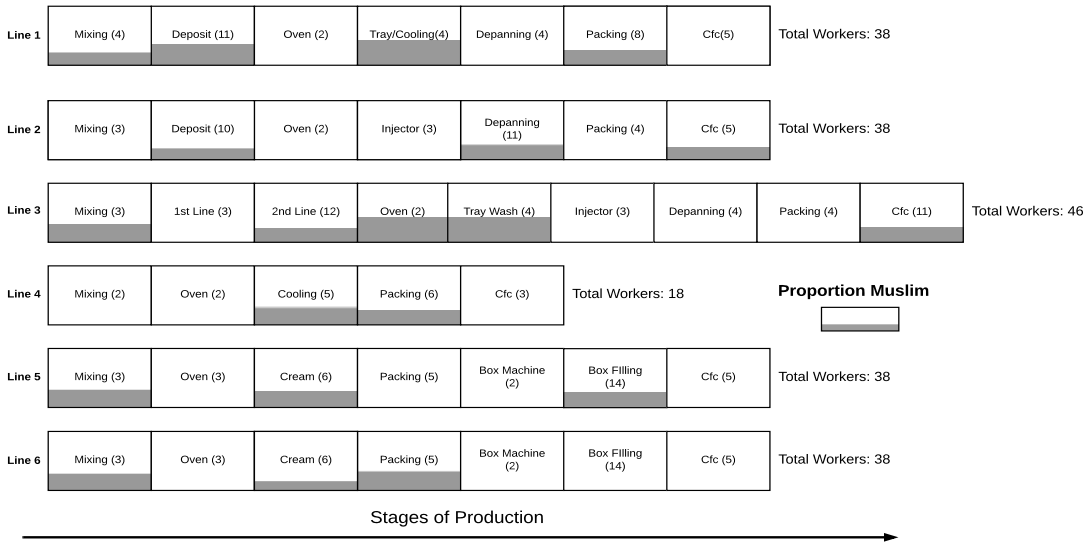
Production lines, sections and shifts

The factory produces packaged bakery products. There are six production lines in total, each of which produces a different product. Figure 1 illustrates the structure of the production lines. Each line is sub-divided into sections (small blocks in the figure) *based on the production task* that is undertaken in that section. The numbers in parenthesis denote the count of workers in each of these sections.[12] Production occurs in three different shifts: morning, afternoon and

[12] Some of the production lines can produce multiple products and these numbers can vary (though only very

night. There are three cohorts per production line, who *as a team* rotate shifts on a weekly basis.[13] As a result, workers have fixed teams at both the line-level and line-section-level i.e., their co-workers do not typically change, only their shift of work as a team changes weekly (each worker therefore has to do morning, afternoon and night shifts on a weekly basis).[14]

Figure 1: Structure of production lines



Note: This figure shows the structure of all six production lines in the factory. The numbers in parentheses denote the count of workers in each section per cohort. Each production line has three cohorts working on it in each of the three shifts in a day. The color shades denote the proportion of Muslim workers in each section in one particular cohort at baseline.

Religious composition of production lines

Table 1 reports the proportion of Muslim workers in each line-level team across the three cohorts at baseline. Line 4 only has two cohorts while all the other lines have three cohorts each. While there is variation in the proportion of Muslims across teams, it is clear from this table that Hindus and Muslims are not segregated in particular lines or cohorts in the factory. On average, each line and cohort roughly have between 15%-25% Muslim workers, which is very close to the overall share of Muslims in the factory. This is formally shown in Figure C.6 – I regress

little) depending on the exact product/product variety being manufactured. Figure 1 is based on the number of people in each section during the baseline survey. The numbers during the intervention were slightly different for some sections.

[13] Teams move from morning to night to afternoon shifts.

[14] Occasionally workers are moved across shifts and lines. This is determined by worker absenteeism and turnover.

a dummy variable denoting a worker's religion on line and cohort fixed effects and show that balance in religious composition of production lines and cohorts cannot be rejected. In section 3.4, I provide evidence that this did not happen simply by chance, but in fact is a result of the firm's policy of allocating workers to tasks in a quasi-random manner.

Table 1: Proportion Muslim by line-level team and cohort (at baseline)

Line	Cohort 1	Cohort 2	Cohort 3	Average
Line 1	0.30	0.23	0.19	0.24
Line 2	0.11	0.07	0.27	0.15
Line 3	0.24	0.24	0.13	0.20
Line 4	0.22	0.26	-	0.23
Line 5	0.15	0.19	0.07	0.14
Line 6	0.14	0.06	0.23	0.14
Average	0.19	0.18	0.18	0.18

Note: Each production line (apart from Line 4) has a total of three cohorts working in it (one in each of the three daily shifts). This table reports the share of Muslim workers in each line-cohort (across all sections) at baseline. Please note the total number of workers in each line-cohort is shown in Figure 1.

The fact that Muslims are in a minority, together with the structure of production lines that require small section-level worker teams within lines, means that a large section of Hindu workers have little or no contact with their Muslim counterparts. This can be observed in Figure 1, where the religious composition of production sections of all six lines is shown for one particular cohort. A large number of sections (close to 50%) have no Muslim workers at all. The share of Muslim workers in most of the other sections is between 0.1 and 0.3. The composition is similar across the other two cohorts as well. This is important for two reasons. First, the degree of inter-religious contact induced by the treatment (60% Hindus and 40% Muslims in mixed teams) represents a significant change from the baseline level of contact for Hindus. Second, the majority-minority asymmetry in exposure to non-coreligionists at baseline might mean that Hindus and Muslims behave differently when randomized into mixed teams.

Pay structure of workers

Workers at the factory are paid a flat monthly wage based on their experience and level of expertise (skill) on the job. Wages are not dependent on daily team productivity but performance is evaluated frequently; poor performance over a period of time can lead to workers being moved

to a lower skill group. Alternatively, performing well can lead to promotion. Workers are categorized into unskilled, semi-skilled and operator groups. Approximately 77% of the workers are unskilled, 14% are semi-skilled, and only around 9% are operators. Semi-skilled workers undertake the same tasks as unskilled workers, while operators are in charge of handling machines.

Characteristics of Hindu and Muslim workers

Summary statistics of worker characteristics are reported in Table B.1. It is clear that workers are not sorted into HD and LD jobs based on their religious identity. There are however important differences between Hindus and Muslims. Muslim workers have lower schooling, as well as lower tenure at the factory. It has been documented in other studies as well that Muslims on average tend to have lower education relative to Hindus in India (Bhaumik and Chakrabarty, 2009). The difference in average tenure however might be surprising. This can be explained by the fact that in the district where the factory is located, Muslims have traditionally been tailors, which many families still continue to pursue as their business. Since families in this region are typically well-connected, this network allows Muslims to work in the informal tailoring sector, providing them with an outside option of employment. The management often cited this as a factor behind the larger turnover of Muslim workers.

Muslim workers report having much greater contact with Hindus outside of work (as well as at work), which is expected given that Hindus form the majority group in the study area and across India in general. Consistent with this, Muslims report to be more comfortable than Hindus when it comes to communicating with non-coreligionists. Surprisingly, both groups report to be equally uncomfortable taking orders at work from non-coreligionists. Finally, as shown in Table B.1, Hindus are much more likely to support the controversial National Registrar of Citizens (NRC), a bill which is often criticized for discriminating against Muslims.[15]

2.3 Direct Dependency as a measure of production technology

Direct Dependency is defined as the degree of continuous coordination (instantaneous and physical) required amongst workers performing a task to ensure uninterrupted production. I study it as the key aspect of production technology for two main reasons. First, a key distinction between high- and low-dependency tasks relates to a core idea in economics: the degree of complementary of labor inputs. Worker efforts have a high degree of complementary in HD

[15]The NRC is a list of people who can prove that they came to India before 24th March, 1971. It is a widely held view that together with the Citizenship Amendment Act (CAA), the NRC may be discriminating against Muslims (Chapparban, 2020).

tasks, while they have a lower degree of complementarity in LD tasks. Second, the degree of complementarity in labor inputs affect incentives to interact, suggesting that this might matter for the effects of religious divisions. Some key characteristics of high- and low-dependency sections (or tasks) are listed in Table 2. Figure C.4 provides a visual illustration of HD and LD tasks and figure C.5 shows all six production lines at the factory broken down into HD and LD sections.

Table 2: Characteristics of High- and Low-Dependency tasks

Work condition	**High-Dependency (HD)**	**Low-Dependency (LD)**
Task coordination	High and Continuous	Low and Intermittent
Control over breaks	Low	High
Physical mobility	Restricted	Good
Repetitive monotony	High (Machine Speed)	Low (Occasionally paced by machine)

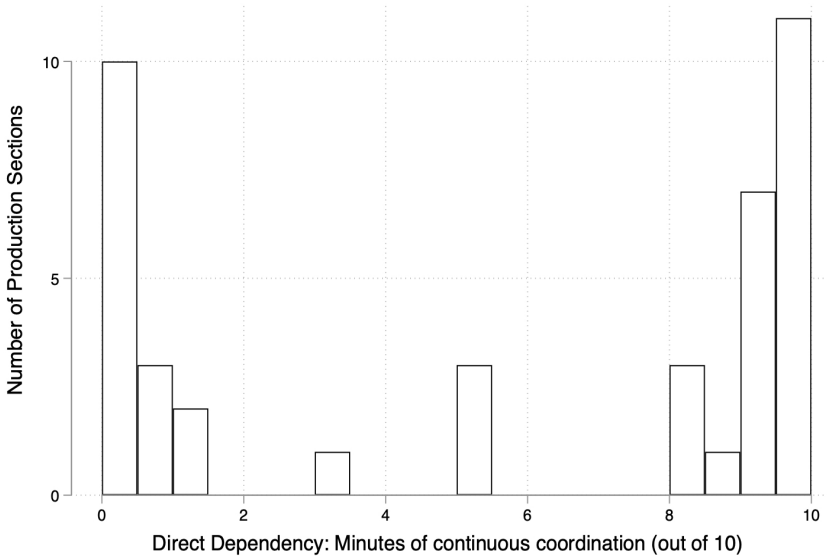
Note: This table lists some key differences in characteristics between High- and Low-Dependency tasks.

Task coordination

The first and the most important distinction between high- and low-dependency tasks is in the amount of coordination required amongst co-workers. I quantify this with time-use data. Research assistants recorded minutes (out of ten) of direct physical coordination (as opposed to working independently) required amongst workers for production to continue without interruption in each section. HD sections typically require workers to coordinate continuously for nine to ten minutes (out of ten), whereas the average in LD sections is only two minutes. Sections *above* the median value (≥ 9) on this scale are classified as HD sections and those *below* as LD sections.[16] The distribution of Direct Dependency is shown in Figure 2. Most tasks require either high continuous coordination (nine or ten minutes out of ten) or less than two minutes of continuous coordination – which leads to the bi-modal distribution in the figure. This allows easy classification of tasks into HD and LD types, an important (third) reason to pick this measure over others.

[16] In Table B.9 I consider various various alternative cut-off values and show that the results are not sensitive to this choice.

Figure 2: Distribution of Direct Dependency



Note: This figure shows the distribution of Direct Dependency. Enumerators visited every section of each production line and took stopwatch measures of the number of minutes (out of 10) for which workers were continuously dependent on each other for production to occur. The figure is generated from these stopwatch records at the line-section-level.

Control over breaks/relief time

The second key distinction between HD and LD tasks is about control over breaks during the production process. Due to dependence on co-workers every minute of the production process, each worker individually has little control over when they can take a break in HD tasks. Sub-groups of workers need to provide “relief" to other workers – a concept known as “relief time". There are often disagreements amongst workers regarding how to schedule these as well as arguments when some workers take more time than allocated. Supervisors reported such disruptions to be a common cause for lower productivity. By contrast, in LD sections each worker has much greater control over scheduling breaks, though the concept of relief time still exists.

Physical mobility

Physical mobility is restricted in HD sections. For example, workers are typically required to stand close to each other on conveyor belts and pick products up as they move on the belt. Coordination with others doing the same is therefore key. In LD sections, greater individual control over the production process allows workers greater physical mobility.

Repetitive monotony

Repetitive monotony is higher in HD sections compared to LD sections since work cycles are shorter. The machine speed set by the supervisor often determines the speed of work, allowing workers little control over the process. If workers do not perform up to the mark, supervisors may need to reduce machine speed causing loss in output. Informal interviews with the supervisors made it clear that it is not uncommon for them to vary machine speed in these areas. This could happen due to worker absenteeism leading to changes in teams, as well as due to workers simply not coordinating as expected on certain days of production. In LD sections, workers typically have more control over process speed, and can re-allocate their time across different sub-tasks to a greater extent.

Direct Dependency and other task-level characteristics

In Table B.2, summary statistics of various aspects of the physical environment of HD and LD sections are presented. I focus on factors that could serve as potential confounders to production function differences explaining my results. I measure the degree of non-work interaction (time workers spend chatting), temperature and noise levels in each section of each production line and rule out that HD and LD sections are systematically different on these aspects of the physical work environment.

Worker status

One might be concerned that LD tasks provide higher status and self-esteem (Bursztyn et al., 2017) thereby lowering intergroup conflict. However, we will see in section 5.2 that frictions amongst workers happen at a similar rate in HD and LD tasks. Despite this, team output is less sensitive to these frictions in LD tasks, likely due to the nature of production.[17] Furthermore, there are no explicit pay or skill-designation differences between HD and LD tasks, suggesting that status or self-esteem is unlikely to play a major role in this context.

3 Research design

This section discusses the research design. I first go through the randomization process and then present balance checks over a range of worker characteristics across the different treatment arms. Before the intervention, workers were informed that their teams would be changed

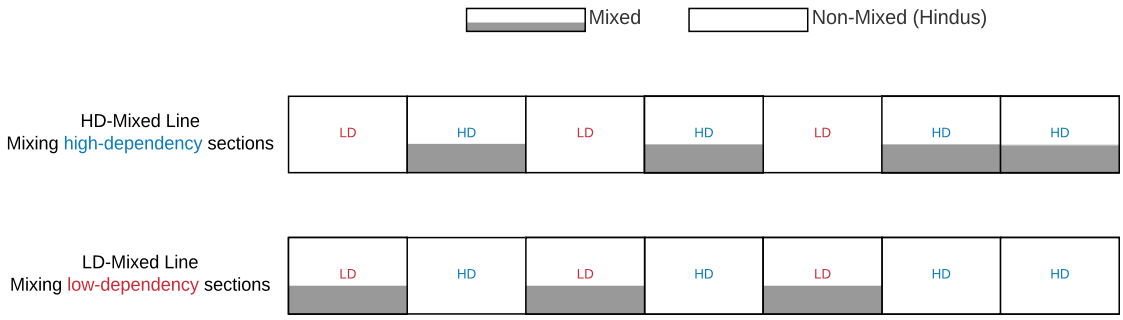
[17] Anecdotally, there are various aspects of LD tasks, such as heavy lifting in certain sections or working in mixing rooms that have unpleasant smell which workers reported to dislike.

in order to assess the effect of team-switching on firm performance. The new team lists (post randomization) were printed and posted on the production floor. Religion of teammates could be directly inferred from their names in this context (no additional information was provided to make this aspect salient).[18]

3.1 Treatment and randomization

As mentioned earlier, the factory operates in three shifts (morning, afternoon, night) and an entire cohort of workers moves from one shift to the next on a weekly basis. A new set of workers come to work in each shift on a particular day. Therefore, each line has three different cohorts working on it each day of the week. For the purpose of randomization, I moved workers across cohorts, holding their production line and section of work fixed.[19]

Figure 3: Randomized team structure



Note: This figure shows the two different types of line-level teams after randomization. Sections are partially shaded to denote religiously mixed teams. HD-Mixed lines had all their HD sections mixed and LD sections non-mixed. The opposite is true for LD-Mixed lines.

Individual workers were randomized into line-section-level teams in order to achieve two distinct types of teams (treatments) at the line-level. The first type comprised of line-level teams with religiously mixed groups only in HD sections (HD-Mixed lines), while the second type had religiously mixed groups only in LD sections (LD-Mixed lines). Two of the randomized cohorts within each line were of one team type while the third cohort was of the other type. Figure 3

[18] Whether a person is a Hindu or a Muslim can be determined from their first name itself in the Indian context. In very few cases where the first name maybe ambiguous, the last name would certainly reveal one's religion. My sample consists of only Hindus and Muslims.

[19] For a small share (7.9%) of workers this was not the case. Some workers had to be (randomly) moved from their tasks at baseline to achieve the desired line-level team types. However, such task-shifting is not correlated with treatment status. Section 5.3 includes a discussion on this.

provides a visual illustration of the two types. I use Line 2 from Figure 1 for this illustration. Section names are replaced by HD and LD labels to denote section (task) type. The first type of line-level team has all its HD sections mixed (partly shaded in grey) while its LD sections are comprised of only Hindu workers (HD-Mixed line). The structure in the second type is exactly the opposite – LD sections have religiously mixed teams while HD sections have only Hindu workers (non-shaded) (LD-Mixed line).[20] This leads to four different types of line-section-level teams: 1. HD Mixed 2. HD Non-Mixed 3. LD Mixed and 4. LD Non-Mixed. Production data are available for both line-level as well as line-section-level teams. Therefore, any differences in overall line-level performance between teams can be disaggregated to line-section-level performance.

Randomization was constrained by one key limitation – the number of workers switching their section of work (their task) had to be minimized. Even though the induction of workers to specific tasks takes only a few days, it is impossible to train all workers in new tasks simultaneously – this would lead to substantial interruptions and breakdown in production. The management was unwilling to do this. As a result, the randomization process was designed such that did it not require the majority of workers to change their section of work and hence their task type (HD or LD) from baseline. I address concerns with respect to selection of workers (at baseline) into HD and LD tasks subsequently.

The first step in the randomization process involved determining the final (target) number of Hindus and Muslims in each section of each production line (in order to achieve HD-Mixed and LD-Mixed line structures across the three cohorts that would be formed). Since workers were not moved across production lines for randomization, this was typically constrained by the overall number of Hindus and Muslims in a line across the three cohorts at baseline. The share of Muslims in each production line at baseline was close to the overall share of Muslims in the plant (see Table 1). After randomization, the share of Muslim workers in mixed sections (both HD and LD) of all six lines was typically between 35%-40% (this was of course balanced between HD and LD sections).[21]

The second step in the process involved sorting workers by section × religion × skill[22] (pool-

[20] At the line-section-level, religiously mixed and Hindu-only teams are the ones that are naturally formed at baseline (recall Figure 1).

[21] Note that the religious composition of a particular section in a line would be exactly the same across all cohorts if they belonged to the same line-level team type. In other words, if cohorts A and B in Line 1 were such that all their HD sections were mixed and LD sections were non-mixed, then each of their HD sections would have exactly the same ratio of Hindu to Muslim workers i.e. *Packing* in cohort A would have exactly the same number of Hindus and Muslims as *Packing* in cohort B. Non-mixed teams of course only have Hindu workers.

[22] Workers are classified into three skill levels: unskilled, semi-skilled and operator. Each section typically has an operator or a semi-skilled worker (depending on the type of work), and the rest are all unskilled workers. The

ing across all three cohorts in a line) and randomly shifting (some) workers across sections in order to ensure that each section of each line had enough Muslim workers (summing across cohorts) that are required for randomization (as determined in the first step). This had to be done at baseline because not all sections of all lines had enough Muslim workers (sometimes none), such that the desired line-level team structures in Figure 3 could be achieved. For example, the *Injector* section in Line 3 had no Muslim workers at all across the three cohorts. In such cases, some randomly chosen Hindu workers in that section were shifted out and replaced with randomly chosen Muslim workers from another similar section with enough Muslims. This process meant that at the end of step 2, all sections of all lines had both Hindu and Muslim workers[23] who would then be randomly allocated to line-section-level teams. This also satisfied the management's requirement of minimum section (task)-shifting.

Lastly in the third and final step, workers were sorted by their new section (post step 2) × religion × skill and randomly allocated into line-section-level teams in order to achieve the line-level team structures shown in Figure 3. Line-level teams were then randomly allocated to one of the three shifts. A detailed description of each step involved in the randomization process is presented in Appendix A. Figure A.1 provides a visual illustration of the same, especially focusing on how section-shifting allows formation of the desired line-level team structures.

3.2 Data collection, experiment timeline and attrition

Data used for the analysis in this paper come from two main sources. I use administrative records of production obtained directly from the firm's management to estimate treatment effects of diversity on line-level output. The firm records total output at the line-level line in each shift; this measure is tied directly to the revenues of the firm. In addition, supervisors were also trained by the production manager to rate the performance of each line-section-level team. These ratings are used to estimate the effect of diversity on output in HD tasks separately from LD tasks.

Workers participated in an in-person survey at baseline but only a phone survey could be conducted at endline due to COVID-19 related restrictions in India. The baseline survey included a wide set of questions ranging from employment related ones such as tenure, history of past teams, attitudes towards taking orders from and interacting with non-coreligionists, to objective worker characteristics such as age and schooling. I also asked workers about their po-

randomization process did not alter this structure.

[23] This is required because for each section of each line there would at least be one line-level team where that section would have to have a mixed group.

litical preferences, focusing on factors that could capture taste discrimination towards religious groups. These include preference for political parties that are associated with favoring a particular religious group and support for bills that are widely criticized for discriminating against Muslims.

The focus of the endline survey was primarily on interactions (accusations, blame, providing relief time etc.) that happened during the intervention and on worker attitudes that could capture the effects of inter-religious contact in HD and LD environments on inter-group relations. Summary statistics of key variables are presented in Table B.1; differences in characteristics of Hindus and Muslims have already been discussed in section 2.2. Figure 4 presents the timeline of the intervention and sample size by treatment arm. There are 15 line-level teams[24] (7 HD-Mixed Lines and 8 LD-Mixed Lines) and 113 line-section-level teams (23 HD-Mixed, 33 LD-Mixed, 29 HD Non-Mixed and 28 LD-Mixed). A total of 586 workers were part of the intervention distributed in the following way in line-section-level teams: 175 in HD-Mixed, 117 in LD-Mixed, 196 in HD Non-Mixed and 98 in LD Non-Mixed. A total of 546 workers could be reached at endline for the phone survey (attrition rate 6.8%).[25]

3.3 Randomization check

Balance checks in Table 3 show that randomization was successful. Outcomes are divided into two broad categories - (1) those that are relevant at work (Panel A) and (2) general characteristics and attributes (Panel B). The unit of analysis here is an individual. The main regressors are the interaction terms Mixed × LD and Mixed × HD which denote the type of line-section-level team and hence the treatment status of an individual. Line × Section fixed effects are included in these specifications, whereby the main effect of HD versus LD is not separately identified. The omitted group is therefore all workers assigned to non-mixed teams.[26] Across a range of characteristics that include factors that are relevant at the workplace (such as tenure and past

[24] Note that at full capacity the firm would have 17 line-level teams as shown in Table 1. However, in the experiment there are 15 line-level teams only. This is because during the period of the intervention, the firm decided to operate at lower capacity due to low product demand compared to previous years (even though the experiment was timed to coincide with the period when, in terms of seasonality, the firm usually experiences the highest demand). As a result, production lines 1-3 had three cohorts each whereas lines 4-6 only had two cohorts each (Figure 1). This change occurred before the randomization began, so the experiment was not affected by it.

