

Caste and the Indian Economy ^{*}

Kaivan Munshi[†]

October 2016

Abstract

Caste plays a role at every stage of an Indian's economic life; in school, university, the labor market, and into old age. The influence of caste extends beyond private economic activity into the public sphere, where caste politics determines access to public resources. The aggregate evidence indicates that there has been substantial convergence in education, occupations, income, and access to public resources across caste groups in the decades after independence. Some of this convergence is likely due to affirmative action, but caste-based networks could also have played an equalizing role by exploiting the opportunities that became available in a globalizing economy. It is possible that caste will cease to be salient in the long-run. In the interim period, however, it is important to understand the positive and negative consequences of caste involvement across a variety of spheres in the Indian economy.

Keywords. Caste. Affirmative Action. Community networks. Migration. Mobility. Education. Business. Risk-Sharing. Public Goods.

JEL. H41. I25. J12, J62. J71. O15.

^{*}Research support from Cambridge-INET, the National Science Foundation through grants SES-0431827 and SES-0617847, and the National Institutes of Health through grant R01-HD046940 is gratefully acknowledged. I am responsible for any errors that may remain.

[†]University of Cambridge

1 Introduction

Caste plays a role at every stage of an Indian's economic life. His caste will determine the type of school he has access to, the way he is treated by his teachers, and his interactions with his classmates. In young adulthood, his caste will determine whether or not he benefits from affirmative action in higher education and (later) access to government jobs. Over the course of his working life, his caste will determine how he is assessed by potential employers, while, in parallel, networks organized around his caste help him find new jobs. Overlapping caste-based networks will provide him with credit, help him start a business if he is endowed with entrepreneurial talent, and provide him with insurance against income shocks and major contingencies into old age.

The influence of the caste extends beyond private economic activity. Electoral politics in post-independence India has had a strong caste component, with political parties targeting public resources to particular castes in return for their votes. Caste politics extends down to the local level. While it is well known that local politicians target their own caste group, new research shows, in addition, that cooperation within the caste can lead to an increased supply of non-excludable public goods.

The importance of caste in India's economy and polity should not come as a surprise, given the structure of its society. Hindu society is stratified into four hierarchical classes or *varnas*, with a large sub-population of untouchables excluded from this system entirely. Within each of these classes, and among the untouchables, are thousands of castes or *jatis*. The central rule in Hindu society is that individuals must marry within their own caste. Recent genetic evidence indicates that this rule has been strictly followed for over 2,000 years. Each caste consists of approximately 250,000 members spread over a wide area covering many villages (and, sometimes, urban locations). Spatial segregation within the village results in a high degree of local social connectedness, with caste clusters in distant villages linked to each other through ties of marriage over many generations.

The unusual social architecture described above has given rise to caste-based economic networks of exceptional size and scope. While the initial purpose of these networks, going back many centuries, was to smooth consumption in the face of income fluctuations, they later expanded the domain of their activity to the urban labor market and into business when new opportunities became available under colonial rule. Caste networks have continued to play an important role in the Indian economy after independence in 1947. This period witnessed, in parallel, the establishment of the world's most aggressive affirmative action program, whose stated objective was to eliminate the historical inequities

of the caste system. Internal cooperation within castes supports economic networks, which compete with each other in the market without regard to their position in the social hierarchy. The much more visible affirmative action program, in contrast, was explicitly targeted at historically disadvantaged castes, giving them preferred access to higher education, government jobs, and political representation. These two forces have worked in tandem to shape the evolution of the modern Indian economy.

The aggregate evidence indicates that there has been substantial convergence in education, occupations, income, and access to public resources across caste groups in the decades after independence. Some of this convergence is likely due to affirmative action, but caste-based networks could also have played an equalizing role by exploiting the opportunities that became available in a globalizing economy. The convergence between upper castes and lower castes in India contrasts with the experience in other countries, such as the United States where the racial gap has remained unchanged for many decades after Civil Rights. It is possible that the caste-based affirmative action programs will no longer be needed at some point in the future. The caste networks will also decay in time, once the market economy begins to function efficiently. In the interim period, however, it is important to understand the positive and negative consequences of caste involvement across a variety of spheres in the Indian economy.