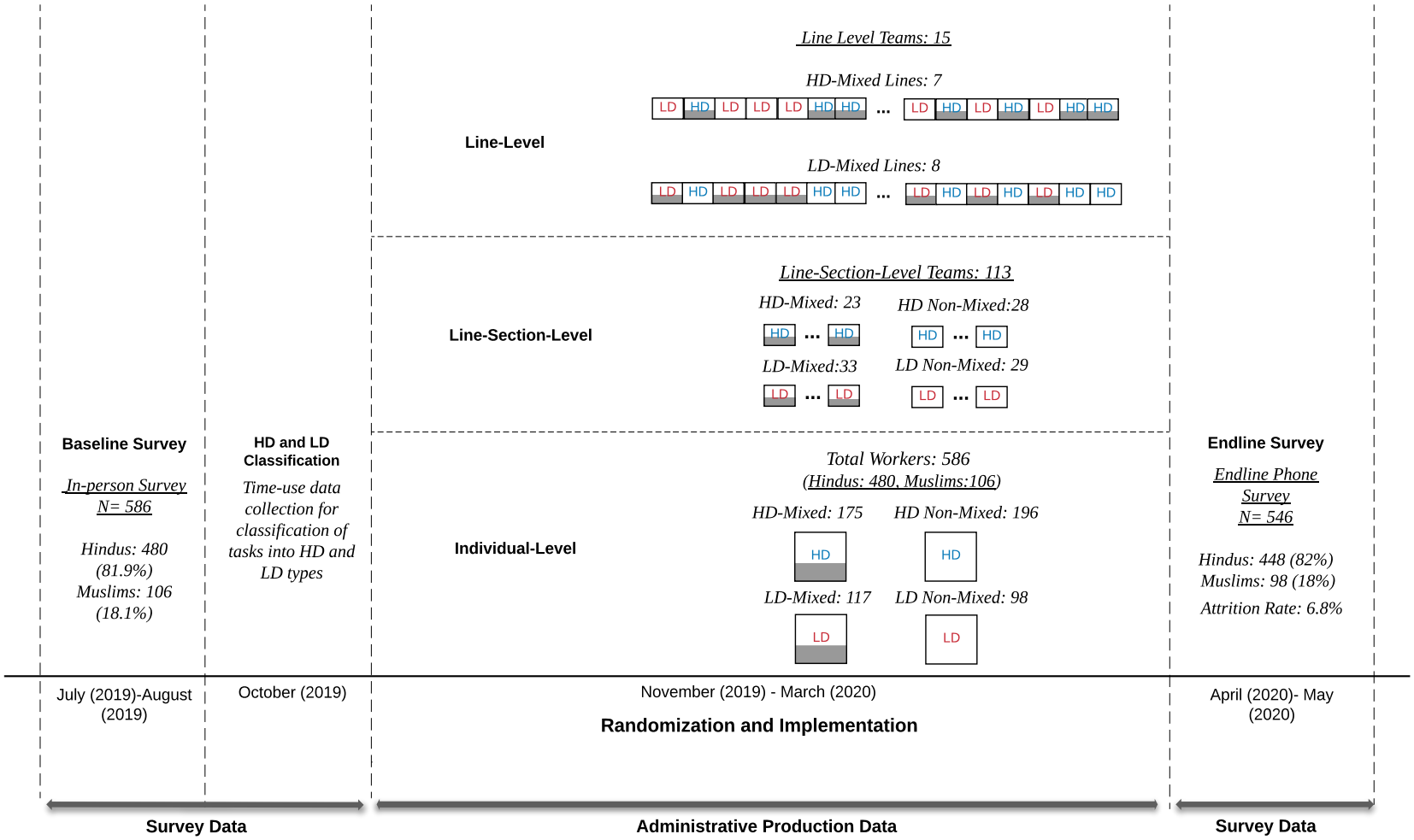
[25] In Table B.21, I show that attrition is balanced across treatment arms.

[26] I use this particular specification for balance checks because the same specification is used to estimate treatment effects at the line-section-level on team production, as well as on individual-level survey outcomes. As a robustness check, I use Line fixed effects instead of Line × Section fixed effects in Table A.4 (whereby the main effect of HD versus LD is identified) and show that worker characteristics are balanced across HD and LD sections. I also show balance in individual characteristics across line-level teams (i.e. HD-Mixed lines versus LD-Mixed lines overall) in Table A.5.

contact with non-coreligionists), as well as general attributes (such generalized trust, altruism and contact outside work), workers are similar across the treatment arms.

Finally, it is also important to show that the proportion of Muslim workers is balanced across mixed HD and LD teams, to rule out that the treatment effects are driven by different “degrees" of religious mixing across the two types of tasks. This is formally shown in Table A.3.

Figure 4: Experimental design and timeline



Note: Shaded boxes denote religiously mixed teams. The share of Muslim workers in mixed teams is (roughly) between 35%-40% (balanced across HD and LD sections, see Table A.3). This diagram shows the timeline of the intervention. The baseline survey was completed between July and August in 2019. Time-use data in order to classify tasks into HD and LD types were collected in October 2019. The experiment was conducted between November 2019 and March 2020. Details of sample size by treatment arms are presented in the figure. A phone survey was conducted at endline in April and May of 2020 due to COVID-19 related restrictions.

Table 3: Randomization check

	Panel A: Outcomes relevant at work				**Panel B: General characteristics and attributes**				
	Tenure	Muslim co-workers *Hindus*	Taking Orders	Communicating	Age	Schooling	Trust	Altruism	Inter-religious contact outside work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mixed × LD	0.1013	0.0037	0.0591	-0.0775	1.5458	-0.2256	0.4444	0.0289	0.0408
	(0.3357)	(0.0096)	(0.0550)	(0.0513)	(1.3956)	(0.5021)	(0.3353)	(0.2187)	(0.0472)
Mixed × HD	-0.0745	-0.0067	0.0004	0.0248	0.4142	0.1904	-0.0990	0.0466	0.0025
	(0.3283)	(0.0083)	(0.0440)	(0.0459)	(0.7993)	(0.3375)	(0.3161)	(0.1495)	(0.0477)
p(Mixed × HD = Mixed X LD)	0.68	0.41	0.37	0.11	0.45	0.46	0.20	0.94	0.54
Mean Dep. Var	4.44	0.16	0.73	0.53	34.21	7.84	3.82	6.66	0.45
Line × Section Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	585	480	586	586	586	586	583	583	586
Adj. R^2	0.124	0.300	0.012	0.051	0.075	0.061	-0.004	-0.010	0.098

* p<0.10, ** p<0.05, *** p<0.010. The unit of observation is an individual worker. Standard errors clustered at the line-section-level team. "Tenure" and "Schooling" are measured in years and as highest grade completed respectively. "Taking Orders" is a dummy variable coded 1 if the respondent reported to be “Always comfortable" taking orders from non-coreligionists and 0 if they reported to be “Sometimes uncomfortable" or “Always uncomfortable". "Communicating" is coded 1, 0.5 and 0 for the responses “Always comfortable”, “Sometimes uncomfortable” and “Always uncomfortable” when asked about being comfortable communicating with non-coreligionists. Survey questions on “Trust" and “Altruism" are used from the World Value Survey (WVS). The dependent variable "Inter-religious contact" refers to the degree of cross-religion interaction that workers had at baseline, outside of work. The variable is coded 1, 0.5 and 0 if a worker mentioned that during the daily course of their life they: 1) interact with more than 5 non-coreligionists 2) interact with 1 to 5 non-coreligionists, or 3) do not interact with anyone outside their religion, respectively.

3.4 Quasi-random allocation of workers to tasks at baseline

Since the majority of workers continued to work in their original tasks (i.e. the area of work was not randomized), one might worry about distinguishing between the effects of task types versus worker types (on team productivity) from religious mixing. This is particularly important if workers are able to self-select into high- or low-dependency sections. The randomization check already provides evidence against such systematic sorting. Nevertheless, I address this concern in more detail in Appendix A.3. I argue that worker characteristics are balanced across HD and LD tasks due to the firm's hiring and worker allocation policy and not simply by chance. The HR manager always has a pool of job applicants who are called upon on a first-come-first-served basis, when vacancies become available. As a result, workers do not have the option to choose their area of work when they join. However, workers may quit at different rates across the two types of tasks, leading to possible selection bias. If that were the case, this would be reflected in the average tenure of workers in HD and LD sections. As shown in Table A.4, this is not the case – tenure is balanced between workers in HD and LD sections. I then show that only a handful of workers (15.2%) have switched their area of work from when they first joined the firm. Finally, I show that these switches are not correlated with observable characteristics of the workers and have happened largely due to organizational requirements at the firm.

4 Econometric specification

Outcomes in this paper are measured at three levels: 1. Production line-level, 2. Production line-section-level, and 3. Individual-level. Line-level real output data are linked to the firm's revenues. Line-section-level ratings were recorded by production supervisors daily during the period of the experiment only. These data help investigate the source of line-level differences in real output. Survey measures at baseline and endline are at the individual worker level. I use these to study worker interactions during production as well as treatment effects on attitudes.

Line-Level specification

I compare line-level output between HD-Mixed and LD-Mixed lines as shown in Figure 3. The specification used is:

$$Y_{klst} = \beta_1 T_k + \alpha_l + \alpha_s + \alpha_t + \epsilon_{klst}, \tag{1}$$

where Y_{klst} is output from line-level team k, in line l, in shift s on day t. T_k denotes the treatment status (1 if HD-Mixed line and 0 if LD-Mixed line). The coefficient β_1 denotes the line-level

treatment effect. α_l, α_s and α_t are line, shift and day fixed effects respectively. I include production line fixed effects to control for product type, shift fixed effects to account for differences in worker productivity at different times of the day (morning, afternoon, night) and day fixed effects to control for factory-wide shocks to demand. Standard errors are clustered at the line-cohort-level (or in other words at the line-level team). Since there are only 15 clusters at the line-level, I also present wild cluster bootstrap standard errors (Cameron et al., 2008) for these regressions.

Line-Section-Level specification

Supervisors assigned a daily rating (out of 5) to each line-section-level team, independent of the performance of other sections in the line. I use this data to evaluate the source of line-level differences in output. The following baseline specification is used:

$$Y_{mklst} = \beta_1 Mixed_{mkl} \times LD_{ml} + \beta_2 Mixed_{mkl} \times HD_{ml} + X_{mkl} + \alpha_{ml} + \alpha_s + \alpha_t + \epsilon_{mklst}, \quad (2)$$

where Y_{mklst} is the performance rating of section m of team k in line l in shift s on day t. $Mixed_{mkl}$ denotes whether the section has a religiously mixed or homogeneous team (which is determined by line-level team type k). LD_{ml} and HD_{ml} are dummies coded 1 if the section is classified as HD and LD respectively (this is defined by line l and section m only). I use the interaction terms $Mixed_{mkl} \times HD_{ml}$ and $Mixed_{mkl} \times LD_{ml}$ to identify effects of having mixed teams in HD and LD sections respectively (given by the coefficients β_1 and β_2). Since *line* × *section* effects α_{ml} are included in these regressions, the dummies HD_{ml} and LD_{ml} are not separately introduced. X_{mkl} is a vector of line-section-level controls. α_s and α_t are shift and day fixed effects respectively.

Individual-level specification

I surveyed workers both at baseline and endline. I use the baseline data for randomization checks as shown in section 3 and also for heterogeneous treatment effects which follow in section 5. As mentioned earlier, the endline data is used to evaluate treatment effects on worker attitudes and interactions between teammates during production. The main specification is:

$$Y_{imkl} = \beta_1 Mixed_{mkl} \times LD_{ml} + \beta_2 Mixed_{mkl} \times HD_{ml} + X_{imkl} + \alpha_{ml} + \epsilon_{imkl}, \quad (3)$$

where Y_{imkl} is the outcome of interest for individual worker i of section m of team k in line l.

X_{imkl} is a vector of individual-level controls. All other variables are described exactly as before. The treatment effects are estimated by coefficients just as in the line-section level specification described above.

5 Results

5.1 Production data

This section begins by showing that HD-Mixed lines produce lower output than LD-Mixed lines, but this effect attenuates over time. I then proceed to the line-section-level analysis and show that line-level differences in output are driven largely by losses from religious mixing in HD sections, while mixing has little effect in LD sections.

5.1.1 Line-Level

Production supervisors record total output from each production line at the end of each shift. Table 4 shows that HD-Mixed lines produced lower output compared to LD-Mixed lines. Observations in this regression are at the line-cohort-day-level. The outcome variable in Column (1) is the log of total output (in pieces) produced by a line-level team in a particular shift of a day. Column (1) shows that HD-Mixed lines on average produced 5% lower output compared to LD-Mixed lines over the period of the intervention. This effect is economically large. Given average output per shift of 450,000 pieces (across all lines) and the typical product priced at Rs 10 ($ 0.13), the results suggest that the firm's revenue would increase by up to Rs.225,000 ($2725) per shift, from having only LD-Mixed lines relative to having only HD-Mixed lines in the factory.

Table 4: Treatment effect on line-level output

	(1) **Log Output (Pieces)**	(2) **Log Output (Boxes)**
HD-Mixed vs LD-Mixed Line	-0.0561***	-0.0440**
	(0.0176)	(0.0202)
Bootstrap (Wild Cluster) C.I.	[-0.102, -0.018]	[-0.10, 0.009]
Day F.E.	Yes	Yes
Shift F.E.	Yes	Yes
Production Line F.E.	Yes	Yes
Mean Dep Var.	10.81	6.97
	(1.02)	(0.970)
N	1018	1018
Adj. R^2	0.725	0.649

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily output produced by line-level teams. Standard errors clustered at the line-level team in parenthesis. Wild cluster bootstrap (Cameron et al., 2008) confidence intervals in square brackets. HD-Mixed Line is a dummy coded 1 for a line-level team with all HD sections religiously mixed and LD sections non-mixed, and 0 for exactly the opposite line-level structure, i.e. LD sections mixed and HD sections non-mixed (LD-Mixed Line).

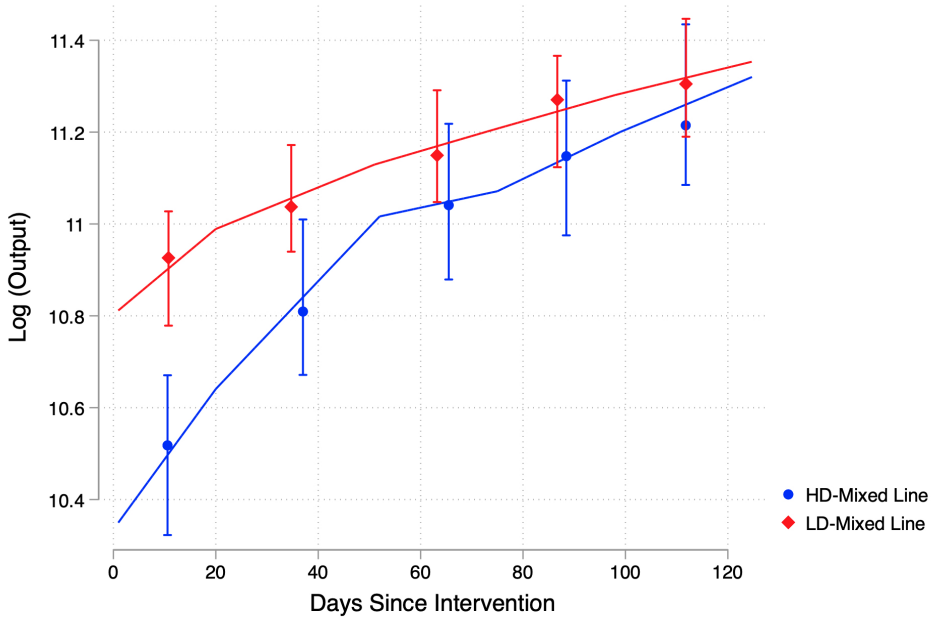
The firm also records total output using the number of boxes with final products that are packed at the end of a shift. These boxes are used to ship products to the market and each box typically includes multiple pieces of a product. The effects are robust to using this variable as the outcome instead (Column 2). Since each production line can manufacture more than one variant of the same product, I show robustness to the inclusion of *line* × *variety* fixed effects in Table B.3. Finally, I also include *line* × *day* fixed effects in Table B.4 – the magnitudes remain similar.

Over the entire period of the intervention, HD-Mixed lines produced lower output than LD-Mixed lines – but how did the treatment effect evolve over time? This would inform us whether repeated interaction with the same set of non-coreligionists can help ameliorate some of the negative effects of mixing on output. In Figure 5, I present a dynamic plot of output (logged) produced over the period of the intervention, by team type.[27] These are from binned regressions using the same specification as in section 4, with the treatment period split into five equal sized bins. The difference in output produced by HD-Mixed and LD-Mixed lines was the largest at the beginning of the intervention and it gradually attenuated over time. Interestingly, output from both HD-Mixed as well as LD-Mixed lines followed an upward trajectory throughout the

[27] This plot is created using the STATA command binsreg, which implements binscatter estimation with robust inference proposed in Cattaneo et al. (2019).

four months of the intervention. This might be because of two reasons: 1. The firm was itself adjusting to new teams and therefore only gradually increased production targets as workers became more comfortable with each other or 2. The experiment was timed to coincide with a period during which the factory faces high demand for its products; so that production remains uninterrupted, absenteeism is low and teams don't disintegrate. This could have also led to the firm setting higher output targets in each subsequent month of the intervention.

Figure 5: Treatment effect on line-level output (dynamic)



Note: This figure is generated from binned regressions using exactly the same controls variables as in Table 4. The treatment period is divided into 5 equal sized bins. The outcome variable is output produced in pieces (logged). Bars denote 95% confidence intervals.

Overall, these results imply that religious diversity is *relatively* more costly in HD tasks than in LD tasks. But the overall line-level differences (between HD-Mixed and LD-Mixed lines) could be driven by religious mixing lowering output in both types of tasks but more in HD, or mixing increasing output in both types of tasks but more in LD. Another possibility is that it affects output negatively (or has no effect) in HD tasks, but positively in LD tasks. Finally, it is also possible that mixing only (negatively) affects output in HD tasks but not in LD tasks. I cannot distinguish between these possibilities using line-level data as there are no homogeneous line-level teams by design. I take this up next in the line-section-level analysis, where such comparisons are possible due to the presence of teams composed of only Hindu workers.

5.1.2 Line-Section-Level

Recall that there are four different types of teams at this level: HD Mixed, HD Non-Mixed, LD Mixed and LD Non-Mixed. The performance of each section was rated (between 0 to 5) daily by production supervisors. These ratings were based on established metrics of time-use efficiency, which assist supervisors in monitoring the performance of teams and individual employees. The benchmarks were different across tasks. For example, *Mixing* sections were rated on the number of batches prepared in a given time period, while most other sections downstream until *Packing* were rated on the number of trays with unfinished products that were sent onto the following section every hour, accounting for the number of trays received from the previous section. *Packing* sections were rated on the number of boxes packed with final goods as well as on packaging material wastage.[28] While the supervisors were trained (by the production manager) to rate each section independent of the performance of the entire line or other sections in the line (to ensure that no section was penalized for the actions of sections upstream), it is still possible that these ratings do not appropriately take into account spillover effects from upstream to downstream sections. In section B.3 (Appendix), I present various checks using sub-samples for which spillovers are less likely to be a concern, and show that the results do not change.

Before turning to the results, I emphasize several validation exercises that establish supervisor ratings as a reliable proxy for assessing line-section-level performance. Firstly, I demonstrate that average ratings (mean over all sections in a line) exhibit a strong correlation with overall line output (Figure C.8). Secondly, the firm systematically records *forming* or *deposit* quantity for each line, which is influenced (largely) by the performance of the Deposit section (for Lines 4 and 5, it is Oven since there is no separate Deposit section). In Table B.5, I present correlations between forming quantity and ratings received by both the Deposit section and Non-Deposit sections in a line. Ratings received by Deposit sections are considerably more predictive of a line's forming quantity. This suggests that the ratings are likely to accurately represent the performance of the respective sections. Lastly, I show that ratings received by both HD and LD (as well as mixed and non-mixed) sections within a line similarly impact overall line-level output (Figure C.9). This is important because it demonstrates that line-level differences in output cannot be explained by LD sections contributing less to overall output, and that supervisors do not discriminate in their ratings between mixed and non-mixed sections.[29]

[28] Table B.23 gives a short explanation of the main tasks across the different production lines, and also explains how team performance was rated in each task.

[29] Informal conversations with supervisors did not suggest that there were systematic differences between HD and LD sections in terms of their contribution to overall output. The fact that the ratings corroborate this observation is reassuring. It also implies that there is unlikely to be discrepancies in rating accuracy between the two

Table 5: Treatment effect on section ratings

	Rating (Raw)		**Rating > Median**	
	(1)	(2)	(3)	(4)
Mixed	-0.0249**		-0.0233***	
	(0.0104)		(0.0083)	
Mixed × LD		-0.0141		-0.0073
		(0.0135)		(0.0128)
Mixed × HD		-0.0363**		-0.0403***
		(0.0162)		(0.0106)
p(Mixed × HD = Mixed × LD)		0.296		0.051
Mean Dep. Var.	3.86	3.86	0.46	0.46
	(0.65)	(0.65)	(0.50)	(0.50)
Day F.E.	Yes	Yes	Yes	Yes
Shift F.E.	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes
N	7627	7627	7627	7627
Adj. R^2	0.601	0.601	0.364	0.365

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Section fixed effects are included in all specifications; as a result the main effect of HD versus LD is not separately identified in columns (2) and (4).