2 Caste in Indian Society

The caste system is a system of social stratification in India that goes back as far as 1500-500 BCE. The caste system is comprised of four hierarchical classes or *varnas*, the Brahmins, Kshatriyas, Vaishyas, and Shudras. Certain population groups, known today as *Dalits*, were historically excluded from the *varna* system and were regarded as untouchables. Within each *varna* are thousands of castes or *jatis*, which are believed to have crystallized later and were originally organized as occupational guilds. A key feature of the caste system is that individuals are only permitted to marry within their *jati*.

The origins of the caste system and the initial purpose it served are controversial. The view among scholars of the British colonial era, as articulated by Dumont (1970), was that the caste hierarchy was largely symbolic, determined by differences in ritual purity maintained by upper and lower castes. More recent research contests this view, emphasizing the material exploitation that has historical been, and continues to be, associated with vertical stratification (Dirks 2001, Gupta 2000). The origins of the castes or *jatis* are similarly contentious. Gupta traces their emergence to the feudal period (7th to

12th century), whereas other scholars have claimed that caste identity and the rigid boundaries between castes that exist today were a creation of the British colonial government (De Zwart 2000, Dirks 2001).¹

It is certainly true that the first attempt to classify the Indian population by caste was made by the British colonial government, starting from the 1881 census (Bayly 2001). It is also true that the British used this classification to assign government jobs to particular groups, possibly creating a new caste identity. However, recent genetic evidence indicates that marriage within castes, the lynch-pin of the caste system, was rigidly adhered to long before the colonial period. This evidence indicates that while there was free mixing in the Indian population in ancient times, there was an abrupt change 1,900-4,200 years ago, after which strict endogamy within narrowly defined population groups became the norm (Moorjani et al. 2013). The Indian population today consists of 4,000 distinct genetic groups, each of which is a caste (or its non-Hindu equivalent kinship group) making India one of the most genetically diverse countries in the world.² The number of genetic groups that have been identified is in line with the number of castes (approximately 3,000) listed in the colonial censuses. Thus, while colonial policies may have enhanced caste affiliations, they did not create them.

With 4,000 castes and a national population of one billion, each caste comprises 250,000 members on average. Although castes are spread over a wide area, sometimes an entire state, a single caste will not have a presence in all village in the area that it covers. It will cluster, instead, in select villages, which are often located far from each other. This clustering shows up clearly in the most recent round of the Rural Economic Development Survey (REDS) conducted in 2006. The REDS is a nationally representative survey of rural Indian households that has been administered by the National Council of Applied Economic Research at multiple points in time over the past four decades. Each survey round includes a census of all villages that are covered, followed by a detailed survey of a sample of households in each village. The census data from the 2006 REDS indicate that the mean number of castes per state is 64, while the mean number of castes per village is 12. With 340 households on average in a village, this implies that a caste will have about 30 households in the select villages where it locates. This is a large enough number to support a local caste community within the village, while marriage ties link the caste across villages. Marriage in India is endogamous (within-caste) and patrilocal; ie. the bride moves into her husband's household. Most women leave their natal village when they marry.

¹Bidner and Eswaran (2012) propose an economic explanation for the emergence of the caste system that is based on the complementarity between a woman's productivity and her husband's skill. Their theory can explain many features of the caste system, including its hereditary nature, the key rule of marital endogamy, and its hierarchical structure.

²Muslims marry within *biradaris* and Christians continue to marry within their original (pre-conversion) *jatis*. Other religious communities such as the Muslim Bohris and the Jains also practice strict endogamy.

A dense web of marriage ties thus forms over many generations, allowing caste networks to span a wide area while maintaining their connectedness.

Although the population census of India has not collected information on caste affiliation since 1931, recent rounds of nationally representative surveys such as the 1999 REDS and the 2005 India Human Development Survey (IHDS) report that over 95% of Indians continue to marry within their caste (or equivalent kinship community). This is not surprising, given the genetic evidence that has recently become available, and the same patterns of strict endogamy have been reported by respondents in other sample surveys conducted in rural and urban India (Munshi and Rosenzweig 2006, Munshi 2011, Luke and Munshi 2011). Figure 1 uses retrospective information from the 1999 REDS to report rates of out-marriage; i.e. marriage between members of different castes, in rural India for the children and siblings of household heads over the 1950-1999 period. Out-marriage is just around 5% of all marriages – this number appears with remarkable consistency across all the studies listed above – and has remained stable over time.