Table 5 presents the core results from the line-section-level analysis. In column (1), I regress supervisor ratings on a dummy variable that denotes whether a line-section-level team is religiously mixed or not (Mixed). The coefficient on Mixed is negative and statistically significant suggesting that mixed teams perform worse overall. Note that line × section effects are included in all specifications in the line-section-level analysis, whereby the identifying variation comes from within the same line-section across different treatment cohorts (teams). These are important to include because of the different benchmarks used to rate each section. In column (2), I introduce the interaction terms (Mixed × HD) and (Mixed × LD) to estimate the effect of having a mixed team in a HD section separately from a LD section. The coefficient on Mixed × LD in column (2) is small and not statistically significant while that on Mixed × HD is negative and statistically significant. This suggests that having mixed teams lead to lower ratings in HD sections but not in LD sections. In columns (3) and (4), the outcome variable is coded 1 if the rating received is above median and 0 if lower.[30] The effects with a binary dependent variable

section types.

[30] A large fraction of ratings is concentrated between 4 and 5 (see Figure C.7), making a binary dependent variable also appropriate for this specification.

are similar to those with raw ratings. Overall, this is evidence that lower output in HD-Mixed lines (relative to LD-Mixed) is caused predominantly by lower output in religiously mixed HD sections, while in LD sections, mixing is less costly.

I perform a few additional checks to verify the robustness of these findings. First, in Table B.6, I run separate regressions for HD and LD sections and find similar results. Second, I include supervisor fixed effects in Table B.7, as there might be some degree of subjectivity in the ratings that is supervisor-specific, but the results look identical. Third, since Muslims have lower schooling and tenure relative to Hindus and these factors may influence team productivity, I also include them as controls[31] in Table B.8 — once again, the results do not change. Finally, I consider various alternative cutoff values (of continuous coordination in Figure 2) to distinguish HD and LD tasks, and as the findings in Table B.9 show, the results are not sensitive to this.

I next examine whether there is convergence in line-section-level performance over time between mixed and non-mixed HD teams. This is likely given that line-level output differences between HD-Mixed and LD-Mixed lines attenuate over time (recall Figure 5). I split the intervention period into five equal sized bins (exactly as in the line-level analysis), and show that this is indeed the case. The results are presented in Table B.10.

The baseline effect is reported in column (1), which shows a large, negative and statistically significant effect of having a mixed HD team. In column (2), I introduce interaction effects with the event bins. Coefficients on earlier bins are larger (negative) and they gradually reduce in magnitude. This suggests that the largest negative effects of religious mixing on HD section output occurred at the beginning of the experiment when the new teams were first formed; and performance ratings of mixed and non-mixed teams gradually converged over time. The baseline effect and interactions with the event bins are presented for LD sections in columns (3) and (4) respectively. The baseline effect is small and not statistically significant, while the interactions are noisy with no clear dynamic pattern. Overall, these results are re-assuring in that they line up closely with the line-level event-study analysis, but using granular production data at the line-section-level.

[31] In Table B.13, I show that conditional on schooling and tenure, Muslims are as likely to be promoted as Hindus. Therefore, to the extent that the results here may be driven by Hindu-Muslim differences in productivity, these controls are likely to be able to account for that.

5.2 Endline phone survey

The endline survey focused on two main sets of outcomes: 1. Those that capture actual interactions between workers during production and 2. Attitudes towards non-coreligionists coworkers. Only a phone survey could be conducted at endline because of restrictions related to COVID-19. As a result, a large set of outcomes that I was interested in, including political preferences that respondents maybe uncomfortable discussing over the phone, could not be recorded.[32] I take up each of the two sets of survey outcomes in turn.

Worker interactions

In Table 6, I focus on the first set of factors. These collectively proxy for the degree of cohesion and coordination in a line-section-level team. There are three main outcomes variables. The first question asked respondents to point out workers who did not contribute sufficient effort at any point during the intervention ("Identified teammate as contributing low effort"). If a worker identifies his teammate to have not contributed to the team as much as other workers did, or to the extent that is expected, then this outcome is coded 1. I then asked workers to identify teammates who have blamed them in the past for not performing up to the mark ("Blamed by teammate"). The outcome variable is coded 1 for teammates who have blamed the respondent at least once during the intervention period. The final question asked workers to pick teammates who they would give up their relief time for, if asked, or have already done so in the past. Relief time refers to breaks that each worker is entitled to at regular intervals during their shift. In HD sections, workers typically need to coordinate on breaks to a greater degree than in LD sections. The outcome variable is coded 1 for teammates that workers are *not* willing to give up their relief time for ("Unwilling to give up relief time"). Note that these questions were asked retrospectively in lieu of more high frequency data, since many workers reported to have had

[32] In addition, one might be worried about social desirability bias in the responses, since the outcomes I study are self-reported (even though both the baseline and endline surveys were conducted one-to-one with the respondents and anonymity and confidentiality were emphasized). To deal with this, I correlate baseline responses to survey questions (that were asked again at endline and used as outcomes in Table 7) with scores from an Implicit Association Test (IAT) that the workers took. The test involved associating Hindu and Muslim names with positions in the firm hierarchy (worker, operator, supervisor, production manager etc). A positive score on this test denotes a bias towards having Hindus in higher positions, while a negative score shows preference towards Muslims. I correlate these scores with workers' reported attitudes towards taking orders from non-coreligionsts as well as their comfort in communicating with non-coreligionists (Figure C.10). Hindu workers with a larger positive score are less likely to say they are comfortable taking orders and communicating with Muslims. Similarly, Muslim workers with a larger negative score are less likely to say they are comfortable taking orders and communicating with Hindus. This suggests that workers' responses are correlated strongly with their actual preferences and improves confidence in the self-reported survey outcomes.

problems with their teammates in the past but also mentioned that they subsided over time.[33]

Observations in Table 6 are at the *worker-teammate-level* for line-section-level teams. In other words, there are $(N-1)$ observations for each worker, where N denotes the total number of workers in the line-section-level team. I include line × section fixed effects and therefore compare similar size teams doing the same task. Columns (1), (3) and (5) show that mixed teams perform worse on all of these measures. Workers in mixed teams are 4.2 percentage points (30%) more likely to identify a teammate as contributing low effort, 4 percentage points (50%) more likely to have been blamed by a teammate and 6.4 percentage points (25.6%) less likely to give up their relief time for a teammate. In columns (2), (4) and (6), I introduce the interaction terms Mixed × HD and Mixed × LD to test for differential effects by task type. Clearly, having mixed teams in HD sections lead to greater frictions. Surprisingly however, I find that workers in mixed LD sections report to have been blamed more by co-workers than those in mixed HD sections. Individual performances are more easily observable to supervisors in LD tasks (and hence it is easier to blame co-workers), which might explain this pattern.[34] Note that both of these effects are statistically significant on their own.

More generally, it can be observed that mixed teams in LD sections also suffer from these frictions to a greater extent than homogeneous teams – the effects on the interactions Mixed × LD are positive and meaningful in magnitude though not precisely estimated. In fact, one cannot statistically reject that the effects in LD sections are different from those in HD, though the effects in HD sections tend to be larger. Crucially however, these do not translate into mixed teams performing any worse than non-mixed teams in LD sections, which is the case in HD sections, as shown in Table 5. The sample is restricted to only Hindu respondents (since the Mixed vs. Non-mixed overall variation comes only from Hindu workers) in Table B.11 and a similar pattern is observed.

[33] For example, workers were asked if they have been blamed by a teammate at least once in the past, or asked to identify workers who they thought were not contributing effort at any point during the intervention.

[34] It is also plausible that by endline these frictions had subsided more in HD sections than in LD ones.

Table 6: Treatment effect on worker interactions

	Identified teammate as contributing low effort		**Blamed by teammate**		**Unwilling to give up relief time**	
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed	0.0420***		0.0400**		0.0640*	
	(0.0137)		(0.0158)		(0.0365)	
Mixed × LD		0.0317		0.0817***		0.0339
		(0.0226)		(0.0228)		(0.0563)
Mixed × HD		0.0445***		0.0301*		0.0719*
		(0.0154)		(0.0175)		(0.0423)
p(Mixed × HD = Mixed × LD)		0.62		0.05		0.57
Mean Dep. Var	0.14	0.14	0.08	0.08	0.25	0.25
Worker Skill F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Religion F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3696	3696	3684	3684	3727	3727
Adj. R^2	0.016	0.016	0.013	0.014	0.072	0.072

* p<0.10, ** p<0.05, *** p<0.010. Observations are at the worker-teammate level for line-section-level teams i.e. there are (N-1) observations per worker, where N denotes the number of workers in the section. Standard errors clustered at the line-section-level team. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in columns (2), (4) and (6). The outcome variables are as follows – workers were asked to choose teammates who they: (1) thought did not contribute sufficient effort at any point during the intervention (2) have been blamed by during the intervention and (3) have (or would be willing to) given up their relief time for.

These results are consistent with the treatment effects on output and inform us of actual interactions between workers that led to those effects. Coordinating closely as a team on a wide set of issues is important in HD tasks and lower team cohesion caused by these frictions can reduce team output. While mixing in LD tasks also leads to some frictions, the production technology is such that team output is less sensitive to these problems, which likely explains the null effects in LD tasks.[35] But despite these frictions, output differences between HD-Mixed and HD Non-Mixed sections attenuate over time. The next set of results study treatment effects on attitudes of workers towards non-coreligionists at endline, and formally tests whether the

[35] Of course, extreme events such as religious violence after the passing of the Citizenship Amendment Act (CAA) did affect output in mixed LD sections (see Table B.20). But overall, the lower sensitivity of team output to these frictions is perhaps also why there is little incentive for workers to try to overcome their differences. This is reflected in the next set of results where I show reductions in negative out-group attitudes from mixing for Hindu workers, but only in HD tasks.

attenuating output effects are accompanied by improved inter-group relations.

Attitudes at endline

For attitudes, treatment effects are restricted to Hindu workers only, since Muslim workers are always in mixed teams.[36] I use three main outcome variables, two of which are questions also asked at baseline (the baseline values are included as controls). Workers were asked if they are comfortable taking orders from non-coreligionists ("Taking Orders"), whether they find communicating with non-coreligionists (in general) as comfortable as co-religionists ("Communicating") and finally if they prefer to be in mixed or all-Hindu groups if teams were to change again in the future ("Co-working"). While the first two questions were unincentivized, for the third question, enumerators mentioned to the workers that their responses would be recorded by the research team (but not shared with the firm) and kept in mind for future team changes.

I first report the main effect of being randomized into a mixed team. Outcomes in Table 7 are at the individual worker level. All outcomes show positive effects from mixing. Relative to those in homogeneous teams, Hindu workers in mixed teams are 11.5 percentage points more likely to report that they are comfortable taking orders from Muslims (Column 1) and 8.7 percentage points more likely to be comfortable communicating with Muslims (Column 2). Finally, Column 3 shows that they are 11.3 percentage points more likely to *not* express preference for being in a Hindu-only team. These effects are economically significant in magnitude and suggest large gains for Hindu workers from repeated contact with Muslim colleagues. In Columns (2), (4) and (6), I introduce the interaction terms Mixed × HD and Mixed × LD. The effects are entirely driven by contact in HD sections.

The coefficients on Mixed × HD are economically large in magnitude and statistically significant at the 1% level. The coefficients on Mixed × LD are small and not statistically significant, suggesting a zero effect in LD sections. The differences between the effects in HD and LD sections are large and statistically significant. These findings on positive attitude changes of Hindu workers towards Muslims (from mixing in HD tasks) are consistent with attenuating output differences between mixed HD and non-mixed HD teams (Table B.10), as well as with the overall convergence in line-level output between HD-Mixed and LD-Mixed lines (Figure 5).

Summarizing the main results

Overall, it is insightful and non-obvious that the largest positive effects of treatment on attitudes occurred in teams that also suffered the largest negative output shocks. This suggests

[36] The questions were of course still administered to Muslim workers.

that working in close quarters even with some frictions (in HD teams) leads to more positive effects on intergroup relations than working in LD teams. These results emphasize the importance of contact that forces people to learn to work together in overcoming existing differences leading to reduced prejudice. Purely from a profit maximizing point of view however, firms may have little incentive to mix workers in HD tasks if it leads to output loss. This suggests that discrimination might persist in equilibrium and emphasizes the need for targeted management practices to mitigate them.

Table 7: Treatment effect on attitudes at endline: Hindus

	Attitudes towards Muslims					
Comfortable:	**Taking Orders**		**Communicating**		**Co-working**	
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed	0.1147**		0.0872**		0.1132***	
	(0.0452)		(0.0405)		(0.0336)	
Mixed × LD		0.0163		-0.0807		0.0191
		(0.0777)		(0.0617)		(0.0627)
Mixed × HD		0.1714***		0.1840***		0.1674***
		(0.0557)		(0.0409)		(0.0406)
p(Mixed × HD = Mixed × LD)		0.11		0.00		0.07
Mean Dep. Var.	0.74	0.74	0.49	0.49	0.61	0.61
Sample Mean	Baseline		Baseline		Endline non-mixed teams	
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Worker skill F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes	Yes	Yes
N	448	448	448	448	448	448
Adj. R^2	0.066	0.072	0.066	0.088	0.063	0.068

* p<0.10, ** p<0.05, *** p<0.010. The unit of observation is an individual worker. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. The main effect of HD versus LD is not separately identified in columns (2), (4) and (6) because Line × Section fixed effects are included. “Taking Orders” is a dummy variable coded 1 if the respondent reported to be “Always comfortable" taking orders from Muslims, and 0 if they reported to be “Sometimes uncomfortable" or “Always uncomfortable". “Communicating” is coded 1, 0.5 and 0 for the responses “Always comfortable”, “Sometimes uncomfortable” and “Always uncomfortable” respectively, when asked about being comfortable communicating with Muslims. For “Co-working" the outcome is coded 1,0.5 and 0 for the responses "Mixed team", "Indifferent" and "Hindu-only team" when asked about respondents' preferred team type for future changes.

5.3 Robustness: Threats to identification

In this section, I discuss potential threats to the identification strategy and describe how they are dealt with. I discuss factors linked to the research design (such as the absence of Muslim-only teams), as well as those that randomization cannot directly account for (such as differences between mixed and homogeneous teams on demographic dimensions other than religion).

5.3.1 Religion and productivity

First, I address concerns regarding potential bias that could stem from religion simply proxying for differences in other dimensions (education, tenure, etc.) between mixed and non-mixed teams induced by randomization. One might be worried that it is not the interaction between religious mixing and the production technology that leads to productivity loss, but that this interaction simply proxies for these other differences between Hindus and Muslims (or HD and LD tasks).[37] For example, it might be the case that differences in schooling between Hindus and Muslims are not important in LD sections, but might be a problem in HD sections, given the nature of contact. In other words, it is differences in education between Hindus and Muslims that matter in some tasks and not others, as opposed to the production technology being the important factor.

To deal with this, I introduce interactions between the dummy variable Mixed and these variables as controls, in addition to the interaction terms Mixed × HD and Mixed × LD in the line-section-level specification. I specifically use three variables: group size, tenure of workers and schooling of workers. HD sections tend to have more workers, and one might be concerned about differences in responses of workers from being mixed in larger groups as opposed to smaller groups. For example, diversity might be costly when groups are larger because there is likely to be a wider set of issues that require coordination on. It is also possible that supervisors discriminate (in their ratings) when the absolute number of Muslim workers in a team is large. The other two are more obvious choices given that they may affect productivity and there *are* differences amongst Hindu and Muslim workers on these dimensions (but conditional on these two factors, Muslims are as likely to be promoted has Hindus, see Table B.13). The results are reported in Table B.12 – I introduce the interacted controls sequentially. Column 4 reports results from the specification with all the controls. Reassuringly, the interaction term Mixed × HD remains negative and significant after the inclusion of these controls. Note that the interactions Mixed × Schooling and Mixed × Tenure are both positive and meaningful in magnitude (though not statistically significant). This suggests that higher tenure and schooling can dampen some of the negative effects of diversity. Note also in columns (3) and (4), the coefficients on the interaction term Mixed × LD are statistically significant suggesting that diversity might be costly in LD tasks as well, if workers have very low tenure or schooling.

By design, Muslims workers are only in mixed teams in this experiment. In other words, the

[37] A more fundamental worry may be that due to lower schooling and tenure, Muslims have lower productivity. If that were the case however, we should find religious mixing in LD tasks to lower output as well – but we do not find that. I control for these factors in the line-section-level analysis as a robustness check (see Table B.8), and find that the results are not affected.

treatment of being in a mixed team is perfectly collinear with the presence of Muslims. This was done for two main reasons. First, Muslims comprise of only 18% of all workers in the factory, whereby forming homogeneous Muslim teams would lead to significant loss of statistical power in estimating the effects of religious mixing. Second, at baseline, there were no homogeneous Muslim teams to begin with; therefore experimentally generating such teams could raise ethical concerns. The concern this raises is that Muslim workers may have lower productivity, and this could be driving my findings. However, the null effect of mixing in LD tasks suggests that this is unlikely. Second, there is significant heterogeneity in how mixed teams perform at HD tasks. When Hindus have been in mixed teams with Muslims in the past, I find the negative effects of diversity to be muted significantly (see Figure C.1). If Muslims generally had lower productivity, and especially so at HD tasks, it is unlikely that the negative effects of mixing in these tasks would attenuate so significantly when analyzing heterogeneity *by characteristics of Hindu workers* in mixed teams. These results are discussed in more detail in the following section. Finally, I find that Hindus and Muslims were equally like to be promoted as semi-skilled personnel or as operators at the factory at baseline (conditional on schooling and tenure, see Table B.13). Since the skill-designation of workers affect salary, this suggests that the firm does not perceive Hindus and Muslims to be differentially productive either.[38]

5.3.2 New versus old teammates

One might be concerned that the finding that religious diversity negatively affects productivity is driven in part by the difficulty of working alongside new co-workers, as opposed to the frictions that arise when working alongside non-coreligionists. This would be problematic if the share of new co-workers was not balanced between HD- and LD-Mixed teams, as well as between mixed teams (HD or LD) and Hindu-only teams.

I formally reject this possibility in Table B.14. These are individual worker-level regressions where the outcome variable is the proportion of workers in one's current team (randomized team) that were also in their line-section-level team pre-randomization. The mean of the outcome variable is 0.34, which is expected since workers in each production line-section[39] were randomized between three different cohorts – whereby roughly a third of the workers would be known to each other after new teams were formed. Importantly, as shown in Columns (1) and

[38] Further, the results in Table 7 (with only Hindus), do not suffer from this collinearity problem. They suggest that an environment that forces people to learn to work together is important to alleviate group-level differences. The positive effects on attitudes of Hindu workers would be unlikely if Muslims did not perform well at HD tasks.

[39] Of course few workers did not remain in their original tasks (line-section) as explained in section 3.1, but the overwhelming majority did.

(2), the proportion of new workers is balanced across mixed and non-mixed line-section-level teams. Further, the interactions Mixed × HD and Mixed × LD are small in magnitude and not statistically significant. This suggests that the findings in this paper do not simply result from the inability of workers to coordinate with new colleagues, since workers on average had the same proportion of new teammates irrespective of treatment status.

5.3.3 Treatment status and section changes due to randomization

The randomization process involved moving 7.9% of the workers from their original sections (tasks) at baseline so that the line-level team structures in Figure 3 could be achieved. While this is a small share of workers, it is nevertheless important to show that treatment status is not correlated with the probability of section-switching. If that were the case one could argue that the treatment effects are potentially contaminated. For example, if mixed HD teams have a greater share of workers who changed their sections, it is possible that it is in fact the time required to adjust to new tasks that explains the results. To rule this out, in Table B.15, I regress a dummy denoting whether the section (task) of a worker was changed due to randomization, on the treatment dummies. In columns (1) and (2), I include only a dummy for whether the team is religiously mixed or not (Mixed) and then in columns (3) and (4) I include its interactions with section type (HD or LD). I include line × baseline section effects in columns (1) and (3) and line × section effects in columns (2) and (4). The coefficients across the different specifications are small and not statistically significant. Only in column (4), the coefficient on Mixed × HD is negative and marginally significant, suggesting that the probability a worker switched their baseline section is actually marginally *lower* for those in mixed HD sections. This exercise therefore rules out the possibility that the treatment effects are driven by differential rates of section-switching across treatment arms during the randomization process.

6 Channels

The main findings of this experiment are that religious mixing lowers team output only in HD tasks, but output differences between mixed and non-mixed HD teams attenuate over time, and finally, attitudes of Hindus towards Muslims improve from mixing in HD tasks but not in LD. I consider three possible channels for these effects: differences in productivity between Hindus and Muslims, communication/coordination issues, and minority-stereotyping and discrimination. Unfortunately, the evidence does not allow a conclusive judgement on the importance of these channels, but I do provide suggestive evidence in favor of the last channel and against the

first two.[40]

6.1 Assortative (mis)matching in complementary tasks

If Muslims have lower productivity, positive assortative matching (only all-Hindu and all-Muslim teams) would be the output maximizing allocation of workers in HD tasks. While this can explain the static results of mixing, there must additionally be on the job learning or skill transfer from Hindus to Muslims in this framework to explain the dynamic results. However, evidence presented in section 5.3 suggests that Hindus and Muslims are unlikely to be differentially productive. Additional tests in section 6.3 also provide evidence against this as the main channel.

6.2 Communication

Religious mixing could also lead to lower output in HD tasks due to pure communication problems amongst Hindus and Muslims. And over time, improved communication can bring about production gains. An important strength of my setting is that there are no linguistic differences amongst religious groups[41] — majority of the workers in my sample are born in the same district and speak the same language. It is therefore unlikely that the inability to communicate effectively with non-coreligionists is the primary channel either.

While this is not direct causal evidence, I evaluate how the line-section-level results differ for groups that had high outgroup contact (outside of the workplace) at baseline versus those that did not (Table B.16).[42] The results indicate that baseline outgroup exposure does mitigate the adverse impacts of religious diversity to a certain extent. However, the effects remain substantial enough to suggest that this channel is unlikely to be of primary importance and exposure specifically at the workplace may be important.

6.3 Favored mechanism: Minority-stereotyping and discrimination

Having established that Hindu-Muslim differences in productivity and communication breakdown are unlikely to be primary channels, I focus on stereotyping and discrimination as the

[40] One reason for this is the lack of individual-level productivity data. The firm does not collect such data and they were not collected specifically during the experiment due to IRB restrictions.

[41] Of course, communication problems can stem from issues beyond language. However, the fact the outgroup exposure outside the firm does little to mitigate negative productivity effects suggests that this may not be the primary channel.

[42] High baseline exposure may enhance communication and coordination capabilities among diverse groups, thereby improving effectiveness in joint production.

potential main mechanism. I present a conceptual framework (the full model is presented in Appendix D) of minority-stereotyping (by the majority group) to rationalize the core results and present tests for it's empirical implications.

Outline of the conceptual framework

A key distinction is made between Hindu and Muslim workers in this framework based on the asymmetry in their exposure to non-coreligionists at baseline. A large section of Hindu workers in the factory have had very little exposure to Muslims, while all of the Muslim workers have worked with Hindus (recall Figure 1). Based on this, together with general evidence on discrimination against Muslims in access to education and labor markets in India (Kalpagam et al., 2010; Basant, 2007),[43] I assume that Hindus (mistakenly) believe Muslims may have lower productivity. Muslim workers do not make this distinction between in-group and out-group workers. This asymmetry in baseline priors matters in HD interactions due to complementarities in the production function.[44]

Workers interact in teams for a given length of time and have choice over high or low effort, with the former being more costly. Hindus and Muslims are identical in all aspects, other than the fact the Hindus (depending on past exposure) start off with the belief that Muslims may exert low effort.[45] This causes them to exert low effort in HD interactions. Hindu workers update their beliefs based on their own effort and realized team output (teammate effort is not directly observable). Muslims workers, given their past experience of working with Hindus, are forward-looking and can take actions to "shift" their Hindu teammate's prior (by exerting high effort).

While it is optimal for Muslims to also exert low effort if Hindus exert low effort, given a fixed period of interaction, Muslim workers exert high effort if and only if Hindu workers' priors are not below a certain level. The intuition behind this is that if Muslims exert high effort there would be greater realizations of high output events than Hindus expect (given their belief and action), and as a result they would gradually update their priors, leading to both groups eventually exerting high effort. Muslims only find this worthwhile if initial beliefs of Hindus are

[43] Muslims in my sample have significantly lower schooling than Hindus (Table B.1).

[44] In HD tasks, the joint effort of all workers determines the likelihood of high (and low) output. In LD, total output is modelled as the sum of individual expected output (output is still a stochastic function of individual effort) and therefore the priors of Hindu workers are inconsequential in determining effort.

[45] Hindus think Muslims may be behaviorally disposed to exerting low effort (stereotyping) or face infinite cost of high effort — in terms of the framework these two are equivalent. In general, taste-based stereotyping/discrimination could also lead to similar effects. In fact, I do find that the negative effects of religious mixing are larger if Hindus have political preferences that may capture anti-Muslim sentiments (see Table B.22).

sufficiently high, such that their initial investment cost is compensated for by sufficient periods of high-output payoff. While workers are not paid a daily piece-rate, team performance does influence promotion and pay. Therefore (from a forward-looking Muslim worker's perspective) there are incentives to improve team performance. (Note that Hindu workers simply operate based on their beliefs).

Empirical tests

An important implication of this framework is that Hindu workers with a high initial belief that Muslims also exert high effort are less likely to discriminate against them. Therefore, Hindus who in the past have had Muslim co-workers (for a sufficiently long period of time), should continue to optimally exert high effort (based on their higher priors), when randomized into a mixed HD team. I exploit the quasi-random nature of allocation of workers to teams at baseline, and the resulting variation in past contact that workers had in different teams to test this.[46]

I divide the dataset into two distinct groups for the analysis: (1) teams where the share of Muslim co-workers that Hindus had at baseline was above-median, and (2) teams where the share of Muslim co-workers for Hindus was below-median. Figure C.1 presents dynamic effects (relative to homogeneous teams) for these two groups separately. Notably, the adverse effects of diversity appear more pronounced in the second group, where Hindu have had fewer Muslim teammates in the past (but they attenuate over time). In contrast, the effects for the first group are small and not statistically significant.[47]

Let's now turn our attention to Muslims' past exposure to Hindus. Muslims being the minority group invariably have high exposure to Hindus. Thus, the comparison is between teams where Muslims have had exceptionally high contact with Hindus (median Hindu-exposure for Muslims is 80%) and teams where past exposure *is closer to the actual experiment* (below 80%). Interestingly, if Muslims have been part of teams with a similar Hindu-Muslim composition as in the experiment (relative to those with very high exposure), production losses are less pronounced (see Figure C.3). (However, these differences are not as marked as those observed when analyzing exposure from the Hindu perspective.) This is likely because diversity only matters when the share of Muslims in a team is sufficiently high, such that the priors that Hindus have about Muslims bear significant importance for their own effort. This necessitates a response from the Muslim workers, and those with past experience in such dynamics tend to be more responsive during the experiment. The fact that teams where Muslims have had very

[46] The firm lacks comprehensive historical data on the teams workers have been a part of. Instead, they maintain records of the current team allocation, which are infrequently updated and only when significant changes occur.

[47] The effects for LD sections are presented in Figure C.2. There effects are small and not statistically significant.

high exposure to Hindus under-perform, further implies that the observed effects are unlikely to be due to initial lower productivity of Muslims (and is in fact driven by Hindus) and subsequent learning from Hindus. If this were the case, one would anticipate Muslims with higher exposure to Hindus in the past to have actually benefitted more from such skill transfer.

Next, I use endline survey data to test the implication that the minority group "invests" in the majority group to ameliorate negative stereotypes about them. First, the data indicates a higher overall likelihood for Muslims (in comparison to Hindus) to sacrifice their relief time (an important marker of individual effort) for teammates (see Table B.17). This is direct evidence at the individual-level in support of the notion that Muslims exerted higher overall effort during the intervention. Second, in Table B.18, I use dummy variables for the religion of the respondent, that of the person being referred to in the survey question, and their interaction as the main regressors to decompose the effects on worker interactions. Columns (1), (3) and (5) show that depsite exerting higher effort, Muslim workers (at least initially) are more likely to be identified as not contributing effort, be blamed, and have fewer co-workers willing to give up relief time for them (even though they themselves are less likely to criticize their co-workers). Importantly, the coefficients on the interaction terms introduced in Columns (2), (4) and (6) show that the criticism of Muslim workers come from their Hindu counterparts (consistent with Hindus having lower priors about the ability of Muslims), while Muslim workers are *more* likely to give up relief time for Hindu co-workers (consistent with Muslim workers initiating integration and overall higher effort exerted by them). This decomposition lends support to the idea that it is indeed stereotyping of minorities by the majority group which results in lower team cohesion and output initially; while the minority group initiates the integration process.[48] Finally, it is worth highlighting that Hindu workers' improved preferences towards co-working *generally* with Muslims in the future (see Table 7) is also consistent updated priors of Hindus, and inconsistent with the Muslims initially having lower productivity and learning from Hindus.

Taken together, these results support minority-stereotyping and discrimination are the primary mechanisms behind the core findings. They also suggest that factors such as communication, coordination, or Hindu-Muslim differences in productivity may only play a secondary role in this context.

[48] In Table B.19, I further decompose the findings of Table B.18 into HD and LD sections and show that the effects discussed above are driven by HD sections and less so by LD sections.

7 Policy discussion: Firm supervisor survey

Do firm supervisors understand the costs of diversity and how they depend on the production function? Can they predict the findings from this experiment, and if so, do they suggest integration of workers only in LD tasks or do they recommend other management practices to ameliorate possible negative effects of religious mixing in HD tasks? To analyze these policy relevant questions, I surveyed supervisors and operators (personnel with some leadership role) of five different processed food manufacturing plants in April, 2021.

Participants were first asked to denote which of the two tasks (HD or LD in Figure C.4): (1) requires greater coordination and communication amongst co-workers and (2) is likely to cause more frictions and arguments amongst workers. They picked the HD task more frequently for both of these questions (Figure C.11). Interestingly, while close to 80% of the supervisors chose the HD task for (2), a fair share of them also picked either the LD task (17%) or mentioned both HD and LD (35%) for (1). This reiterates an important point about LD tasks and the mechanism behind the core results in this paper: workers are not typically inclined to sabotage or undermine the efforts of their out-group members (which is possible in LD tasks as well). Rather, a negative perception of out-group members causes frictions which are costly when working in production environments that require workers to be significantly dependent on each other.

Participants were then asked to predict whether a religiously homogeneous or mixed team would be more productive at each task. They were informed that I have conducted an experiment to test this and that they would be rewarded with Rs 25 bonus (about 30% of their hourly wage) if their answer matches with my findings — this was done to reduce social desirability bias. Between 40%-45% of the supervisors mentioned that religiously mixed teams would be more productive in both tasks (Figure C.12). This could still be because of social desirability bias or as I show next, supervisors perhaps consider issues beyond direct productivity arising from segregation of workers (by religion), which prompts them to answer in this manner. Nevertheless, a significantly higher share of respondents mentioned that a homogeneous team would be more productive at the HD task (30%) than the LD task (8%). Overall, about a third of the supervisors predicted correctly that homogeneous teams would be more productive in HD tasks and about half of them correctly mentioned that mixing would be inconsequential in LD tasks.

While it is possible that a large share of supervisors actually do not understand the costs of diversity and consequently do not segregate workers, it is also possible that there are additional costs that do not justify segregation. To understand this systematically, respondents were finally asked if they are willing to segregate workers by religion and/or age if workers do not perform

well as a team because of these differences. I use age as a natural benchmark because in the Indian context age differences could be an important source of conflict amongst teammates. The supervisors generally seem to be averse to segregation on either dimension, but they are especially opposed to segregation by religion (Figure C.13), despite the potential for losses. About a quarter of the supervisors correctly (as I find) cite negative effects of diversity dissipating over time as reasoning behind this choice. However, the first order concern is about segregation actually raising tensions further. Informal conversations with supervisors suggest that some of the concerns they have in mind are with respect to such segregation creating a hostile environment in common areas of interaction (canteen, tea room), in addition to tensions on the production floor.

In sum, this survey shows that roughly between a third to a half of the supervisors correctly predicted the results of the intervention. However, it is clear that despite the possibility of losses, the majority of supervisors are averse to segregation of workers by religion. Many of them, being aware of the long-term gains from repeated contact are willing to trade off potential short-run costs of non-segregation to productivity. But they are also concerned about costs to segregation that are typically hard to identify as a researcher by simply analyzing production data.

8 Conclusion

My findings suggest that both the nature and duration of contact are important in understanding how religious diversity in firms may impact productivity. Teamwork that requires high coordination creates incentives for workers to invest in building social capital with out-group members. This brings positive attitudinal change and productivity gains over time, but it might be unprofitable in the short-run through lost output. Overall, my results suggest a potential tension between the goal of maximizing short-run productivity and that of improving intergroup relations. This suggests that in equilibrium we could observe a lot of integration at work without intergroup relations improving – the integration might only occur in contexts where contact is socially ineffective.

Beyond conceptual contributions, this paper has a few important implications for policy. First, firms with high-dependency production should minimize team switching in order to mediate possible negative effects of diversity. Second, in firms with low-dependency production, exposure to non-coreligionists might not necessarily reduce negative out-group attitudes. While this may not lower output, a less cohesive work culture can lead to problems outside of daily production. Thus, additional measures to ensure a collaborative environment for workers

to interact in might be worthwhile in LD firms. This can even be achieved outside the workplace, for example though sports teams (Lowe, 2021; Mousa, 2018). An open question remains whether that can also lead to productivity gains at the workplace. If that were the case, the cost to output from mixing workers in HD tasks to integrate them could be avoided.[49] However, if belief updating with respect to specifics about co-workers' effort levels at work is the driving factor (as suggested in the theoretical framework), contact outside the firm might not be able to entirely mitigate the negative effects of diversity.

With structural transformation, the nature of production changes, potentially influencing the type of inter-ethnic interactions. In agricultural societies, land cultivators largely work in LD environments with limited contact with new people, but manufacturing activity involves a higher share of HD work (construction work, small firms etc.), as well as regular contact with new people, making diversity costly. In services, with a comparatively higher share of LD work and a regular set of colleagues, these costs might be low again. This suggests that policies that promote trust in ethnically diverse societies can help ease the transition from agriculture to formal manufacturing where diversity may be costlier (A.Churchill and Danquah, 2020).

Finally, the finding that minorities (Muslims) bear the cost of integration in this experiment is generalizable to other settings. For example, the finding that African-Americans are rewarded less for their effort (relative to the average American), requiring them to work harder to achieve similar career goals (DeSante, 2013) or that Asian immigrants in the U.S., being aware of their unequal racial status, adopt a normative path to success and assimilation (by achieving model-minority status) (Zhou, 2004 and Zhou and Xiong, 2005), relate closely to my results in the Indian context. Minority groups, despite being discriminated against, may thus play a crucial role in the process of nation-building through initiating economic and social integration in diverse societies.

References

A.Churchill, S. and M. Danquah (2020). Ethnic diversity and informal work in Ghana. Technical report, WIDER 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*, 102445.

[49] Of course arranging such organized collaborative contact might itself be costly for a firm in terms of time and resources.

Afridi, F., A. Dhillon, and S. Sharma (2020). The ties that bind us: Social networks and productivity in the factory. *CEPR Discussion Paper No. DP14687*.

Alam, M. S. (2010). Social exclusion of Muslims in India and deficient debates about affirmative action: Suggestions for a new approach. *South Asia Research 30*(1), 43–65.

Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature 43*(3), 762–800.

Alesina, A. and E. Spolaore (1997). On the number and size of nations. *The Quarterly Journal of Economics 112*(4), 1027–1056.

Allport, G. W., K. Clark, and T. Pettigrew (1954). The nature of prejudice. *Addison-wesley Reading, MA*.

Altonji, J. G. and C. R. Pierret (1998). Employer learning and the signalling value of education. *Internal Labour markets, Incentives and Employment*, 159–195.

Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics 116*(1), 313–350.

Artiles, M. (2020). Within-group heterogeneity in a multi-ethnic society. *Unpublished Working Paper*.

Asher, S., P. Novosad, and C. Rafkin (2018). Intergenerational mobility in India: Estimates from new methods and administrative data. *World Bank Working Paper*.

Ashraf, N. and O. Bandiera (2018). Social incentives in organizations. *Annual Review of Economics 10*, 439–463.

Baggio, M. and M. M. Cosgel (2023). Racial diversity and team performance: Evidence from the american offshore whaling industry. *Available at SSRN 4398892*.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *The Review of Economic Studies 77*(2), 417–458.

Bandiera, O., I. Barankay, and I. Rasul (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association 11*(5), 1079–1114.

Basant, R. (2007). Social, economic and educational conditions of Indian Muslims. *Economic and Political Weekly*, 828–832.

Bazzi, S., A. Gaduh, A. D. Rothenberg, M. Wong, et al. (2017). Unity in diversity?: Ethnicity, migration, and nation building in Indonesia. *Centre for Economic Policy Research*.

Becker, G. (1957). The economics of discrimination. *Chicago: University of Chicago Press*.

Bhalotra, S. R., I. Clots-Figueras, L. Iyer, and J. Vecci (2018). Leader identity and coordination. *IZA Discussion Paper*.

Bhattacharjee, S. and A. R. Chaudhuri (2022). Religion in the labor market: evidence from india. *Applied Economics 0*(0), 1–36.

Bhaumik, S. K. and M. Chakrabarty (2009). Is education the panacea for economic deprivation of Muslims?: Evidence from wage earners in India, 1987–2005. *Journal of Asian Economics 20*(2), 137–149.

Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles (2006). Empathy or antipathy? The impact of diversity. *American Economic Review 96*(5), 1890–1905.

Bursztyn, L., T. Chaney, T. A. Hassan, and A. Rao (2021). The immigrant next door: Exposure, prejudice, and altruism. *NBER Working paper* (w28448).

Bursztyn, L., B. Ferman, S. Fiorin, M. Kanz, and G. Rao (2017, 12). Status Goods: Experimental Evidence from Platinum Credit Cards*. *The Quarterly Journal of Economics 133*(3), 1561–1595.

Bursztyn, L. and R. Jensen (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics 9*, 131–153.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Carpenter, J. and E. Seki (2011). Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay. *Economic Inquiry 49*(2), 612–630.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2019). On binscatter. *arXiv preprint arXiv:1902.09608*.

Census, . (2011). Census of India 2011 provisional population totals. *New Delhi: Office of the Registrar General and Census Commissioner*.

Chapparban, S. N. (2020). Religious identity and politics of citizenship in South Asia: A reflection on refugees and migrants in India. *Development 63*(1), 52–59.

Chen, R. and Y. Chen (2011). The potential of social identity for equilibrium selection. *American Economic Review 101*(6), 2562–89.

Churchill, S. A., M. R. Valenzuela, and W. Sablah (2017). Ethnic diversity and firm performance: Evidence from China's materials and industrial sectors. *Empirical Economics 53*(4), 1711–1731.

Corno, L., E. La Ferrara, and J. Burns (2022, December). Interaction, stereotypes, and performance: Evidence from South Africa. *American Economic Review 112*(12), 3848–75.

DeSante, C. D. (2013). Working twice as hard to get half as far: Race, work ethic, and America's deserving poor. *American Journal of Political Science 57*(2), 342–356.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics 112*(4), 1203–1250.

Eckel, C. C. and P. J. Grossman (2005). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization 58*(3), 371–392.

Farber, H. S. and R. Gibbons (1996, 11). Learning and Wage Dynamics. *The Quarterly Journal of Economics 111*(4), 1007–1047.

Fershtman, C. and U. Gneezy (2001). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics 116*(1), 351–377.

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*(2), 212–216.

Hamilton, B. H., J. A. Nickerson, and H. Owan (2012). Diversity and productivity in production teams. In *Advances in the Economic Analysis of participatory and Labor-managed Firms*. Emerald Group Publishing Limited.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics 129*(4), 1899–1946.

Jakiela, P., E. Miguel, and V. te Velde (2011). Combining field and lab experiments to estimate the impact of human capital on social preferences. *Mimeo, UC Berkeley*.

Jha, S. (2013). Trade, institutions, and ethnic tolerance: Evidence from South Asia. *American Political Science Review*, 806–832.

Kalpagam, U. et al. (2010). Are Muslims discriminated against in the labour market in India? *Indian Journal of Labour Economics 53*(1).

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.

Khan, J. I. (2019). Muslims in Indian labour market: Access and opportunities. *Sage Publications Pvt. Limited*.

Lange, F. (2007). The speed of employer learning. *Journal of Labor Economics 25*(1), 1–35.

Lazear, E. P. (1998). Personnel economics for managers. *Wiley New York*.

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review* (6), 1807–44.

Macchiavello, R., A. Menzel, A. Rabbani, and C. Woodruff (2020). Challenges of change: An experiment promoting women to managerial roles in the bangladeshi garment sector. Technical report, National Bureau of Economic Research.

Marx, B., V. Pons, and T. Suri (2021). Diversity and team performance in a Kenyan organization. *Journal of Public Economics 197*, 104332.

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

Miguel, E. (2004). Tribe or nation? Nation building and public goods in Kenya versus Tanzania. *World Politics 56*(3), 327–362.

Mitra, A. and D. Ray (2014). Implications of an economic theory of Conflict: Hindu-Muslim violence in India. *Journal of Political Economy 122*(4), 719–765.

Montalvo, J. G. and M. Reynal-Querol (2017). Ethnic diversity and growth: Revisiting the evidence. *Review of Economics and Statistics*, 1–43.

Mousa, S. (2018). Overcoming the trust deficit: Inter-group contact and associational life in post-ISIS Iraq. *Science. Vol. 369, Issue 6505, pp. 866-870.*.

Nath, S. and S. R. Chowdhury (2019). Mapping polarisation: Four ethnographic cases from West Bengal. *Journal of Indian Anthropological Society 54*, 51–64.

Paluck, E. L., S. A. Green, and D. P. Green (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy 3*(2), 129–158.

Parrotta, P., D. Pozzoli, and M. Pytlikova (2012). Does labor diversity affect firm productivity? *IZA Discussion Paper*.

Pettigrew, T. F., L. R. Tropp, U. Wagner, and O. Christ (2011). Recent advances in intergroup contact theory. *International Journal of Intercultural Relations 35*(3), 271–280.

Pillalamarri, A. (2019). The origins of Hindu-Muslim conflict in South Asia. *Washington, DC: The Diplomat*.

Rao, G. (March, 2019). Familiarity does not breed contempt: Generosity, discrimination and diversity in Delhi schools. *American Economic Review Vol. 109, No. 3*.

Talbot, I. and G. Singh (2009). *The partition of India*. Cambridge University Press.

Zhou, M. (2004). Are Asian-Americans becoming "White?". *Contexts 3*(1), 29–37.

Zhou, M. and Y. S. Xiong (2005). The multifaceted American experiences of the children of Asian immigrants: Lessons for segmented assimilation. *Ethnic and Racial Studies 28*(6), 1119–1152.

A Experiment details (Online Appendix)

This section describes details of the experiment. I first discuss the randomization process and then move on to other experimental details related to subject selection and implementation.

A.1 Randomization steps and implementation timeline

In terms of allocating workers to teams, I first obtained the baseline (existing) allocation of workers from the management, and then undertook the following steps to allocate workers to their new teams. Each step involved in the randomization process is described in detail below.

> **Step 0: Determine religious composition of each section in each line**
> *For each section of each line, first decide the final number of Hindus and Muslims (typically 35%-40% Muslims in mixed sections) to achieve the desired line-level team structures (HD-Mixed or LD-Mixed lines)*
> *s.t.* $\sum H_s = \bar{H}$ *and* $\sum M_s = \bar{M}$, where $\bar{H}$ and $\bar{M}$ denote the total number of Hindus and Muslims in the line across all three cohorts.

Workers were not moved across production lines for randomization. Therefore, the religious composition of line-section-level teams was constrained by the overall number of Hindus and Muslims in the whole line (across the three cohorts) at baseline. Since the proportion of Muslim workers in each line was very close to the overall share of Muslims in the factory, mixed sections (both HD and LD) ended up with roughly 35%-40% Muslim workers after randomization.

> **Step 1: Section Shifting**
> *Suppose 2 additional Muslim workers are required in a section to achieve the desired religious composition (35%-40% Muslims in mixed teams). Then the following steps are taken:*
> *a) Randomly order workers within section × religion × skill*
> *b) Find a section with enough Muslims*
> *c) Randomly pick 2 Muslim workers to shift in*
> *d) Randomly pick 2 Hindu workers to shift out*

This step is perhaps the most crucial in order to achieve the desired line-level treatment types described in Figure 3. At baseline, not all sections of all lines (across all three cohorts) had enough Muslim workers to achieve 35%-40% Muslim share in mixed line-section-level teams after randomization. Therefore, workers were moved across sections in this manner to achieve that. This also meant that only the *minimum* number of workers required were moved, satisfying the firm's requirement of minimizing section-switching.