In addition to marital endogamy, the distinctive feature of the caste system is the restrictions it places on social intercourse between castes (Ghurye 1932). These restrictions result in spatial segregation, with members of a caste typically living together in a particular area of the village; indeed, streets in Indian villages are often referred to by their caste names. Data from the 2006 REDS can be used to describe the spatial distribution of castes within villages. Focusing on the 91% of REDS villages for which information on local government (*panchayat*) wards is available, each caste makes up on average 6% of the population of a village. Within the ward, which is a smaller spatial unit, the average caste's share increases to 14%, indicative of the spatial clustering along caste lines that characterizes the Indian village.

Given the strong social ties within castes and the weak social interactions between castes, we would expect informal economic transactions in the village to be concentrated within the caste. The 2006 REDS village census asked each household head to list the names of three individuals in the village that he would approach for food, in the event of a temporary shortfall, and for a short-term loan. If these individuals were approached without regard to caste affiliation, we would expect on average that 6% of the individuals providing food transfers and loans would belong to the respondent's caste. Based on the identity of the first listed individual, the REDS data indicate that as much as 50% of food transfers and loans are from individuals belonging to the same caste. The REDS data also tell us that more than half of caste loans (by value) originate outside the village. If we included those loans, the fraction

of individuals within the caste that would be approached for a loan would increase well above 50%. Informal transactions are evidently concentrated within the caste in the Indian village.

The insider-outsider dichotomy, characterized by close intra-group relations and weak (or even adversarial) inter-group relations, is a feature of all traditional societies.³ This dichotomy is especially pronounced in Indian society due to its unique caste-based structure, as discussed above, and is also evident from cross-country comparisons based on the World Values Survey (WVS). Figure 2 reports cross-country results from Wave 5 of the WVS (conducted between 2005 and 2009 in different countries) on questions relating to trust, and tolerance of outsiders. Restricting the sample to countries with a population in excess of 20 million that are classified by the World Bank as low, lower-middle, or upper-middle income (with the addition of Russia), we see in Figure 2A that India ranks close to the top of the list with regard to trust in neighbors. Given the spatial segregation that characterizes the Indian village, the Indian respondents are essentially reporting that they have a high degree of trust in their fellow caste members living nearby. However, the ranking reverses when we look at tolerance of outsiders – speaking a different language or following a different religion – in Figures 2B and 2C, respectively. India now lies at the very bottom of the list and, based on the preceding discussion, we would expect Indians to be similarly intolerant of neighbors belonging to a different caste (this is certainly reflected in how they live).

The sections that follow will examine the role played by caste communities in a wide variety of economic activities. The insider-outsider dichotomy will be a common thread running through the discussion in each section. While internal cooperation allows caste communities to support the economic activity of their members when market institutions are missing or function imperfectly, the disregard (and even active hostility) towards outsiders, creates new distortions in the economy. There is no perfect substitute for a well functioning market, and we will see that while informal caste-based arrangements may increase economic efficiency in the short-run, they can restrict the occupational and spatial mobility of their members in the long-run.

³Akerlof (1976) and Choy (2016) provide game theoretic foundations for why internal cooperation coupled with the exclusion of outsiders can be sustained in equilibrium.

3 Mutual Insurance

3.1 Caste-Based Insurance

Consider an economy with two risk averse individuals, A and B. Each individual earns either 1000 or 2000 per period, with equal probability. Income realizations are independent across individuals, resulting in the following income states: (1000, 1000), (2000, 2000), (1000, 2000), (2000, 1000), where the first term in parentheses refers to individual A's income and the second term refers to individual B's income and each income state occurs with probability 0.25.

Given that individuals are risk averse, they will benefit from a mutual insurance arrangement that smooths their consumption over time. Individuals who receive a positive income shock make transfers to temporarily less fortunate individuals in these arrangements, with the roles reversing in later periods when the income shocks are reversed, leaving everyone better off. An alternative (equivalent) way to characterize mutual insurance arrangements when preferences are logarithmic is that the total income of all participants is pooled in each state or period and then distributed on the basis of a pre-specified income sharing rule.⁴

$$c_i^s = \frac{\lambda_i}{\sum_i \lambda_i} Y^s, \quad (1)$$

where c_i^s is the consumption of individual i in income-state s , Y^s is total income (over all participants in the insurance arrangement) in state s , and $\lambda_i / \sum_i \lambda_i$ is individual i 's share of the pot, which is constant across all states. In our simple example, the income generating process is the same for the two individuals and so it is reasonable to assume an equal income-sharing rule: $\lambda_A = \lambda_B = 1/2$. The usual assumption in models of mutual insurance is that there is no storage or savings, which results in three consumption-states: (1000, 1000) with probability 0.25, (2000, 2000) with probability 0.25, and (1500, 1500) with probability 0.5.