Step 2: Re-randomize

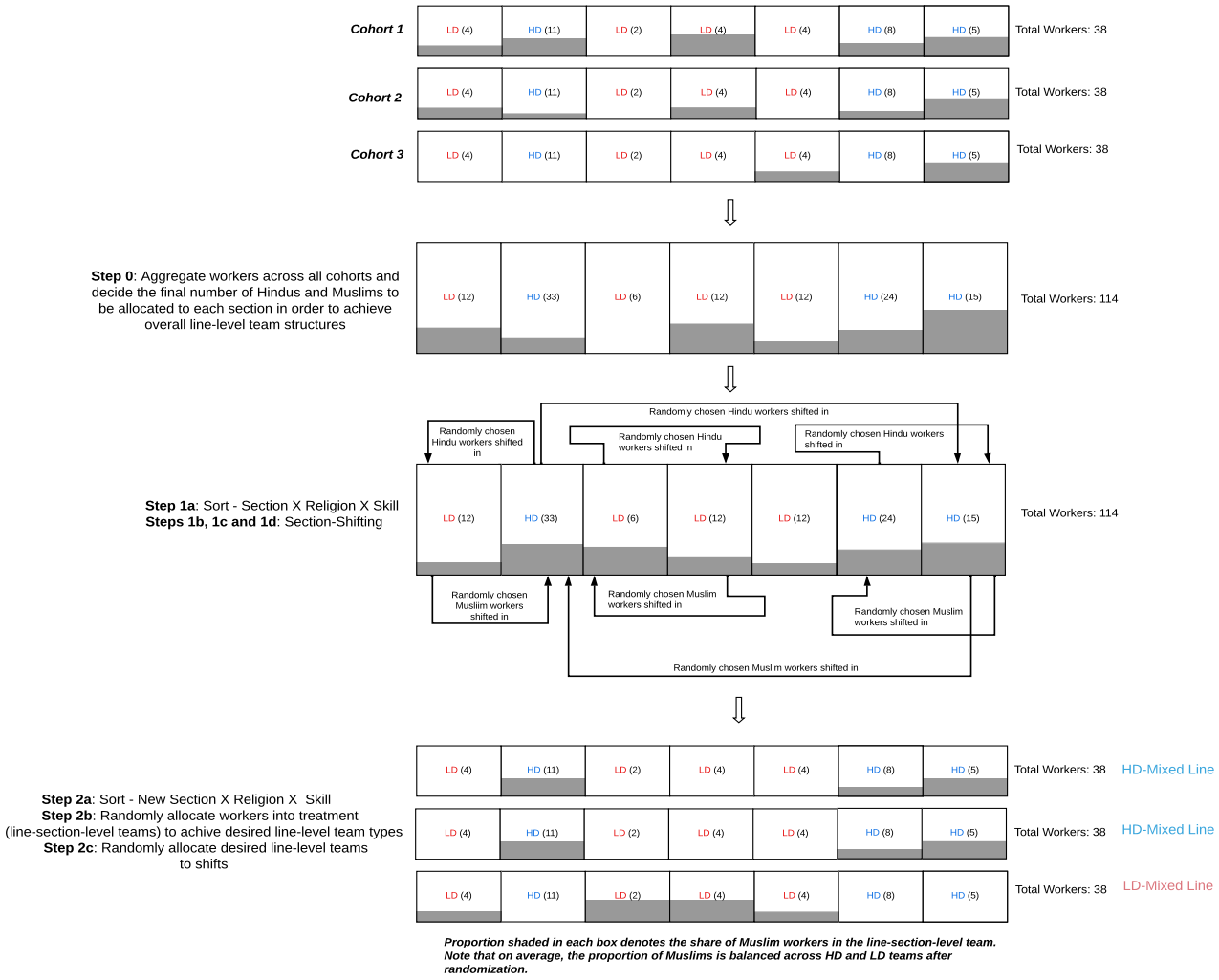
a) Randomly order within new section × religion × skill level

b) Allocate workers into mixed vs homogeneous line-section-level teams as pre-specified to achieve line-level treatment types

c) Randomly allocate line-level teams (lines) to shifts/supervisors

In Step 2, workers were sorted by their new section (only workers who were moved in Step 1 had a different section than at baseline), religion and skill and randomly allocated to line-section-level teams (recall that there are three teams per section in a line – one for each shift) to form the desired line-level team structures, as in Figure 3 (i.e HD-Mixed lines or LD-Mixed lines). Finally, the line-level teams were randomly allocated to one of the three shifts and the usual weekly shift rotations were introduced. Figure A.1 provides a visual representation of these steps.

Figure A.1: Randomized Steps (From baseline structure to randomized teams)



Note: This figure illustrates the steps involved in the randomization process – how, from the given religious composition of sections at baseline the desired line-level team types are achieved. The figure is based on the description of the steps discussed in section A.1. The structure of Line 1 in Figure 1 is used. Team religious compositions shown in the figure are for illustration only. The shades denote the share of Muslim workers in a line-section-level team.

A.2 Other experiment design details

Factory and subject (worker selection): The factory was contacted through a member of the senior management team and multiple different experimental designs were discussed with

the production supervisors before the one actually implemented was finally approved. All the workers at the factory were involved in the experiment.

Implementation of post-randomization teams: During both the baseline and endline surveys enumerators mentioned to the workers that the “project that aims to understand how interaction between workers from different backgrounds in a factory environment affects team performance.” The new teams (post-randomization) were printed and posted on the production floor. Workers were able to infer the religion of their co-workers directly from the names.

Worker compensation on days with no production: Workers in the factory are divided into two groups: permanent and contractual. The majority of workers are contractual workers. On days when there is no production on a particular line, supervisors select a set of contractual workers who are asked to come to work. In principle, these “off-days" are supposed to be equally shared amongst all contractual workers. This ensures all workers are able to get wages on some off-days which helps the firm retain its workforce. In practice, this often happened in a haphazard manner with some workers getting more than their fair share while others getting less. This was changed during the intervention. Workers in each line and cohort were divided into two large groups and the groups were alternately given work on off-days. This helped reduce attrition from the study.

A.3 Identification checks: Quasi-random allocation of workers to tasks at baseline

Hiring at the factory occurs on a rolling basis as and when vacancies become available for each position on a production line. The HR manager always has a pool of job applicants at hand who are called upon on a first-come-first-served basis. As a result, workers do not have the option to choose their area of work when they join. It is possible that workers quit at a different rate across the two types of tasks (HD and LD), leading to possible selection bias. However, if that were the case, this would be reflected in the average tenure of workers in HD and LD sections. As shown in Table A.4 subsequently (below), this is not the case - tenure is balanced between workers in HD and LD sections.

Table A.1: Dependency switches

First Job/Final Job	Low-Dependency	High-Dependency	Total
Low-Dependency	146	29	175
High-Dependency	60	350	410
Total	206	379	585

Note: This matrix reports the number of workers, who, from when they first joined the factory until before the intervention, switched jobs that also involved switching Dependency. 29 workers (4.9%) switched from low- to high-dependency, while 59 workers (10.2%) switched from high- to low-dependency. While 15.2% of the workers switched jobs at least once, 7.2% of them held one or more job between their first and final job at the factory.

While selection into jobs is therefore unlikely at hiring, it is possible that over time, workers are able to sort into their sections of choice. In order to assess if that is the case, workers were asked to report their first job at the factory and their final job immediately before the intervention began. They were also asked to report any other job that they held for a period of more than six months at the factory. Table A.1 reports a matrix of job switches between HD and LD sections. Only 89 out of 585 workers (15.2%) reported to be currently in jobs that involved switching *dependency* from their first job. Only 7.2% of the workers reported to switch jobs more than once, whereby majority of the workers who switched jobs did so only once. Additionally, many of these changes resulted from a closure of one production line at the factory in 2018. As a result, workers from that line were reallocated, typically to similar jobs, in the same shift, but to other existing lines and an additional line which was bought around the same time.

Table A.2: Dependency sorting

	(1) **Switched Dependency**	(2) **High to Low**	(3) **Low to High**
Age	0.0041**	-0.0032	-0.0010
	(0.0019)	(0.0018)	(0.0011)
Tenure	-0.0046	0.0061	-0.0014
	(0.0049)	(0.0042)	(0.0027)
Schooling (Highest Grade)	0.0021	0.0008	-0.0030
	(0.0045)	(0.0038)	(0.0019)
Muslim	-0.0113	0.0173	-0.0056
	(0.038)	(0.0251)	(0.0285)
Line × Section F.E. (First Job)	Yes	Yes	Yes
N	584	584	584
Adj. R^2	0.053	0.042	0.276

* p<0.10, ** p<0.05, *** p<0.010. The unit of observation is an individual worker. Workers were asked to report their first job at the factory and their last job before the intervention began. "Switched Dependency" refers to whether the move (if any) between the first and last job involved changing Dependency as well. Workers were also asked to report if they held any other job in between. Only 7.2 % reported that they did. Workers are categorized into the the following skill categories: unskilled, semi-skilled or operators. Unskilled workers are the omitted group. Standard errors clustered at the worker's initial (first) line-section-level.

Overall, this suggests that only a small share of workers switched the *type* (HD or LD) of job that they had when they first joined. This rules out systematic sorting into tasks over time and potential selection bias resulting from it. Nevertheless, in Table A.2, I test whether observable characteristics of workers are correlated with the probability of moving across task types, based on the few moves that have occurred (as shown in Table A.1). As observed, none of the factors (age, tenure, schooling, religion) which could potentially affect sorting over time, are statistically significant in Column (1) (while age is statistically significant it's magnitude is very small). In Columns (2) and (3), I split up job switches from HD to LD and LD to HD sections. Again, the coefficients on the the usual factors are small in magnitude and not statistically significant. Overall, the evidence suggests that a large majority of workers remained in the task that they joined upon employment, and the few switches that have occurred seemed to be determined by organizational requirements rather than workers selecting into jobs.

Table A.3: Balance in proportion Muslim

	(1) Proportion Muslim
HD vs LD mixed sections	0.0416 (0.0488)
Mean Dep. Var.	0.36
Line F.E.	Yes
N	56
Adj. R^2	0.235

* p<0.10, ** p<0.05, *** p<0.010. The unit of observation is a line-section. Standard errors clustered at the production line-level. This table shows that the proportion of Muslim workers in mixed teams was balanced across HD and LD sections after randomization.

Table A.3 shows that the share of Muslim workers was balanced across HD and LD tasks after randomization. This is important to rule out that the different effects of religious mixing in HD and LD tasks are caused by different “degrees" of mixing rather than the effects being driven by the production technology. Finally, in Table A.4, I report balance in work characteristics across treatment arms without the inclusion of line × section fixed effects. Therefore, unlike in Table 3, the main effect of being in HD versus LD section *is* identified. If workers were able to systematically sort into HD and LD tasks based on certain observable characteristics, then the main effect of HD versus LD should pick these differences up. This however is not the case, it can be observed that worker characteristics are balanced between HD and LD sections overall.

Table A.4: Randomization check (HD vs LD main effect)

	Panel A: Outcomes relevant at work				Panel B: Other outcomes				
	Tenure	Muslim co-workers *Hindus*	Taking Orders	Communicating	Age	Schooling	Trust	Altruism	Inter-religious contact outside work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mixed	0.0827	0.0022	0.0586	-0.0582	1.3766	0.0026	0.5226	0.0693	0.0430
	(0.4613)	(0.0122)	(0.0582)	(0.0679)	(1.4830)	(0.6346)	(0.3723)	(0.2414)	(0.0516)
HD	-0.4920	0.0123	0.0075	0.0045	1.0955	-0.3364	0.2086	0.1120	0.0227
	(0.3993)	(0.0107)	(0.0563)	(0.0636)	(1.1338)	(0.4874)	(0.3752)	(0.2132)	(0.0491)
Mixed × HD	-0.1742	-0.0072	-0.0584	0.0946	-1.0592	0.3178	-0.5677	0.0057	-0.0405
	(0.5609)	(0.0157)	(0.0826)	(0.0858)	(1.6794)	(0.7187)	(0.4923)	(0.2742)	(0.0653)
Mean Dep Var.	4.44	0.16	0.73	0.47	34.21	7.84	3.82	6.66	0.45
Production Line F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Religion F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	585	480	586	586	586	586	583	583	586
Adj. R^2	0.090	0.258	0.0010	0.028	0.024	0.016	0.002	0.005	0.107

* p<0.10, ** p<0.05, *** p<0.010. The unit of observation is an individual worker. Standard errors clustered at the line-section-level team. "Tenure" and "Schooling" are measured in years and as highest grade completed respectively. "Taking Orders" is a dummy variable coded 1 if the respondent reported to be "Always comfortable" taking orders from non-coreligionists and 0 if they reported to be "Sometimes uncomfortable" or "Always uncomfortable". "Communicating" is coded 1, 0.5 and 0 for the responses "Always comfortable", "Sometimes uncomfortable" and "Always uncomfortable" when asked about being comfortable communicating with non-coreligionists. Survey questions on "Trust" and "Altruism" are used from the World Value Survey (WVS). The dependent variable "Inter-religious contact" refers to the degree of cross-religion interaction that workers had at baseline, outside of work. The variable is coded 1, 0.5 and 0 if a worker mentioned that during the daily course of their life they: 1) interact with more than 5 non-coreligionists 2) interact with 1 to 5 non-coreligionists, or 3) do not interact with anyone outside their religion, respectively.

Table A.5: Randomization check (Line-level treatment indicator)

	Panel A: Outcomes relevant at work					Panel B: General characteristics and attributes			
	Tenure	Muslim co-workers *Hindus*	Taking Orders	Communicating	Age	Schooling	Trust	Altruism	Inter-religious contact outside work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
HD-Mixed Line vs	-0.0333	-0.0004	-0.0175	0.0705**	-0.2144	0.0213	-0.1853	0.0079	-0.0106
LD-Mixed Line	(0.4425)	(0.0080)	(0.0266)	(0.0266)	(1.1556)	(0.2377)	(0.3184)	(0.1268)	(0.0366)
Bootstrap (Wild Cluster) C.I.	[-1.26, 2.16]	[-0.02, 0.04]	[-0.09, 0.04]	[-0.03, 0.16]	[-2.91, 3.73]	[-0.59, 0.49]	[-1.11, 0.55]	[-0.31, 0.42]	[-0.20, 0.07]
Mean Dep. Var	4.45	0.159	0.73	0.47	33.88	7.92	3.88	6.68	0.45
Line Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Religion F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	557	457	557	557	557	557	554	554	557
Adj. R^2	0.083	0.211	0.0037	0.0291	0.000	0.019	-0.007	0.011	0.111

* p<0.10, ** p<0.05, *** p<0.010. The unit of observation is an individual worker. Standard errors clustered at the line-level. "Tenure" and "Schooling" are measured in years and as highest grade completed respectively. "Taking Orders" is a dummy variable coded 1 if the respondent reported to be “Always comfortable" taking orders from non-coreligionists and 0 if they reported to be “Sometimes uncomfortable" or “Always uncomfortable". "Communicating" is coded 1, 0.5 and 0 for the responses “Always comfortable”, “Sometimes uncomfortable” and “Always uncomfortable” when asked about being comfortable communicating with non-coreligionists. Survey questions on “Trust" and “Altruism" are used from the World Value Survey (WVS). The dependent variable "Inter-religious contact" refers to the degree of cross-religion interaction that workers had at baseline, outside of work. The variable is coded 1, 0.5 and 0 if a worker mentioned that during the daily course of their life they: 1) interact with more than 5 non-coreligionists 2) interact with 1 to 5 non-coreligionists, or 3) do not interact with anyone outside their religion, respectively. This sample excludes individuals who work in common sections (“Egg" and “Flour") that cater to all production lines, but themselves are not part of any particular line.

B Additional tables referred to in the main text (Online Appendix)

B.1 Summary Statistics

Table B.1: Summary Statistics: Hindu and Muslim workers

Variable	**Hindu**	**Muslim**	**Diff (2) - (1)**	**N**
Panel A: Dependency				
High Dependency (share of workers)	0.610	0.660	0.052	586
	(0.02)	(0.05)	(0.052)	
Panel B: Schooling and Tenure				
Schooling (Grade)	8.05	6.90	-1.152***	586
	(0.16)	(0.33)	(0.370)	
Tenure	4.83	2.71	-2.116***	585
	(0.15)	(0.28)	(0.345)	
Panel C: Cross-religion interaction and attitudes				
Cross-religion interaction (outside work)	0.39	0.73	0.338***	586
	(0.02)	(0.03)	(0.040)	
Comfortable taking orders from non-coreligionists	0.72	0.75	0.030	586
	(0.02)	(0.040)	(0.048)	
Would live next door to non-coreligionists	0.54	0.87	0.334***	541
	(0.02)	(0.02)	(0.039)	
Equally comfortable communicating with non-coreligionists	0.49	0.69	0.196***	586
	(0.02)	(0.04)	(0.046)	
Panel D: Political				
Supports National Registrar of Citizens (NRC)	0.43	0.24	-0.196***	444
	(0.02)	(0.04)	(0.049)	

* p<0.10, ** p<0.05, *** p<0.010. Standard errors in parentheses. "Cross-religion interaction (outside work)" is a categorical variable coded 1, 0.5 and 0 if an individual reported to come in contact with “greater than 5", “between 1 and 5" or “0" non-coreligionists respectively in their daily life outside of work. "Comfortable taking orders from non-coreligionists" is a dummy variable coded 1 if the respondent reported to be “Always comfortable" taking orders from non-coreligionists and 0 if they reported to be “Sometimes uncomfortable" or “Always uncomfortable". "Equally comfortable communicating with non-coreligionists" is coded 1, 0.5 and 0 for the responses “Always comfortable”, “Sometimes uncomfortable” and “Always uncomfortable” respectively.

Table B.2: Summary statistics: Mean differences (physical environment)

Variable	**Low-Dependency**	**High-Dependency**	**Diff (2) - (1)**
Panel A: Interaction (Minutes out of 10)			
Direct Dependency	2.14	9.48	7.339***
	(0.63)	(0.10)	(0.669)
Non-work Interaction	0.61	1.11	0.501
	(0.18)	(0.24)	(0.302)
Panel B: Noise Level (Decibels)			
Avg Noise (Db)	79.17	75.95	-3.230
	(1.36)	(1.42)	(1.969)
Panel C: Temperature (Celsius)			
Section Temperature (°C)	28.51	30.63	2.123
	(1.00)	(0.78)	(1.288)
N	22	20	42

* p<0.10, ** p<0.05, *** p<0.010. This table reports mean differences in various characteristics of HD and LD tasks.

B.2 Robustness checks and additional results

Table B.3: Treatment effect on output (Line × Variety fixed effects)

	(1) **Log Output (Pieces)**	(2) **Log Output (Boxes)**
HD-Mixed Line (LD Non-Mixed)	-0.0555***	-0.0520***
	(0.0141)	(0.0129)
Bootstrap Wild cluster C.I.	[-0.088, -0.021]	[-0.082, -0.0163]
Day F.E.	Yes	Yes
Shift F.E.	Yes	Yes
Production Line × Variety F.E.	Yes	Yes
Mean Dep Var	10.81	6.97
	(1.02)	(0.912)
N	1018	1017
Adj. R^2	0.887	0.855

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily output produced by line-level teams. Standard errors clustered at the line-level team in parenthesis. Wild cluster bootstrap confidence intervals in square brackets. HD-Mixed Line is a dummy coded 1 for a line-level team with all HD sections religiously mixed and LD sections non-mixed, and 0 for exactly the opposite line-level structure (LD-Mixed Line).

Table B.4: Treatment effect on output (Line × Day fixed effects)

	(1) **Log Output (Pieces)**	(2) **Log Output (Boxes)**
HD-Mixed Line (LD Non-Mixed)	-0.0520** (0.0184)	-0.0533* (0.0255)
Bootstrap Wild cluster C.I.	[-0.087, -0.006]	[-0.111, 0.012]
Shift F.E.	Yes	Yes
Production Line × Day F.E.	Yes	Yes
Mean Dep. Var.	10.81 (1.02)	6.96 (0.912)
N	1018	1018
Adj. R^2	0.898	0.822

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily output produced by line-level teams. Standard errors clustered at the line-level team in parenthesis. Wild cluster bootstrap confidence intervals in square brackets. HD-Mixed Line is a dummy coded 1 for a line-level team with all HD sections religiously mixed and LD sections non-mixed, and 0 for exactly the opposite line-level structure (LD-Mixed Line).

Table B.5: Deposit/Forming quantity and line-section-level ratings

	Log(Forming Quantity)
	(1)
Rating: Deposit	0.2023*** (0.0502)
Rating: Non-Deposit	0.1246 (0.0786)
Observations	1045
Mean Dep. Var	3.55
Day Effects	Yes
Line Effects	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.010. Observations are daily deposit/forming quantities produced by line-level teams. Robust standard errors are reported in parenthesis. Rating: Deposit denotes the rating received by the Deposit section of a line-level team, whereas Rating: Non-Deposit denotes the average rating received by all other sections sections.

Table B.6: Treatment effect on section ratings

	HD Sections		LD Sections	
	Rating	Rating > Median	Rating	Rating > Median
	(1)	(2)	(3)	(4)
Mixed	-0.0453***	-0.0445***	-0.0071	-0.0055
	(0.0170)	(0.0105)	(0.0128)	(0.0127)
Mean Dep. Var.	3.89	0.43	3.89	0.43
	(0.67)	(0.50)	(0.67)	(0.50)
Day F.E.	Yes	Yes	Yes	Yes
Shift F.E.	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes
N	3713	3713	3914	3914
Adj. R^2	0.614	0.391	0.601	0.336

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. “Education" and “Tenure" control for average levels of schooling and tenure of workers in the line-section-level team. In this table, the sample is split into HD and LD sections and it can be observed that while religious mixing leads to lower ratings in HD sections, the effects are small and not statistically significant in LD.

Table B.7: Treatment effect on section ratings (with supervisor fixed effects)

	Rating (Raw)		**Rating > Median**	
	(1)	(2)	(3)	(4)
Mixed	-0.0240**		-0.0249***	
	(0.0109)		(0.0090)	
Mixed × LD		-0.0114		-0.0023
		(0.0151)		(0.0123)
Mixed × HD		-0.0386**		-0.0511***
		(0.0184)		(0.0138)
p(Mixed × HD = Mixed × LD)		0.28		0.012
Mean Dep. Var.	3.86	3.86	0.47	0.47
	(0.65)	(0.65)	(0.50)	(0.50)
Day F.E.	Yes	Yes	Yes	Yes
Shift F.E.	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes
Supervisor Effects	Yes	Yes	Yes	Yes
N	7590	7590	7590	7590
Adj. R^2	0.608	0.608	0.366	0.366

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Section fixed effects are included in all specifications; as a result the main effect of HD versus LD is not separately identified in columns (2) and (4).