The preceding discussion serves as the basis for a test of full risk-sharing. Taking logs in equation (1),

$$\log(c_i^s) = \log(Y^s) + \log\left(\frac{\lambda_i}{\sum_i \lambda_i}\right). \quad (2)$$

The preceding equation can be used to derive two testable predictions.

⁴More generally, with full risk-sharing, the ratio of marginal utilities between any two members of the arrangement is constant in all income states. With log preferences, this implies that the ratio of consumptions are constant. Summing over all network partners, this implies that each individual receives a constant fraction of the total income that is generated in each state.

(1) Replacing states, s , with time periods, t , individual consumption will co-move with total income (or total consumption) in the group around which mutual insurance is organized.

(2) Conditional on total income, Y^s , notice that individual income, y_i^s , has no bearing on consumption, c_i^s . This is evident from our simple example; consumption is equal to 1500 in states with unequal incomes, regardless of whether the individual earns 1000 or 2000.

In our example, income-shares are equal. More generally, individuals earning higher incomes on average will have higher income shares. This is accounted for in the test of full risk-sharing by including individual fixed effects, which capture the last term in equation (2). Conditional on group income in a given period, shocks to individual income; i.e. deviations from average income, should have no effect on individual consumption with full risk-sharing.

Starting with Townsend's (1994) seminal contribution set in village India, studies from across the developing world have documented extremely high levels of risk sharing (close to full risk sharing) in rural communities; e.g. Grimard (1997), Fafchamps and Lund (2003), Angelucci, Di Giorgi, and Rasul (2015). These communities are evidently able to smooth income-risk extremely effectively, despite the fact that private credit and government safety nets are largely absent in the economies within which they operate. While this research has shed new light on the important role played by informal non-market institutions in smoothing income-risk in developing economies, it suffers from two limitations.

The first limitation is that the social group within which risk sharing is assumed to take place is often chosen on the basis of convenience. For example, Townsend treats the village as the relevant social group, whereas our characterization of Indian society would suggest that the caste (within and across villages) is the social group around which mutual insurance arrangements are organized in rural India. The test of full risk sharing can be implemented with data from any sub-set of individuals belonging to the relevant social group. However, it will erroneously reject full risk sharing if it includes individuals outside the group, such as individuals outside the caste residing in the same village.

The available evidence supports the view that the caste, rather than the village, is the relevant group around which mutual insurance arrangements are organized in rural India. Caldwell, Reddy, and Caldwell (1986) surveyed nine villages in South India after a two-year drought and found that nearly half (46 percent) of the sampled households had taken consumption loans during the drought. The sources of these loans (by value) were government banks (18 percent), moneylenders, landlord, employer (28 percent), and relatives and members of the same caste community (54 percent), emphasizing the importance of caste loans for smoothing consumption. The REDS, which provides detailed information

on gifts and loans for a sample of villages covering all the major Indian states can be used to corroborate these findings. Focusing on the 1982 REDS round, which included a detailed classification of the source and purpose for each loan, the key findings are (i) that caste loans are second in overall value after bank loans, and (ii) that caste loans are the primary source of support for meeting major contingencies such as illness and marriage. Gifts make up the bulk of transfers within the caste (67% by value). If we included gifts, the caste would be by far the most important source of support for smoothing consumption and meeting major contingencies in rural India. The same pattern holds up in later REDS rounds, such as the 1999 round, and in other nationally representative surveys such as the 2005 IHDS. Providing the most direct evidence that the caste rather than the village is the relevant risk-sharing unit, Mazzocco and Saini (2012) use ICRISAT data to establish full risk-sharing at the level of the caste but not the village.

The second limitation of the standard test of full risk-sharing is that it does not account for other important aspects of consumption smoothing. Returning to our simple example, suppose now that there are N individuals in the mutual insurance arrangement. As before, income realizations are independent across participants and each individual earns 1000 with probability 0.5 and 2000 with probability 0.5. The income generating process is the same for all individuals, so we continue to assume an equal income-sharing rule.

It is straightforward to show that mean consumption with insurance, M_I , is equal to mean income or, equivalently, mean consumption under autarky, M_A , which is 1500 in our example:

$$M_I = E \left(\frac{1}{N} \sum_i y_i^s \right) = \frac{1}{N} (NM_A) = M_A. \quad (3)$$

Although mean consumption is independent of the number of participants, the variance of consumption with insurance, V_I , is decreasing in N :

$$V_I = V \left(\frac{1}{N} \sum_i y_i^s \right) = \frac{1}{N^2} (NV_A) = \frac{V_A}{N}. \quad (4)$$

Extremely high levels of risk sharing have been documented throughout the developing world. What distinguishes caste-based insurance arrangements from their counterparts elsewhere is their size and scope. Large N reduces the variance of consumption, and the wide spatial scope of caste networks diversifies income shocks, reducing the variance of consumption even further. This feature of consumption smoothing is not captured by the test of full risk sharing and may explain, in part, the persistence of the caste system in India. Without adequate access to private credit and government safety nets, the

exceptionally well-functioning caste-based insurance network provides a vital and irreplaceable service (in addition to other services that will be discussed in the sections that follow).⁵ This may be why individuals continue to marry within their caste, to the extent that such marriages are a pre-requisite for access to the caste network, and why a social system that was most likely set up to serve other objectives continues to play such an important role in Indian economic life.

3.2 Caste-Based Insurance and Restrictions on Mobility

The preceding view of caste-based insurance is that it has emerged, and persisted, because market credit and government safety nets are inadequate. While this view would suggest that caste-based insurance networks are efficiency enhancing, Munshi and Rosenzweig (2016) show that they can inadvertently generate distortions in other markets; specifically, the labor market. Their research is motivated by two stylized facts: (i) The rural-urban wage gap in India is exceptionally high relative to other large developing countries and has remained this way for many decades, and (ii) the rate of permanent male migration has, nevertheless, remained exceptionally low over this period. This implies that there must be some friction that prevents rural workers from moving to the city and taking advantage of what appears to be an arbitrage opportunity.

Munshi and Rosenzweig posit that the source of this friction is the loss in (rural) network insurance when one or more members of a household migrates permanently to the city. Households with migrant members will have reduced access to rural caste networks for two reasons. First, migrants cannot be as easily punished by the network, and their family back home in the village now has superior outside options (in the event that the household is excluded from the network). It follows that households with migrants cannot credibly commit to honoring their future obligations at the same level as households without migrants. Second, an information problem arises if the migrant's income cannot be observed. If the household is treated as a collective unit by the network, it always has an incentive to misreport its urban income so that transfers flow in its direction. If the resulting loss in network insurance from migration exceeds the income gain, then large wage gaps could persist without generating a flow of workers to higher-wage areas. This distortion is paradoxically amplified when the informal insurance

⁵Data from the 1982 REDS indicate that loan terms are substantially more favorable for caste loans on average. Caste loans are significantly more likely to be interest-free than loans from banks, employers, and moneylenders. They are also significantly more likely to be collateral-free than loans from banks. Among the caste loans received in the year prior to the 1982 survey, 20 percent by value required no interest payment and no collateral. The corresponding statistic for the alternative sources of credit was close to zero, except for loans from friends where 4 percent of the loans were received on similarly favorable terms. Similar results are obtained with the 2005 IHDS.

networks work exceptionally well because rural households then have more to lose by sending their members to the city.

The simplest test of the hypothesis that the potential loss in network services restricts mobility in India would be to compare migration rates in populations with and without caste-based insurance. This exercise is infeasible, given the pervasiveness of caste networks. What Munshi and Rosenzweig do instead is to look within the caste and theoretically identify which households benefit less (more) from caste-based insurance. They then proceed to test whether it is precisely those households that are more (less) likely to have migrant members.

The literature on mutual insurance is concerned with *ex post* risk sharing, taking the size of the network and the sharing rule as given. To derive the connection between insurance networks and migration, however, it is necessary to take a step back and model the *ex ante* participation decision and the optimal design of the income sharing rule. Under the assumption that there is full risk-sharing, households can either remain in the village and participate in the insurance network or send one or more of their members to the city, increasing their income but losing the services of the network. The sharing rule that is chosen in equilibrium determines which households benefit more from the insurance network and, therefore, choose to stay.