Table B.8: Treatment effect on section ratings (with education and tenure controls)

	Rating (Raw)		**Rating > Median**	
	(1)	(2)	(3)	(4)
Mixed	-0.0210**		-0.0237***	
	(0.0108)		(0.0088)	
Mixed × LD		-0.0122		-0.0072
		(0.0128)		(0.0124)
Mixed × HD		-0.0310*		-0.0425***
		(0.0173)		(0.0114)
p(Mixed × HD = Mixed × LD)		0.39		0.034
Mean Dep. Var.	3.86	3.86	0.47	0.47
	(0.65)	(0.65)	(0.50)	(0.50)
Education and Tenure Controls	Yes	Yes	Yes	Yes
Day F.E.	Yes	Yes	Yes	Yes
Shift F.E.	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes
N	7627	7627	7627	7627
Adj. R^2	0.602	0.602	0.365	0.365

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Section fixed effects are included in all specifications; as a result the main effect of HD versus LD is not separately identified in columns (2) and (4). "Education" and "Tenure" control for average levels of schooling and tenure of workers in the line-section-level team.

Table B.9: Treatment effect on section ratings: Different thresholds to define HD vs LD tasks

	HD (> 4 mins)	HD (> 6 mins)	HD (> 8 mins)
	(1)	(2)	(3)
Mixed × LD	-0.0012	-0.0120	-0.0061
	(0.0184)	(0.0150)	(0.0144)
Mixed × HD	-0.0368***	-0.0351**	-0.0409***
	(0.0129)	(0.0148)	(0.0150)
Observations	7459	7459	7459
p(Mixed × HD = Mixed × LD)	0.12	0.28	0.1
Mean Dep. Var	3.86	3.86	3.86
Day Effects	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Section fixed effects are included in all specifications; as a result the main effect of HD versus LD is not separately identified.

Table B.10: Treatment effect on section ratings: Dynamic effects

	Raw Ratings			
	HD Sections		LD Sections	
	(1)	(2)	(3)	(4)
Mixed	-0.0453***		-0.0071	
	(0.0170)		(0.0128)	
Mixed × 0-25 days		-0.1034*		0.0254
		(0.0683)		(0.0594)
Mixed × 26-50 days		-0.0713**		-0.0948*
		(0.0352)		(0.0501)
Mixed × 51-75 days		0.0284		-0.0172
		(0.0334)		(0.0363)
Mixed × 76-100 days		-0.0635**		0.0651*
		(0.0309)		(0.0278)
Mixed × 101-120 days		-0.0240		-0.0462
		(0.0417)		(0.0344)
Mean Dep. Var.	3.89	3.89	3.84	3.84
	(0.67)	(0.67)	(0.64)	(0.64)
Day F.E.	Yes	Yes	Yes	Yes
Shift F.E.	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes
N	3713	3648	3914	3847
Adj. R^2	0.614	0.613	0.601	0.600

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Section fixed effects are included in all specifications; as a result the main effect of HD versus LD is not separately identified in columns (2) and (4). “Education" and “Tenure" control for average levels of schooling and tenure of workers in the line-section-level team. This table is based on specification 2; interactions of “Mixed" with day bins are added.

Table B.11: Treatment effect on worker interactions: Hindus respondents only

	Identified teammate as contributing low effort		**Blamed by teammate**		**Unwilling to give up relief time**	
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed	0.0399***		0.0393**		0.0565	
	(0.0139)		(0.0165)		(0.0369)	
Mixed × LD		0.0309		0.0656**		0.0317
		(0.0274)		(0.0260)		(0.0636)
Mixed × HD		0.0421***		0.0330*		0.0631
		(0.0159)		(0.0191)		(0.0441)
p(Mixed × LD = Mixed × HD)		0.665		0.282		0.573
Mean. Dep. Var.	0.14	0.14	0.08	0.08	0.24	0.24
Worker skill F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes	Yes	Yes
N	3020	3020	3009	3009	3056	3056
Adj. R^2	0.015	0.015	0.013	0.013	0.079	0.079

* p<0.10, ** p<0.05, *** p<0.010. Observations are at the worker-teammate level for line-section-level teams i.e. there are (N-1) observations per worker, where N denotes the number of workers in the section. Standard errors clustered at the line-section-level team. The outcome variables are as follows – workers were asked to choose teammates who they: (1) thought did not contribute sufficient effort at any point during the intervention (2) have been blamed by during the intervention and (3) have (our would be willing to) given up their relief time for.

Table B.12: Treatment effect on section ratings: Adding key controls

	Rating (Raw)			
	(1)	(2)	(3)	(4)
Mixed × LD	-0.0141	-0.0358**	-0.1140***	-0.1224**
	(0.0135)	(0.0171)	(0.0408)	(0.0480)
Mixed × HD	-0.0363**	-0.0795***	-0.1563***	-0.1644***
	(0.0162)	(0.0243)	(0.0413)	(0.0476)
Mixed × Group Size		0.0062*	0.0096***	0.0093**
		(0.0032)	(0.0036)	(0.0038)
Mixed × Tenure			0.0130**	0.0121*
			(0.0058)	(0.0062)
Mixed × Schooling				0.0017
				(0.0050)
p(Mixed × HD = Mixed × LD)	0.29	0.05	0.05	0.05
Mean Dep. Var.	3.86	3.86	3.86	3.86
	(0.65)	(0.65)	(0.65)	(0.65)
Day F.E.	Yes	Yes	Yes	Yes
Shift F.E.	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes
N	7627	7627	7627	7627
Adj. R^2	0.607	0.607	0.607	0.607

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Section fixed effects are included in all specifications; as a result the main effect of HD versus LD is not separately identified. “Education" and “Tenure" control for average levels of schooling and tenure of workers in the line-section-level team.

Table B.13: Determinants of worker promotion

	Promoted (Semi-skilled/Operator)
	(1)
Schooling	0.0089*
	(0.0050)
Tenure	0.0629***
	(0.0065)
Age	-0.0026
	(0.0023)
Muslim	-0.0442
	(0.0380)
Observations	584
Mean Dep. Var	0.23
Line × Section (First Job) F.E.	Yes

* p<0.10, ** p<0.05, *** p<0.010. The outcome is a dummy variable coded 1 if a worker’s skill designation is semi-skilled or operator (as opposed to being unskilled). Standard errors clustered at the worker’s initial (first) line-section-level.

Table B.14: Proportion of old teammates

	Proportion of old teammates	
	(1)	(2)
Mixed	-0.00890	
	(0.0163)	
Mixed × LD		-0.0291
		(0.0314)
Mixed × HD		0.00223
		(0.0195)
p(Mixed × HD = Mixed × LD)		0.425
Mean Dep. Var.	0.34	0.34
Religion F.E.	Yes	Yes
Line × Section F.E.	Yes	Yes
N	586	586
Adj. R^2	0.580	0.580

* p<0.10, ** p<0.05, *** p<0.010. The unit of observation is an individual worker. Standard errors clustered at the line-section-level team. The outcome variable in these regressions is the share of co-workers in each individual's line-section-level team that were also in their team at baseline. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in column (2).

Table B.15: Section change and treatment status

	Changed Section			
	(1)	(2)	(3)	(4)
Mixed	-0.0344	-0.0349		
	(0.0278)	(0.0252)		
Mixed × LD			-0.0377	0.0123
			(0.0542)	(0.0364)
Mixed × HD			-0.0323	-0.0610*
			(0.0359)	(0.0322)
Mean Dep. Var.	0.085	0.085	0.085	0.085
Religion F.E.	Yes	Yes	Yes	Yes
Line × Section F.E.	No	Yes	No	Yes
Line × Old Section F.E.	Yes	No	Yes	No
N	586	586	586	586
Adj. R^2	0.113	0.042	0.111	0.043

* p<0.10, ** p<0.05, *** p<0.010. The unit of observation is an individual worker. Standard errors clustered at the line-section-level team. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in columns (3) and (4). These are individual worker-level regressions. The outcome variable is a dummy coded 1 if after the randomization process the worker was in a different section (task) than their section of work at baseline.

Table B.16: Baseline outgroup exposure (outside of the factory) and ratings

Baseline Outgroup Exposure:	Low	High
	(1)	(2)
Mixed × LD	-0.0149	-0.0042
	(0.0290)	(0.0348)
Mixed × HD	-0.0797***	-0.0598**
	(0.0299)	(0.0248)
Observations	3931	3696
Mean Dep. Var	3.85	3.88
Day Effects	Yes	Yes
Shift Effects	Yes	Yes
Line × Section Effects	Yes	Yes

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Section fixed effects are included in all specifications. "Baseline exposure (High versus Low)" refers to the degree of cross-religion interaction that workers had at baseline, outside of work. It is derived from a variable is coded 1, 0.5 and 0 if a worker mentioned that during the daily course of their life they: 1) interact with more than 5 non-coreligionists 2) interact with 1 to 5 non-coreligionists, or 3) do not interact with anyone outside their religion, respectively. We use above/below the median value for this variable to denote High/Low exposure.

Table B.17: Hindu-Muslim differences in effort

	Willing to give up relief time
	(1)
Muslim	0.0478**
	(0.0218)
Mean Dep. Var	0.72
Worker Skill F.E.	Yes
Line × Section Effects	Yes
Observations	2035

* p<0.10, ** p<0.05, *** p<0.010. Observations are at the worker-teammate level for line-section-level teams i.e. there are (N-1) observations per worker, where N denotes the number of workers in the section. Robust standard errors in parenthesis. The outcome variable is a dummy that takes the value 1 if a worker responded in the affirmative when asked if they have (or would be willing to) give up their relief time for the teammate specific teammate in question.

Table B.18: Worker interactions: Decomposition (Mixed teams)

	Identified teammate as contributing low effort		**Blamed by teammate**		**Unwilling to give up relief time**	
	(1)	(2)	(3)	(4)	(5)	(6)
Target Muslim	0.0528***	0.0861***	-0.0159	0.0048	0.0474***	0.0090
	(0.0175)	(0.0225)	(0.0126)	(0.0213)	(0.0176)	(0.0332)
Respondent Muslim	-0.0172	0.0152	-0.0009	0.0192	-0.0450	-0.0830**
	(0.0221)	(0.0266)	(0.0199)	(0.0291)	(0.0316)	(0.0356)
Target Muslim ×		-0.0995**		-0.0612		0.1139*
Respondent Muslim		(0.0485)		(0.0406)		(0.0657)
Mean Dep. Var	0.15	0.15	0.10	0.10	0.28	0.28
Worker Skill F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Line × Section F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2033	2033	2029	2029	2035	2035
Adj. R^2	0.018	0.025	0.013	0.016	0.064	0.084

* p<0.10, ** p<0.05, *** p<0.010. Observations are at the worker-teammate level for line-section-level teams i.e. there are (N-1) observations per worker, where N denotes the number of workers in the section. Standard errors clustered at the line-section-level team. The outcome variables are as follows – workers were asked to choose teammates who they: (1) thought did not contribute sufficient effort at any point during the intervention (2) have been blamed by during the intervention and (3) have (or would be willing to) given up their relief time for.

Table B.19: Treatment effect on worker interactions: Decomposition (Mixed teams by dependency)

	Identified teammate as contributing low effort		**Blamed by teammate**		**Unwilling to give up relief time**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: HD Sections						
Target Muslim	0.0711***	0.0987***	-0.0139	0.0056	0.0534**	0.0325
	(0.0185)	(0.0276)	(0.0145)	(0.0360)	(0.0215)	(0.0410)
Respondent Muslim	-0.0202	0.0053	-0.0147	0.0081	-0.0423	-0.0620
	(0.0274)	(0.0328)	(0.0255)	(0.0252)	(0.0408)	(0.0438)
Target Muslim ×		-0.0829		-0.0656		0.0624
Respondent Muslim		(0.0609)		(0.0495)		(0.0815)
Mean Dep. Var.	0.16	0.16	0.09	0.09	0.31	0.31
N	1576	1576	1584	1584	1568	1568
Adj. R^2	0.023	0.025	0.019	0.021	0.107	0.107
Panel B: LD Sections						
Target Muslim	0.0008	0.0502*	0.0392*	0.0672*	0.0290	-0.0635*
	(0.0276)	(0.0284)	(0.0205)	(0.0371)	(0.0291)	(0.0321)
Respondent Muslim	-0.0080	0.0462	-0.0196	0.0055	0.0505	-0.1538***
	(0.0336)	(0.0382)	(0.0250)	(0.0357)	(0.0386)	(0.0290)
Target Muslim ×		-0.1443**		-0.0726		0.2665***
Respondent Muslim		(0.0512)		(0.0628)		(0.0754)
Mean Dep. Var	0.13	0.13	0.12	0.12	0.19	0.19
N	457	457	445	445	467	467
Adj. R^2	0.01	0.02	0.01	0.01	-0.01	0.02

* p<0.10, ** p<0.05, *** p<0.010. Observations are at the worker-teammate level for line-section-level teams i.e. there are (N-1) observations per worker, where N denotes the number of workers in the section. Standard errors clustered at the line-section-level team. The outcome variables are as follows – workers were asked to choose teammates who they: (1) thought did not contribute sufficient effort at any point during the intervention (2) have been blamed by during the intervention and (3) have (or would be willing to) given up their relief time for. All regressions include Worker skill F.E. and Line × Section F.E.

Table B.20: Religious violence and section ratings

	HD Sections		**LD Sections**	
	(1)	(2)	(3)	(4)
Mixed	-0.0453***		-0.0071	
	(0.0170)		(0.0128)	
Mixed × No Violence		-0.0321**		0.0100
		(0.0134)		(0.0166)
Mixed × Violence		-0.1131*		-0.0988*
		(0.0661)		(0.0568)
p(Mixed × No Violence = Mixed × Violence)		0.217		0.104
Mean Dep. Var.	3.86	3.86	3.84	3.84
	(0.68)	(0.68)	(0.64)	(0.64)
Education and Tenure Controls	Yes	Yes	Yes	Yes
Day Effects	Yes	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes
N	3713	3713	3914	3914
Adj. R^2	0.615	0.615	0.601	0.602

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Section fixed effects are included in all specifications. “Education" and “Tenure" control for average levels of schooling and tenure of workers in the line-section-level team. Between 13th-18th December 2019, immediately after the passing of the Citizenship Amendment Act (CAA) violent protests erupted in the district of West Bengal where the factory is located. Hindu-Muslim riots occurred in Delhi between 23rd-28th Feb 2020 during protests against the CAA as well. These days are coded as violent days in these regressions.

Table B.21: Attrition

	Attrited (1)	**Attrited** (2)
Mixed	-0.0164	
	(0.0223)	
Mixed × LD		0.0069
		(0.0279)
Mixed × HD		-0.0292
		(0.0296)
p(Mixed × HD = Mixed × LD)		0.35
Mean Dep. Var	0.05	0.05
Religion Effects	Yes	Yes
Line × Section Effects	Yes	Yes
Observations	586	586

* p<0.10, ** p<0.05, *** p<0.010. The unit of observation is an individual worker. The outcome variable is coded 1 for individuals who left the firm before the end of the experiment. Note that the total number of workers who left the firm is actually lower than the number interviewed at endline. A handful of workers could not be reached by phone during the endline survey.

Table B.22: Political attitudes (Hindus) and section ratings

	HD sections	LD sections
	(1)	(2)
Mixed × BJP/NRC support among Hindus (Low)	0.0235	-0.0206
	(0.0259)	(0.0247)
Mixed × BJP/NRC support among Hindus (High)	-0.0685***	-0.0064
	(0.0155)	(0.0185)
Observations	3713	3914
Mean Dep. Var	3.89	3.84
Day Effects	Yes	Yes
Shift Effects	Yes	Yes
Line × Section Effects	Yes	Yes

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. “BJP/NRC support among Hindus" denotes support for the political party BJP and the National Registrar of Citizens (amongst Hindus in a team), both of which may capture anti-Muslim sentiments. It is an indicator variable that takes the value 1 for above-median (“High") support and 0 for below-median (“Low") support (averaging across all Hindus within a team).

Table B.23: Task descriptions and rating method

Task	Description	Rating Method
Mixing	Mixing raw materials into dough. Typically done in batches.	Average number of batches mixed per unit of time (this varies based on the product) during the shift.
Deposit	Forcing of pourable dough through holes of desired shape in baking trays.	Average number of baking trays prepared per hour during the shift.
Oven	Baking the dough or batter in a commercial or industrial oven, maintaining specific temperature, and time settings.	Average number of baking trays successfully prepared for cooling per unit of time (this varies by product).
Tray/Cooling	Moving the baked goods from the oven to the cooling room, allowing them to cool down and reach room temperature, ensuring they retain their structure and texture.	Average number of baking trays successfully cooled and sent to the next section per unit of time (this varies by product)
Injector	Stuffing baked goods with various filling using a machine. It is important to ensure the accuracy and consistency of the fillings to prevent rejected products.	Average number of trays successfully injected and passed onto the next stage per unit of time (typically an hour).
Depanning	Removing the cooled (and injected) baked products from their baking trays or moulds with causing any damage.	Average rate of Depanning per unit of time (typically an hour). Number of pieces damaged (wastage) is also considered.
Packing	Wrapping or sealing the products in suitable packaging to protect their freshness, taste and quality.	Average number of products successfully packed per unit of time (typically an hour). Packing material wastage is also considered.
Cartonning (Cfc)	Placing the packed bakery products into cartons or boxes for easy handling, storage and distribution.	Average number of cartons prepared per hour.

Note: This table describes the core tasks in each production line and the method used by supervisors to evaluate the performance of teams (and give ratings) in them. Note that supervisors were asked to take into account the performance of upstream sections when evaluating performance of a particular section, which is something that is anyway routine for them. Nevertheless section B.3 addresses concerns related to such spillover effects by conducting analysis on sub-samples where this is less likely to be a concern. Sections “1st Line" and “2nd Line" in Line 3 are exactly the same as Deposit sections, whereas the “Cream" sections in Lines 5 and 6 are the same as “Injector" sections. “Box Machine" and “Box Filling" in Lines 5 and 6 are similar to “Packing" sections.

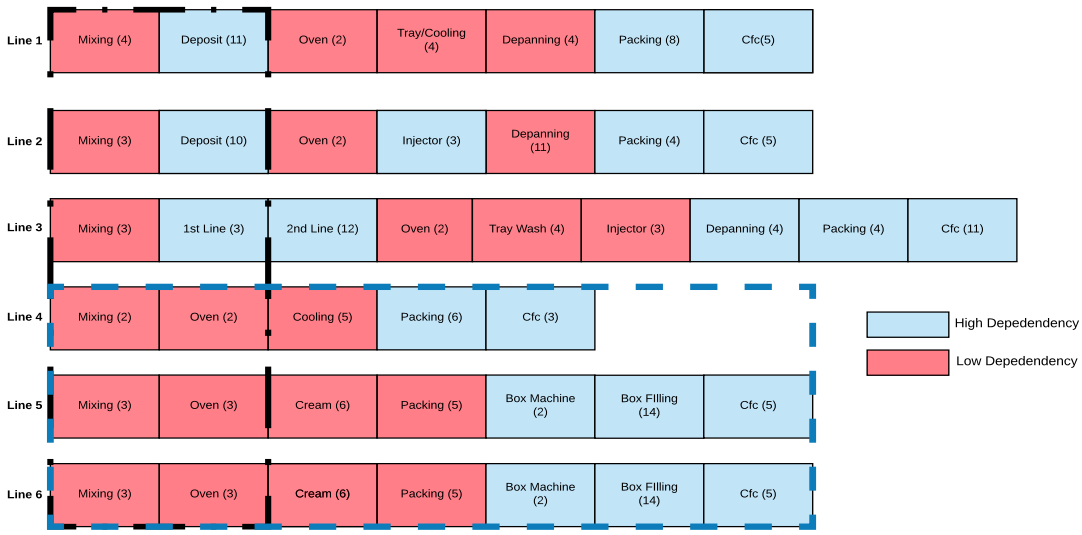
B.3 Spillovers

One concern with the analysis of the treatment effects on output is that there could be spillover effects from upstream to downstream sections, potentially biasing my estimates (even though supervisors tried to each section based solely on it's performance). To understand how this could affect the main findings, I restrict attention to the following two sub-samples (as shown in Figure B.1)

1. Only the first two sections of every line (black dashed-dotted portion)
2. Lines where all HD sections come after all LD sections (blue dashed portion)

With sub-sample 1, I first show that there is no effect of religious mixing in the first section which is always a low-dependency section (and by definition cannot be affected by spillovers) (see columns (1) and (2) of Table B.24.). This is consistent with the main results. Furthermore, this suggests that religious mixing is unlikely to cause differential spillover effects from the first section to the second section based on treatment status. Finally, once the second section of each line is added to the sample, the main section level result (Table B.8) is replicated – we see a negative effect from mixing in HD sections (the magnitude of the effects are also very similar).

Figure B.1: Sub-sample analysis



Note: This figure shows all sections of all lines at the factory split into HD and LD types. Direct Dependency is measured as described in section 2.3. The two relevant sub-samples used for analysis in this section are highlighted by the black dashed-dotted lines and blue dashed lines.

Table B.24: Treatment effect on line-section-level ratings

	Only first section (Mixing, only LD)		**First two sections**	
	Ratings	Ratings > Median	Ratings	Ratings > Median
	(1)	(2)	(3)	(4)
Mixed × LD	-0.0038	-0.0005	-0.0084	-0.0099
	(0.0362)	(0.0260)	(0.0275)	(0.0199)
Mixed × HD			-0.0323	-0.0497***
			(0.0377)	(0.0166)
p(Mixed × HD = Mixed × LD)			0.62	0.14
Mean Dep. Var	3.77	3.77	3.81	3.81
Day Effects	Yes	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes
Observations	1032	1032	2065	2065

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified.

In Table B.25, I show that the main results are replicated with sub-sample 2 as well. This sample is unique in the sense that it only has production lines where all the HD sections come after the LD sections. I once again find that there is no effect of religious mixing in LD sections. The HD sections at the end of the line are therefore unlikely to be affected differentially by spillovers from LD sections (based on whether they are religiously mixed or not). In HD sections, a negative, large and statically significant effect of religious mixing can still be observed. Taken together, the sub-sample analysis is re-assuring in that they convey the same findings as the core results — which is that religious mixing leads to lower team performance but only in HD tasks.

How should we expect line-section-level spillovers to affect the overall line-level treatment effect estimates (Table 4)? Notice that on average production lines have LD sections earlier in the line while HD sections come later. For the line-level effects to be overestimated (in other words the difference in output between HD-Mixed lines and LD-Mixed lines to be more negative than it actually is), it must be the case that there are larger negative spillovers from LD Non-mixed sections to HD-Mixed sections than from LD-Mixed sections to HD Non-Mixed sections, which is unlikely. Therefore, if anything, the line-level results are likely to be underestimated. The fact that the effect sizes in Table B.25 are larger than the main results (Table B.8) is consistent with this.

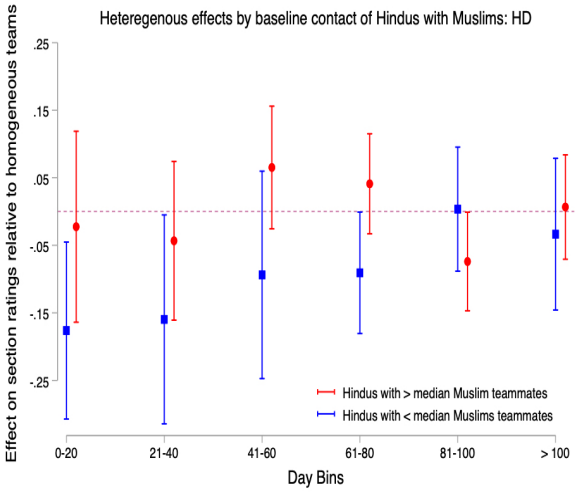
Table B.25: Treatment effect on section ratings (HD after LD)

	Ratings (1)	**Ratings** (2)
Mixed	-0.0353*	
	(0.0181)	
Mixed × LD		0.0009
		(0.0256)
Mixed × HD		-0.0873***
		(0.0302)
p(Mixed × HD = Mixed × LD)		0.04
Mean Dep. Var.	3.95	3.95
Day Effects	Yes	Yes
Shift Effects	Yes	Yes
Line × Section Effects	Yes	Yes
Observations	1929	1929

* p<0.10, ** p<0.05, *** p<0.010. Observations are daily ratings received by line-section-level teams. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in column (2).

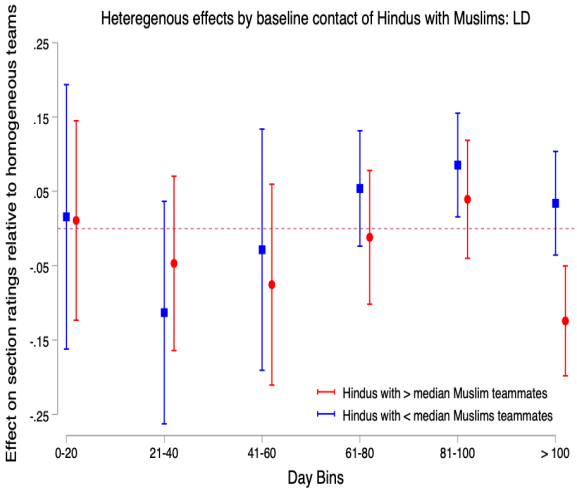
C Additional figures referred to in the main text (Online Appendix)

Figure C.1: Heterogeneous effects by past contact of Hindus with Muslims: HD sections



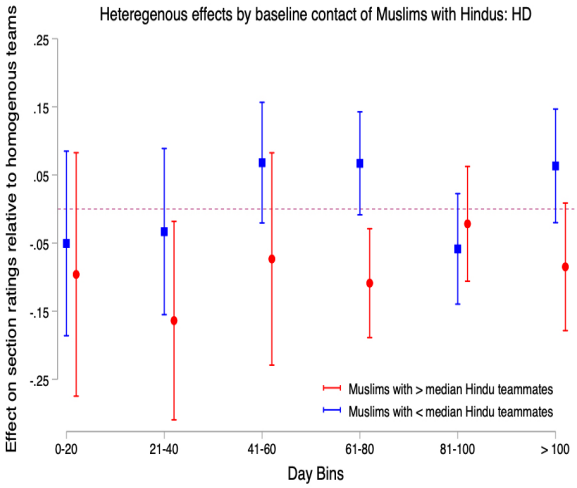
Note: This figure presents dynamic treatment effects of religious mixing in HD sections by past contact that Hindu workers had with Muslims. The intervention period is divided into 6 equal sized bins. 95% confidence intervals are reported.

Figure C.2: Heterogeneous effects by past contact of Hindus with Muslims: LD sections



Note: This figure presents dynamic treatment effects of religious mixing in LD sections by past contact that Hindu workers had with Muslims. The intervention period is divided into 6 equal sized bins. 95% confidence intervals are reported.

Figure C.3: Heterogeneous effects by past contact of Muslims with Hindus: HD sections



Note: This figure presents dynamic treatment effects of religious mixing in HD sections by past contact that Muslim workers had with Hindus. The intervention period is divided into 6 equal sized bins. 95% confidence intervals are reported.

Figure C.4: High- and Low-Dependency tasks



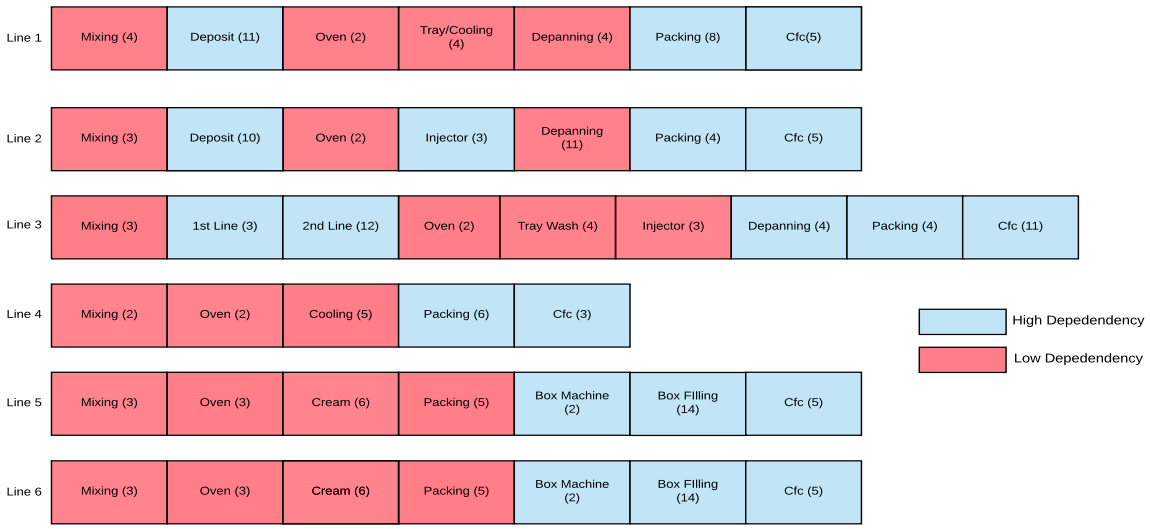
(a) High-Dependency



(b) Low-Dependency

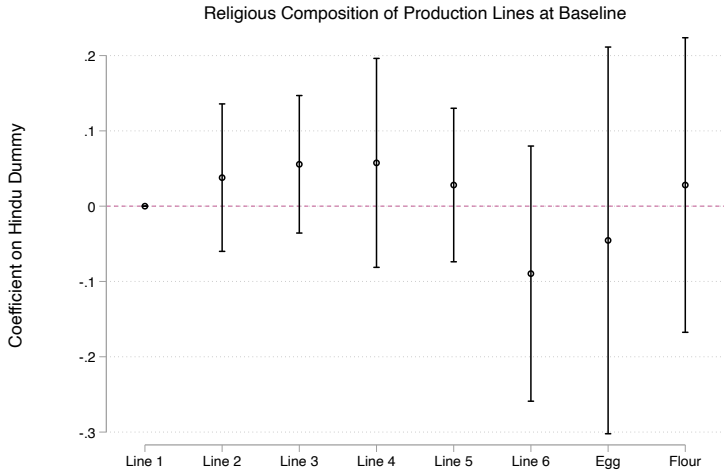
Note: This figure illustrates some key differences between HD and LD tasks. Sub-figure (a) is an example of a HD task. Workers are stood next to each other beside a fast moving conveyor belt. As a group they have to ensure each individual product is put into small packets before they go onto the next stage of production. If the team cannot coordinate and ensure the same, the supervisor has to reduce the speed of the belt to prevent wastage, which in turn would reduce output. Sub-figure (b) is a picture from a mixing room, which is a LD task. One worker is using the weighing scale to weigh raw materials, a second worker is arranging flour buckets and finally a third worker is operating the mixing machine. These workers have to coordinate as well to complete the process, but the frequency of interaction is intermittent and the degree of coordination is much lower relative to the HD task.

Figure C.5: High- and Low-Dependency sections

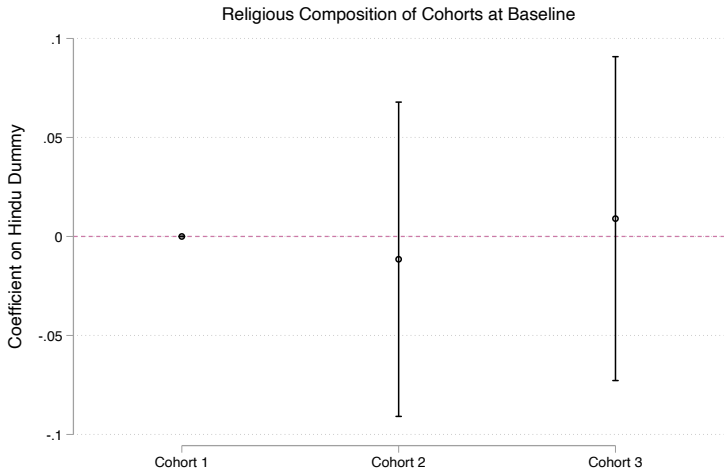


Note: This figure shows all sections of all lines at the factory split into HD and LD types. Direct Dependency is measured as described in section 2.3.

Figure C.6: Religious composition of lines and cohorts at baseline

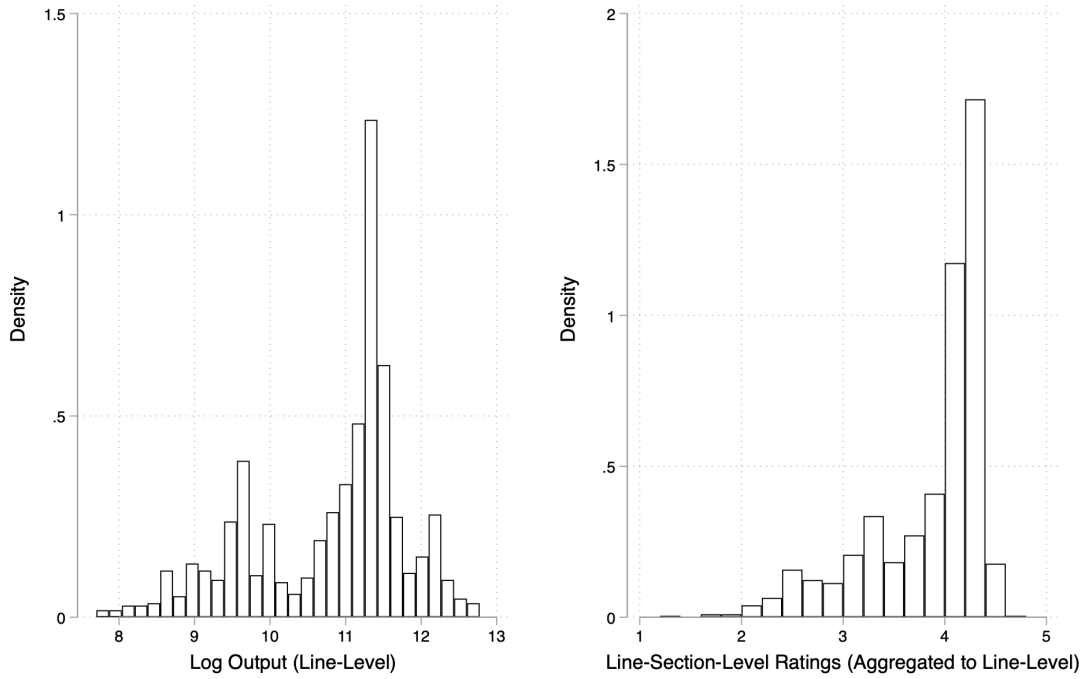


(a) This figure plots coefficients from worker-level regressions. The outcome variable is a dummy coded 1 if the religion of the worker is Hindu and the independent variables are a set of dummy variables denoting each production line. “Egg" and “Flour" refer to production areas where raw materials (eggs and flour) are processed. These production areas are common to all production lines.



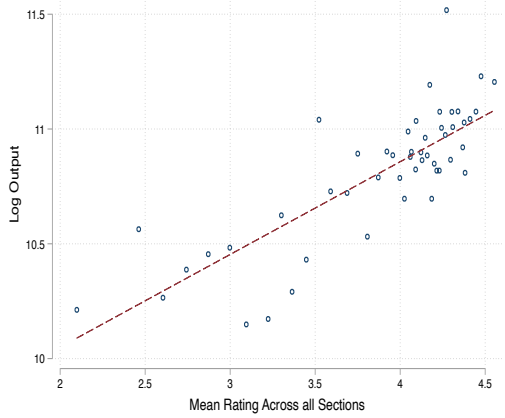
(b) This figure plots coefficients from worker-level regressions. The outcome variable is a dummy coded 1 if the religion of the worker is Hindu and the independent variables are dummies denoting cohorts (groups of workers who work at the factory at the same time).

Figure C.7: Distribution of actual line output and section ratings



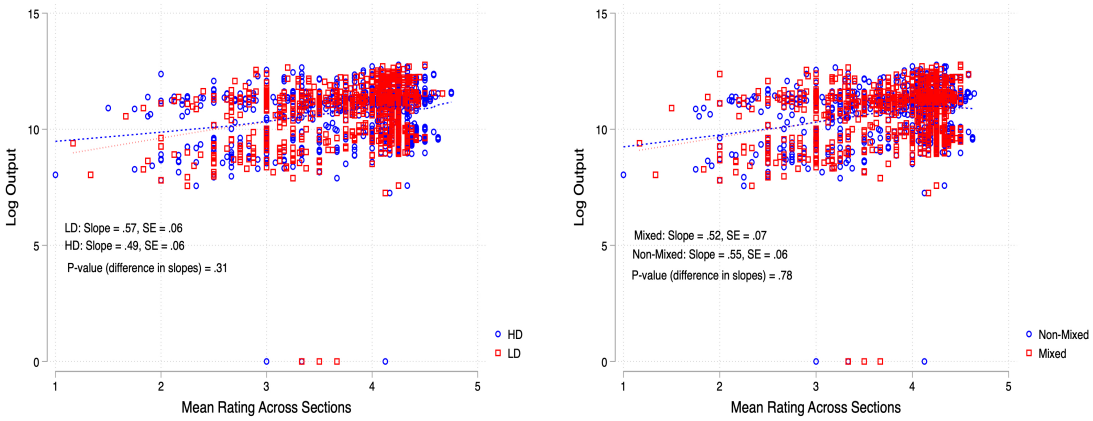
Note: This figure presents the distribution of raw ratings given by production supervisors to line-section-level teams aggregated up to the line-level as well as actual log output at the line-level.

Figure C.8: Line output and section ratings



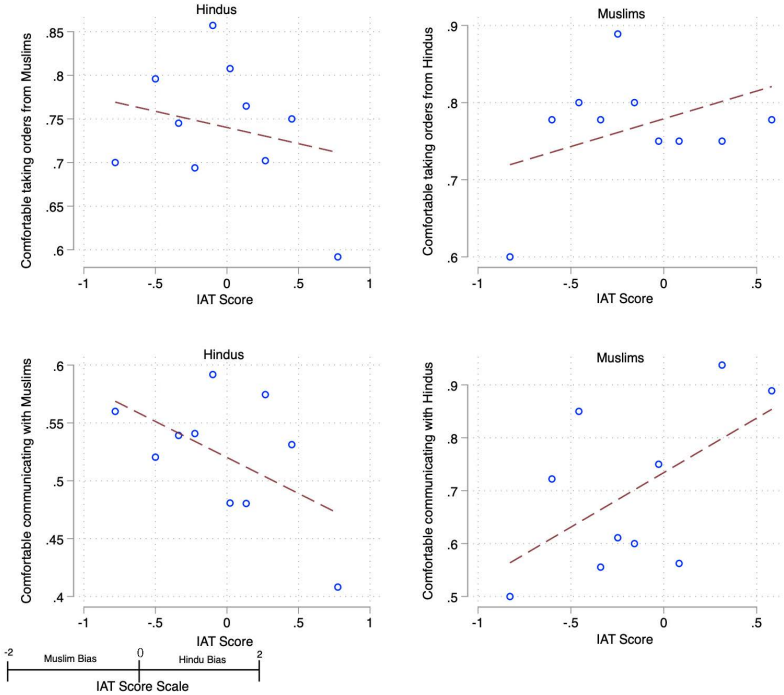
Note: Production line fixed effects are included in this binscatter plot. The variable on the y-axis is daily output (logged) produced by a line-level team, and on the x-axis it is the average value of supervisor ratings received by sections in that line-level team.

Figure C.9: Line output and section ratings (by HD/LD and Mixed/Non-Mixed)



Note: Production line fixed effects are included in these binscatter plots. The variable on the y-axis is daily output (logged) produced by a line-level team, and on the x-axis it is the average value of supervisor ratings received by HD and LD sections in the line-level team.

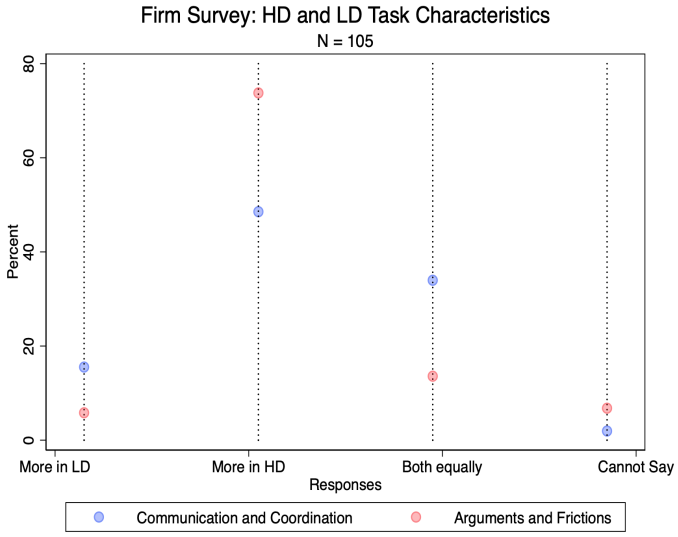
Figure C.10: Correlating IAT scores with survey responses



Note: This figure correlates Implicit Association Test scores (where workers were asked to associate Hindu and Muslim names with positions in the firm's hierarchy) with self-reported survey outcomes of the workers. In the top two figures, the outcome variable on the y-axis is willingness to take orders from non-coreligionists, while in the bottom two figures it is the workers' reported level of comfort in communicating with non-coreligionists. Positive IAT scores denote a bias towards having Hindus in higher positions while a negative value denotes a bias towards Muslims.

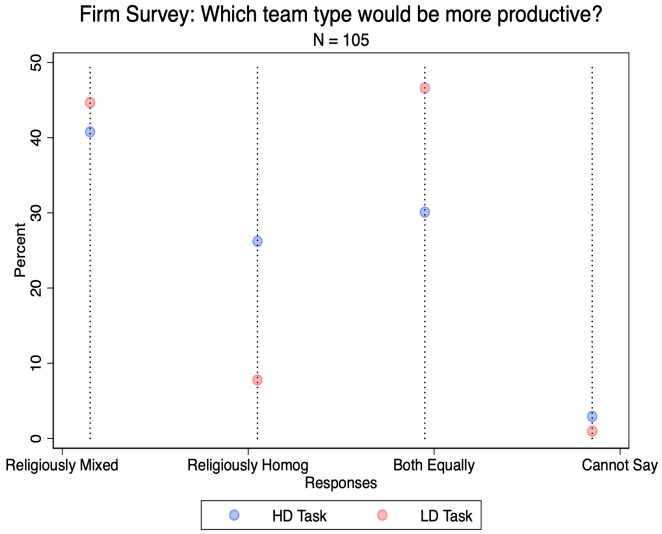
C.1 Figures from firm survey

Figure C.11: Characteristics of HD and LD tasks



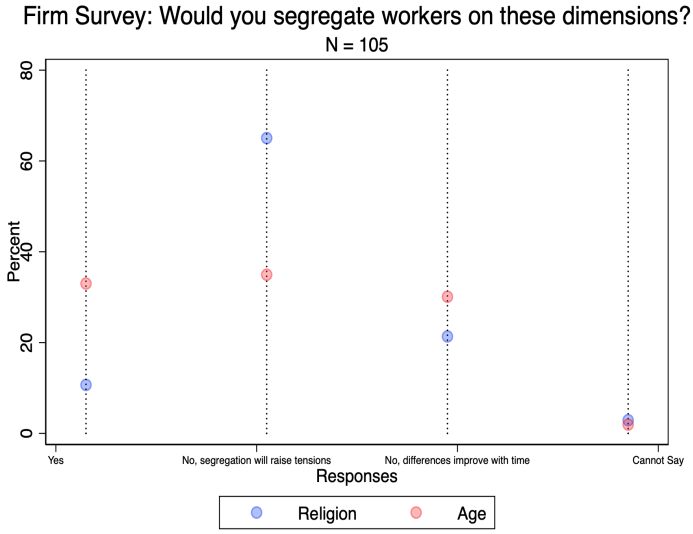
Note: This figure reports the percentage of respondents who picked each option for the following questions: (1) respondents were asked to pick the task that they thought requires greater continuous coordination and communication amongst workers (blue dots), and (2) they were to pick the one that is likely to cause more frictions and arguments amongst workers (pink dots).

Figure C.12: Religious mixing and productivity by task type



Note: This figure reports supervisors' perception of which type of team (religiously homogeneous or mixed) would be more productive in HD vs LD tasks. It reports percentage of respondents who picked each option when they were asked whether a religiously mixed or a homogeneous team would be more productive separately for HD (blue dots) and LD tasks (pink dots).