To derive the sharing rule, and associated migration, we begin with the household's decision rule. With log preferences, the expected utility from consumption can be derived as an additively separable function of mean consumption and its coefficient of variation (squared). With this specification, a household will choose to remain in the rural insurance network and forego its urban opportunities if the following condition is satisfied:

$$\log(M_I) - \frac{1}{2} \frac{V_I}{M_I^2} \geq \log(M_A) - \frac{1}{2} \beta \frac{V_A}{M_A^2} + \epsilon. \quad (5)$$

As defined above, M_I , V_I refer to the mean and variance of consumption with insurance. M_A , V_A are the mean and variance of the household's income; i.e. its consumption in autarky, when all its members remain in the village. This income would increase to $M_A(1 + \tilde{\epsilon})$ if one or more household members moved to the city, where $\epsilon \equiv \log(1 + \tilde{\epsilon})$ measures the household's urban opportunity (in utility units), which is unobserved by the rest of the rural network.

While migration will increase the household's income, this increase must be weighed against the additional consumption risk that a household with migrants must face. The risk parameter β reflects both the decline in income risk through the diversification in household income that accompanies

migration as well as the availability of non-network insurance. We thus expect that $\beta < 1$. Nevertheless, the key assumption of the model is that $\frac{V_I}{M_I^2} < \beta \frac{V_A}{M_A^2}$. This is the wedge that permits large wage-gaps to exist in equilibrium. Retaining the assumption that the rural income generating process is the same for all households, which implies an equal income-sharing rule and, hence, from equation (3) that $M_I = M_A$, a household will only send migrants to the city if $\epsilon > \beta \frac{V_A}{M_A^2} - \frac{V_I}{M_I^2}$. Since we have fixed the income-sharing rule, for the time being, we only need to solve for equilibrium participation; i.e. the threshold ϵ_I above which households will send migrants to the city.

Once we introduce heterogeneity in rural incomes, the income-sharing rule and migration in each income class must be simultaneously determined. Munshi and Rosenzweig assume that the community's objective is to maximize the total surplus generated by the insurance arrangement (placing equal weight on all income classes) taking account of participation in each income class. With diminishing marginal utility, the total surplus generated by the insurance arrangement can be increased by redistributing income so that relatively poor households consume more than they earn on average; i.e. $M_I > M_A$. This gain from redistribution must be weighed against the cost to the members of the network from the accompanying decline in its size, since relatively wealthy households will now be more likely to leave and smaller networks are less able to smooth consumption. The first prediction of the model is that the income sharing rule will nevertheless be set so that there is some amount of redistribution in equilibrium. This implies that relatively wealthy households within their caste benefit less from the network and so will be more likely to have migrant members *ceteris paribus*.

The second prediction of the model is that households facing greater rural income risk will be less likely to have migrant members. Notice from inequality (5) that all households in the insurance network face the same consumption risk, regardless of their income risk. This implies that households facing greater income risk benefit more from the network and so will be more likely to stay. This prediction is especially useful in distinguishing Munshi and Rosenzweig's hypothesis, based on the rural insurance network, from alternative explanations for the large wage-gaps and low migration in India. One alternative explanation is that a small number of castes fortuitously managed to establish networks in the urban labor market, extracting wage rents for their members and excluding outsiders. There are also alternative explanations available for redistribution within the caste and the increased exit from the network by relatively wealthy households. However, none of these explanations imply that households facing greater rural income-risk should be less likely to have migrant members.

Munshi and Rosenzweig begin their assessment of the theory by showing that there is substantial

redistribution of income within castes, using data from the most recent ICRISAT panel survey and from the 2006 REDS. We know very little about the distribution of resources within castes, and this new finding will be corroborated below with data from urban India. Munshi and Rosenzweig then show, using the 2006 REDS village census, that relatively wealthy households within their caste are significantly more likely to report that one or more adult male members have permanently left the village. Highlighting the role that rural income-risk plays in the migration decision, they also find that households with a higher coefficient of variation in their (rural) income – who benefit more from the rural insurance network – are less likely to have migrant members. Additional results, using the 1982 and 1999 REDS rounds, directly support the key assumption of the model, which is that migration should be associated with a loss in network services (and increased marriage out of the caste).

Having found evidence consistent with the theory, Munshi and Rosenzweig proceed to estimate the structural parameters of the model. Migration and the income-sharing rule (the λ 's in equation (1)) are determined jointly in the model. With logarithmic preferences, the λ 's map directly into relative consumption across income classes. The estimated λ 's match closely with relative consumption, which is available from the REDS but not used to estimate the model. This close match effectively validates the model, providing empirical support for the assumptions that preferences are logarithmic and that the community's objective is to maximize the surplus generated by the insurance arrangement (placing equal weight on all income classes). The substantial redistribution that is uncovered appears to be driven entirely by an attempt to equate marginal utilities across income classes. Counter-factual simulations that quantify the effect of formal insurance on migration, leaving the rural insurance network in place, indicate that a 50 percent improvement in risk-sharing for households with migrant members (which is still some way from full risk-sharing) would more than double the migration rate, from 4.3% to 9%. In contrast, doubling the rural-urban wage gap, from 18% to 36%, without any change in formal insurance, increases the migration rate by just 1.5%. These results emphasize the central message of this research, which is that exceptionally well-functioning (caste-based) rural insurance networks coupled with inadequate access to formal insurance, rather than wage differentials as commonly assumed in models of migration, may explain much of the low mobility in India.

One way to circumvent these restrictions on mobility would be for members of the rural community to move to the city as a group. Members of the group could monitor each other and enforce collective punishments, solving the information and commitment problems described above. They would also help each other find jobs at the destination. As described in the sections that follow, the history of

industrialization and urbanization in India is indeed characterized by the formation and the evolution of caste-based urban networks, sometimes over multiple generations. We will see that a limitation of this strategy is that a sufficiently large (common) shock is needed to jump-start the new network at the destination, and such opportunities occur relatively infrequently. Thus, while members of a relatively small number of castes with (fortuitously) well established destination networks can move with ease, most potential migrants will lack the social support they need to move. A second strategy to reduce the information and enforcement problems that restrict mobility is to migrate temporarily. The principal limitation of the temporary migration strategy is that it will not fill the large number of jobs in which there is firm-specific or task-specific learning and where firms will set permanent wage contracts. Nevertheless, seasonal migration has been increasing rapidly over time in India (Morten 2015). The anecdotal evidence; e.g. Dhillon, Iverson, and Torsvik (2013), suggests that caste networks are once again playing an important role in this migration and so this would appear to be an important and open area for future research.

4 Business

4.1 Historical Background

Business success in a developing economy requires connections, know-how, and access to capital. This effectively shuts out most potential entrepreneurs unless they have the good fortune to be born into a business family. The obvious question to ask then is how first-generation entrepreneurs enter business in the first place, particularly in a country like India. In Weber's (1958) view, the stability of the caste hierarchy served as a natural barrier to occupational mobility and entrepreneurial activity. This view has been contested by historians such as Rudner (1994) and Chandavarkar (1994) who take the position that mobility is possible, even in a caste-based society, but that this mobility involves the movement of the entire group. We will see below that caste networks have historically supported, and continue to support, the movement of first-generation entrepreneurs into business. Our objective will be to examine the conditions under which such transitions take place.

One *varna*, the Vaishyas, traditionally controlled money-lending and trade in India, with castes drawn from this *varna* active in different regions of the country. Non-Hindu communities such as the Jains, Bohris, and Khojas were also involved in trade from ancient times. Some of these castes (communities) had migrated overseas, to East Africa and the Persian Gulf, by the seventeenth century

(Mehta 2001). However, mercantile opportunities were relatively limited until the arrival of the British in the eighteenth century. Commercial activity under the British was concentrated around the ports of Bombay and Calcutta.⁶ Parsis, Gujarati Baniyas, Bhatias, and Bohris dominated long-distance trade out of Bombay from the early nineteenth century (Gupta 2014) whereas the Marwaris were only able to establish a foothold in the Calcutta trade (against stiff competition from the British) towards the end of that century (Timberg 1978). Over time, these communities moved from trade to manufacturing; as early as the middle of the nineteenth century in Bombay (Nafziger 1971) and from 1914 onwards in Calcutta (Lamb 1955).

Apart from providing new opportunities in trade and manufacturing, British colonial rule also encouraged the settlement of entire communities in overseas destinations. The Nakarattar Chettiars, for example, served as moneylenders for the British throughout South East Asia, directly providing investment capital to farmers, plantation owners, and mine operators from 1870 to 1930 (Rudner 1994). Starting from the late nineteenth century, the *patidars* (Patels) of Charotar in Gujarat moved to British colonies in East Africa where they set up business and, in time, dominated commercial activity (Mehta 2001). While the Chettiars were forced to return to India in 1930, the Patels remained in East Africa till the 1970s, when they too were forcibly evicted (many of them subsequently moved to the U.K. where they now run successful businesses).