Figure C.13: Willingness to segregate workers by religion/age



Note: This figure presents responses of supervisors when asked if they are willing to segregate workers based on certain demographic dimensions. Percentage of respondents who chose each option for age and religion are denoted by pink dots and blue dots respectively.

D Model (Online Appendix)

This section presents the theoretical framework. The primary objective of the model is to rationalize the core empirical results, especially the mechanism behind the attenuation in output losses in HD-mixed sections over time. The model makes predictions specially with respect to heterogeneous treatments effects based on worker characteristics, which I subsequently test for in the data.

A key distinction is made between Hindu and Muslim workers in this framework. Consistent with majority-minority relations, a large section of Hindu workers at the factory have never worked with Muslims in the past, while 100% of the Muslim workers have worked with Hindus. Based on this asymmetry in exposure at baseline, together with the evidence on discrimination against Muslims in access to education and labor markets in India (Kalpagam et al., 2010; Basant, 2007),[50] I assume Hindus (mistakenly) on average believe that Muslims are not as hard-working as them.[51] Muslim workers do not make this distinction between in-group and out-group workers given that they have had much greater contact with Hindus. This asymmetry in baseline priors leads to multiple equilibria in HD interactions due to complementarities in the

[50] In fact, Muslims in my sample have significantly lower schooling than Hindus (Table B.1).

[51] An implicit assumption here of course is that in reality Hindus and Muslims are equally productive. I present direct evidence for this in section 5.3.

production function, while in LD this does not matter.

As mentioned, Muslims have accurate priors about Hindus, while Hindus might not necessarily have them, it will depend on past exposure. Muslims are aware that they are being stereotyped by Hindus but can “invest" in shifting the priors of Hindus.

D.1 Setup

Production is composed of two types of tasks, HD (high-dependency) and LD (low-dependency). I make the following assumptions about the production process:

1. There are two workers in each type of task (generalizes to multi-worker easily)
2. There are two types of output: High (O_H) and Low (O_L)
3. Worker effort is the only input in production and it is not observed directly by teammates
4. There are two types of effort: High (e_H) or Low (e_L)
5. Output in each task is a noisy function of worker effort
6. Effort is costly: $c(e_H) > c(e_L) = 0$

Assumptions 1,2,4 and 6 are made for simplicity and easily generalizes to settings where output, effort and effort cost are continuous variables and there are multiple workers in each task. There are several factors that workers do not have direct control over which influence productivity and also make perfectly observing teammate's effort difficult. These include machine breakdowns, inadequate raw material planning and unanticipated production stoppages due to supply chain issues. Assumptions 3 and 5 are made based on these factors.

The production function for HD tasks can be written as:

$$y_{HD}(e_{k1}, e_{k2}) = p(e_{k1}, e_{k2})O_H + \{1 - p(e_{k1}, e_{k2})\}O_L \tag{4}$$

where e_{ki} denotes effort level $k = (H, L)$ for worker $i = (1, 2)$. $p(e_{k1}, e_{k2})$ denotes the probability of high output (O_H) conditional on effort. Clearly, the joint effort of both workers determines the probability of high output and the marginal value of effort is thus higher in teammate's effort level.

The probability of high output conditional on effort levels of both the workers are:

$$(e_H, e_H) = p_H \tag{5}$$

$$(e_H, e_L) = (e_L, e_H) = p_{HL} \tag{6}$$

$$(e_L, e_L) = p_L \tag{7}$$

where $p_H > p_{HL} > p_L$. The production function in LD tasks is linear in worker efforts and is written as:

$$y_{LD}(e_{k1}, e_{k2}) = \sum_{i=1}^{2} \{p(e_{ki})o_h + (1 - p(e_{ki}))o_l\} \tag{8}$$

where o_h and o_l denote high and low individual output levels respectively. In LD sections total output is therefore the sum of individual expected output. The probability of high and low output (conditional on e_H and e_L) are p_H and p_L respectively with $p_H > p_L$.

D.2 One shot production

In HD sections, high effort (e_H) is statically preferred by a worker if and only if their teammate also exerts high effort (e_H). In other words, there is no incentive to free-ride when teammate exerts e_H. Mathematically, this condition implies:

$$(p_H - p_{HL})(O_H - O_L) > c(e_H) > (p_{HL} - p_L)(O_H - O_L)^{52} \tag{9}$$

In LD sections, value of a worker's effort is not dependent on teammate's effort level. As a result we assume e_H is the dominant action, implied by:

$$(p_H - p_{HL})(O_H - O_L) > c(e_H)^{53} \tag{10}$$

D.3 Analysis of the model

Workers interact repeatedly for T periods in a task. HD sections are the interesting case here due to complementarity in worker efforts in the production function. For LD sections, it is clear that priors about teammate's effort would not matter for an individual's effort choice.

D.3.1 Hindu workers (majority group)

Hindu workers who are in mixed teams could have been in non-mixed ones, which they believe would be more productive. In other words, they assign probability π_t (in period t, initial belief

[52] This expression is obtained by re-writing: $p_H O^H + (1 - p_H)O^L - c(e_H) > p_L O^H + (1 - p_L)O^L > p_{HL} O^H + (1 - p_{HL})O^L - c(e_H)$.

[53] The expression is $p_h o^h + (1 - p_h)o^l - c(e_H) > p_l o^h + (1 - p_l)o^l$.

is π_0) on their Muslim teammate exerting high effort, which in the case of coreligionists (other Hindus) is 1. Therefore, with probability (1-π_t) Hindus believe Muslims maybe be "lazy" – a behavioural type that always exerts low effort. This can be thought of as Hindus thinking Muslims have infinite cost of high effort or that they are simply behaviourally disposed to exerting low effort.

A Hindu worker's problem at time $t = 1$ is then given by:

$$V = max_{(e_t)_{t=1}^{T}} \sum_{t=1}^{T} P_t(\pi_t, e_t) \tag{11}$$

where e_t denotes the action in period t (choice variable), π_t is the prior at t (state variable) and $P_t(e_t, \pi_t)$ is the expected (perceived) payoff at time t. Each period, given their current prior, their own action and realized output, the Hindu worker's belief about the effort level of their Muslim teammate is updated. The transition matrix at any period t, with current prior π_t, is:

Table D.1: Bayesian Updating (Hindu workers): Prob(Muslim worker exerts e_H)

Own Effort/Realized Output	O^H	O^L
e_H	$\frac{\pi_t p_H}{\pi_t p_H + (1-\pi_t) p_{HL}}$	$\frac{\pi_t (1-p_H)}{\pi_t (1-p_H) + (1-\pi_t)(1-p_{HL})}$
e_L	$\frac{\pi_t p_{HL}}{\pi_t p_{HL} + (1-\pi_t) p_L}$	$\frac{\pi_t (1-p_{HL})}{\pi_t (1-p_{HL}) + (1-\pi_t)(1-p_L)}$

Note: The prior of a Hindu worker in period t is denoted by π_t.

D.3.2 Muslim workers (minority group)

Unlike Hindu workers, Muslim workers have always been in mixed teams. They are used to being stereotyped in this manner. In other words, they are aware that Hindus are operating on incorrect priors.

Muslim workers choose an optimal effort investment path based on the time horizon. At any given time t and set of history s (which determines prior π_t of Hindu teammates), Muslim workers choose an effort level. Their problem can be written as:

$$V = E^{\psi}(\sum_{t=1}^{T} P_t(s_t, a_t) | \pi_1) \tag{12}$$

where ψ denotes the mapping from a set of histories (from 1 to $t-1$) to actions $a_t = (e_H, e_L)$ and P_t denotes expected payoff in each period conditional on the set of history and preferred action choice in that period. State s_t (history of high vs low output events, given own effort) defines the current belief π_t (π_1 is the initial belief) of the Hindu worker regarding their Muslim co-worker.

Note that for any t=k, the problem above can be re-written as

$$V_k^{\psi}(s) = \{ \sum_{s' \in S_{k+1}} P_k(s,a) + \mu_k(s'|s,a) V_{k+1}^{\psi}(s') \}, k = T-1, ..., 1 \tag{13}$$

where ψ denotes the mapping from each possible history $h_t = (s_o, a_o, ..., s_{t-1}, a_{t-1})$ to actions $a_t = \psi_t(h_t)$. $\mu(.)$ denotes the probability of a future state (belief of the Hindu worker) conditional on actions and current state. The optimal effort path for a Muslim worker is then a mapping from state histories to actions ψ^* such that,

$$\psi_k^*(s) \in argmax_{a \in e_H, e_L} \{ \sum_{s' \in S_{k+1}} P_k(s,a) + \mu_k(s'|s,a) V_{k+1}^{\psi *}(s') \} \tag{14}$$

D.3.3 Markov Equilibrium

It is clear from equation (9) that statically there are two equilibria of this game, one where both workers exert e_H and the other where both exert e_L. Since this game is repeated, it is possible to have strategies that are a function of the history of the game, as well as beliefs of Hindu workers. I am not going to rule out the possibility of some complicated equilibria based on such strategies. Instead, I will be looking at an equilibrium where individuals condition behaviour on commonly known beliefs of Hindus.

Definition: *$\bar{P}$ is the probability that a Hindu believes a Muslim is exerting e_H beyond which it is statically payoff maximizing for a Hindu to also contribute e_H.*

$\bar{P}$ is the threshold value such that if π_t is greater than this value, or in other words if Muslims are believed likely enough to be contributing high effort, Hindus will exert high effort in response. Note that $\bar{P}$ is exogenous and is obtained by comparing net expected payoff to a Hindu worker from exerting high effort versus exerting low effort, given π_t.

For a given value of π_t, the payoff from exerting e_H and e_L are as follows:

$$P(e_H, \pi_t) = \pi_t \{ p_H O_H + (1-p_H) O_L \} + (1-\pi_t) \{ p_{HL} O_H + (1-p_{HL}) O_L \} - c(e_H) \tag{15}$$

$$P(e_L, \pi_t) = p_t \{ p_{HL} O_H + (1-p_{HL}) O_L \} + (1-\pi_t) \{ p_L O_H + (1-p_L) O_L \} \tag{16}$$

Comparing (15) and (16) high effort yields high greater payoff iff (15) > (16), which is the case when

$$\pi_t > \frac{c(e_H) - (p_{HL} - p_L)(O_H - O_L)}{(p_H + p_L)(O_H - O_L)} = \bar{P} \tag{17}$$

This gives us $\bar{P}$. From equation (9) it can be seen that the numerator in the RHS is positive.

Given this, I now proceed to characterizing the equilibrium.

***Proposition**: There exists an equilibrium in which, for a given remaining interaction length T, there is a $\pi_T \leq \bar{P}$, such that for $\pi \geq \pi_T$, Muslims will exert e_H. If $\pi < \pi_T$ Muslims exert e_L. Along the equilibrium path, Hindus exert e_L when $\pi < \bar{P}$ and e_H otherwise.*

Proof: A formal proof is provided at the end of the section, I provide the intuition for the proof here. If π_t is lower than $\bar{P}$ (in a certain period t), or in other words if the belief of the Hindu worker is not high enough to exert e_H, it is a static best response for the Muslim worker to also exert e_L. However, the Muslim worker can "invest" in shifting priors of the Hindu worker, if it leads to higher payoff in expectation by transitioning to a high output static equilibrium in the future. In order for that to be worthwhile, there must be enough periods in expectation with $\pi > \bar{P}$, such that the cost of exerting e_H (while the Hindu worker exerts e_L) is compensated for and the net payoff to the Muslim worker is greater than exerting e_L (in expectation). At time t, given remaining interaction length T, π_T is the minimum (threshold) value (of π) for the Muslim worker to find e_H worthwhile.[54] If the belief of the Hindu worker at t (π_t) is below π_T , then not enough interactions are left (in expectation) for the Muslim worker's investment in shifting the beliefs of the Hindu worker to be worth it.[55] e_L is then the best response in that period.

The Hindu worker's belief at time t (π_t) is essentially the state-variable in this Markov equilibrium. The Hindu worker's action along the equilibrium path is thus e_L if in that period $\pi_t < \bar{P}$ and e_H otherwise.

D.3.4 Additional Implications

I now note a few additional implications of this model which are empirically testable.

1. *On average (during the intervention), Hindus would blame low output on Muslims, Muslims would not blame Hindus. Muslims would invest in shifting Hindu beliefs.*
   I collected data on actual interactions between workers (accusations for contributing low effort, blame etc.) which I use to test this prediction.

2. *If Hindus have had past contact with Muslims, we are more likely to see high output in*

[54] π_T is determined by the remaining length of interaction in the game and does not depend on the number of periods that have already elapsed.

[55] A couple of things are worth mentioning here. First, even though exerting e_H might not be worth it at time t, a lucky sequence of high output events can change that, such that in some future period it might be worth it. Alternatively, even though e_H might be worth it in some period, a series of bad outcomes can lead to beliefs of the Hindu worker drifting downwards whereby the Muslim worker may not find it worthwhile to exert high effort anymore in the future.

those mixed teams initially.

The idea here is that if Hindus have had enough experience of working with Muslims in the past, such that their initial belief π is greater than $\bar{P}$, Hindus and Muslims will coordinate immediately on a high output static equilibrium. I use information on pre-randomization teams of individuals as a proxy for past contact with Muslims in order to test this in the data.

3. *Closer the beliefs of Hindus to $\bar{P}$, the faster (in expectation) is convergence to high output.* This is related to the point above. The closer initial beliefs of Hindu workers are to $\bar{P}$, fewer are the number of periods with e_H required from Muslim workers, before the transition to high output static equilibrium is made. This means mixed teams in which Hindu workers have lower priors at baseline, might not see output differences between mixed and non-mixed HD teams complete dissipate during the intervention period.

D.4 Proof of Proposition

Proposition: *There exists an equilibrium in which, for a given remaining interaction length T, there is a $\pi_T \leq \bar{P}$, such that for $\pi \geq \pi_T$, Muslims will exert e_H. If $\pi < \pi_T$ Muslims exert e_L. Along the equilibrium path, Hindus exert e_L when $\pi < \bar{P}$ and e_H otherwise.*

Proof: Suppose a Hindu and a Muslim worker are working together in a team for periods t = $1, ..., T$, where T is finite but can be arbitrarily large. We assume $\pi_1 < \bar{P}$, whereby the Hindu worker exerts low effort e_L initially. At time period 1, the Muslim worker maximizes expected future payoff. High effort is optimal in the beginning for the Muslim worker iff

$$V_k^{e_H} = \{ \sum_{s' \in S_{k+1}} P_k(s, e_H) + \mu_k(s'|s, a) V_{k+1}^{e_H}(s') \} \geq T P^{e_L} \tag{18}$$

for $k = T-1, ..., 1$. Suppose at time 1, (18) is true. Then, any other investment path rather than e_H at the beginning (specifically one where the Muslim worker initially exerts e_L and then e_H) is sub-optimal. To see this, suppose that this were not the case by contradiction. Then there exists some t_1 and t_2 such that,

$$\begin{aligned} t_1 P_{a(H)=e_L}^{e_L} + t_2 P_{a(H)=e_L}^{e_H} + (T - t_1 - t_2) P_{a(H)=e_H}^{e_H} &\geq \{ \sum_{s' \in S_{k+1}} P_k(s, e_H) + \mu_k(s'|s, a) V_{k+1}^{e_H}(s') \}_{k=T-1,...,1} \\ &\geq T P^{e_L} \end{aligned} \tag{19}$$

where t_1 and t_2 respectively denote time periods during which the Muslim worker expects to put low effort and high effort respectively (while the Hindu worker still has not updated their

prior above $\bar{P}$). The notation $P^{j}_{a(H)=k}$ denotes that expected payoff to the Muslim worker in a period he exerts effort j and the Hindu worker's action is k. We can similarly split the payoff from exerting high effort and write the inequality as

$$\begin{aligned} t_1 P^{e_L}_{a(H)=e_L} + t_2 P^{e_H}_{a(H)=e_L} + (T - t_1 - t_2) P^{e_H}_{a(H)=e_H} &\geq \tilde{t_1} P^{e_H}_{a(H)=e_L} + (T - \tilde{t_1}) P^{e_H}_{a(H)=e_H} \\ &\geq T P^{e_L} \end{aligned} \tag{20}$$

where $\tilde{t_1}$ denotes the number of periods until which the the Muslim worker expects to exert e_H while the Hindu worker's $\pi < \bar{P}$. Re-writing the above we have (and ignoring the last inequality),

$$\begin{aligned} &t_1 (P^{e_L}_{a(H)=e_L} - P^{e_H}_{a(H)=e_H}) + t_2 (P^{e_H}_{a(H)=e_L} - P^{e_H}_{a(H)=e_H}) + T P^{e_H}_{a(H)=e_H} \\ &\geq \tilde{t_1} (P^{e_H}_{a(H)=e_L} - P^{e_H}_{a(H)=e_H}) + T P^{e_H}_{a(H)=e_H} \end{aligned} \tag{21}$$

Notice that for the above inequality to be true, t_2 must be larger than $\tilde{t_1}$. But this is not possible because if the Muslim worker starts off with e_L, in expectation it will take longer to shift the prior of the Hindu worker above $\bar{P}$ (so $\tilde{t_1}$ would be greater than t_2).

Next, I show that if the belief of the Hindu worker is not too low, then the Muslim worker will exert e_H. I split the payoff of the Muslim worker into two parts: before and after (expected) period j, such that for all t $\leq j$, $\pi_t \leq \bar{P}$ and $\pi_t > \bar{P}$ for $t > j$. The Muslim worker will exert e_H iff

$$t_j P^{e_H}_{a(H)=e_L} + (T - t_j) P^{e_H}_{a(H)=e_H} \geq T P^{e_L} \tag{22}$$

I show that for a large enough T such j exists. Consider the following extreme scenarios and the consequent priors of Hindu workers: (1) in each period before j, high output is produced and (2) in each period before j, low output is produced. The priors in cases (1) and (2) respectively are:

$$(1) : \frac{\pi_1 p_{HL}^{j-1}}{\pi_1 p_{HL}^{j-1} + (1 - \pi_1) p_L^{j-1}} \tag{23}$$

$$(2) : \frac{\pi_1 (1 - p_{HL})^{j-1}}{\pi_1 (1 - p_{HL})^{j-1} + (1 - \pi_1)(1 - p_L)^{j-1}} \tag{24}$$

(1) and (2) give the lower and upper bound on the beliefs of the Hindu worker about the type of the Muslim worker at time j. The prior at time period j can therefore be written as a linear combination of the expressions above. Re-writing, we therefore have,

$$\frac{\pi_1 p_{HL}^{j-1}}{\pi_1 p_{HL}^{j-1} + (1 - \pi_1) p_L^{j-1}} + A. \frac{\pi_1 (1 - p_{HL})^{j-1}}{\pi_1 (1 - p_{HL})^{j-1} + (1 - \pi_1)(1 - p_L)^{j-1}} \tag{25}$$

$$= \frac{\pi_1}{\pi_1 + (1-\pi_1)(\frac{p_L}{p_{HL}})^{j-1}} + A\frac{\pi_1}{\pi_1 + (1-\pi_1)(\frac{1-p_L}{1-p_{HL}})^{j-1}} \tag{26}$$

where A is a constant that is less than or equal to 0. In period j we therefore must have

$$\frac{\pi_1}{\pi_1 + (1-\pi_1)(\frac{p_L}{p_{HL}})^{j-1}} + A\frac{\pi_1}{\pi_1 + (1-\pi_1)(\frac{1-p_L}{1-p_{HL}})^{j-1}} > \bar{P} \tag{27}$$

Since the L.H.S. is increasing in j, while the R.H.S. is fixed, we clearly have a value of j such that inequality is satisfied. However, it cannot be so large that equation (22) is not satisfied.

In this case π_1, is the starting belief of the Hindu worker. Note that the L.H.S of equation (27) is decreasing in π_1 (since $\frac{1-p_L}{1-p_{HL}} > 1$ and $\frac{p_L}{p_{HL}} < 1$), which suggests if π_1 is too small a larger j is required (if initial prior is low, then more time periods are required). Given T fixed, the smallest value of π that allows equation (22) to be satisfied (i.e. when the equation holds with equality) is essentially the threshold value π_T such that if $\pi_1 \geq \pi_T$, then the Muslim worker exerts e_H in period 1. Note that this threshold is updated every period based on the number of interactions that remain.

A Hindu worker simply operates on their prior π_t (state variable) in each period in this Markov equilibrium. If $\pi_t > \bar{P}$, then Hindus exert e_H and e_L otherwise. Both actions are best responses given priors.

E Pre-registration differences (Online Appendix)

The study was pre-registered (without a formal pre-analysis plan) in the AEA registry with ID #0004270. Initially, a separate design was registered but the attempt to run that experiment failed because of low product demand leading to irregular production (there was no data collected). The design was then changed and the final experiment was timed during a period of relatively higher demand.

The differences between the paper and the pre-registration are the following:

1. As primary outcomes of the study, the pre-registration mentioned output data obtained directly from the firm as well as survey data (including an Implicit Association Test) measuring out group perception and prejudice. While the output data were obtained as planned, at endline only a short phone survey could be conducted due to COVID-19 related restrictions. Therefore, survey questions were restricted to those on inter-group relations at the workplace only. Furthermore, the endline IAT could not be conducted either (there was one conducted at baseline).

2. For secondary outcomes, the pre-registration mentioned survey questions on political attitudes and preferences. While these data were collected during the baseline survey and are in fact used to study heterogeneous treatment effects, the enumerators realized that the workers may not be comfortable answering these questions over the phone. Therefore, they were excluded from the endline survey.

Apart from the above mentioned differences, the experiment was implemented as explained in the pre-registration. In particular, the following main aspects were adhered to without deviation:

1. The definition of inter-group contact as contact between Hindus and Muslims (i.e. inter-religious contact), though data on caste was collected for all Hindu workers.

2. The entire design of the experiment including:

    - The definition of High- (HD) and Low-Dependency tasks with time use data
    - The design of the experiment particularly to study how the effects vary by differences in production technology (HD versus LD tasks) and not any other dimension of difference between tasks

3. Sample Size:

    - Pre-registered: **Line-level teams** - 16 to 18, **Line-section-level teams** - 100 to 110
      Number of workers (by team-type): LD Non-Mixed - 100, LD Mixed - 105 HD Non-Mixed - 190, HD Mixed - 150
    - **Actual**: **Line-level teams** - 15, **Line-section-level teams** - 13
      Number of workers (by team-type): LD Non-Mixed - 98, LD Mixed - 117 HD Non-Mixed - 196, HD Mixed - 175