Although some of the communities listed above, such as the Marwaris and Chettiars, were historically engaged in business-related activities, colonial rule provided the first opportunity for other communities to move into business. These opportunities for group-level mobility have continued after independence in 1947, particularly after the liberalization of the Indian economy in the 1990s, with the emergence of a new business class drawn from a variety of agricultural castes and from castes that were historically concentrated in the bureaucracy and white-collar professions (Damodaran 2008). The striking feature of Indian business history is that the narrative is entirely about castes, or their non-Hindu equivalent kinship communities; as Gadgil (1959: 16) puts it, “the history of the rise and growth of a modern business class in India is largely the history of the activities of members of certain groups.”

Although occupational mobility did occur on occasion, the absolute number of communities that moved into business over time is small. Timberg (1978) reports that 23 of the 37 largest North Indian

⁶These cities were recently renamed as Mumbai and Kolkata, respectively. However, we will retain the historical names throughout to maintain consistency.

owned industrial houses listed in the *Monopolies Inquiry Commission Report* of 1964 were Marwari or Gujarati Bania. More recently, a Times of India article (October 20, 2006) estimates that Gujarati promoted companies account for 17% of the market capitalization of the Bombay Stock Exchange BSE-500 index, followed by Marwari promoted companies with 11% and Parsi promoted companies with 8%. Public sector units, including banks and oil companies, account for 25% and all other companies, including multinationals, just 39% of the market capitalization.⁷ Firms that are listed on the stock exchange, particularly those that are successful enough to be listed on the BSE-500, are operating outside the boundaries of their community network, almost by definition. What these statistics tell us is that the pool from which these successful firms are drawn is not particularly socially diverse. Once a community-based business network has formed, it will strengthen rapidly over time, bringing in hundreds of new entrepreneurs. Those entrepreneurs will support their children, which implies that the community network is less essential from the next generation. The critical challenge appears to be to jump-start the community network. The model laid out below, which is a simplified version of Munshi (2011), characterizes conditions under which business networks will form and grow over time.

4.2 Caste-Based Business Networks

Consider an economy with a continuum of workers of unit mass entering the labor force in each period. There are two occupations in this economy: a traditional occupation and business. The payoff from the traditional occupation is V for all individuals. The payoff from business for individual i who enters the labor force in period t is determined by his ability, ω_i , and the size of his business network, which we measure by the stock of community members who are already in business; $\sum_{\tau=0}^{t-1} \Delta\omega_\tau$, where $\Delta\omega_\tau$ is the flow of individuals into business in period τ . Individual i will then select the business occupation if the following condition is satisfied,

$$\omega_i + \beta \sum_{\tau=0}^{t-1} \Delta\omega_\tau \geq V. \quad (6)$$

As in Munshi and Rosenzweig's (2016) model of insurance and migration, the choice is between the location or sector where the network is active and going it alone. In their model, the insurance network at the rural origin discourages individuals from moving independently to the city. In the current model, the business network at the destination encourages mobility; in each period t , there is a threshold ability, $\underline{\omega}_t$, above which individuals select into the business occupation, with the threshold declining over time

⁷The market capitalization of the BSE-500 index is 92% of the total Bombay Stock Exchange (BSE) market capitalization.

as the network grows in size. Note, however, that there is no inconsistency between the two models; the general point is that community networks support group mobility but discourage the (independent) movement of individual members, and this will show up again in the subsequent section. Where the two models do differ is in the specification of preferences – we assume from now on that individuals are risk-neutral – and that network effects occur with a lag (because individuals need to have some business experience before they can support others).

To initiate the network dynamics, $\Delta\omega_0$ individuals are moved exogenously into the business occupation in period 0. Assume that ability is distributed uniformly on the unit interval; $\omega \sim U[0, 1]$. The flow of individuals into business in any subsequent period is then $\Delta\omega_t \equiv 1 - \underline{\omega}_t$. This allows us to derive the relationship between $\Delta\omega_t$ and the history of flows from equation (6),

$$\Delta\omega_t = \beta \sum_{\tau=0}^{t-1} \Delta\omega_\tau - \alpha, \quad (7)$$

where $\alpha \equiv V - 1$. The assumption here is that $V > 1$, which ensures that no individual would choose the business occupation without the exogenous push in period 0. Equation (7) can be solved recursively to derive a closed-form solution for the flow of individuals into business in each period,

$$\Delta\omega_t = (\beta\Delta\omega_0 - \alpha)(1 + \beta)^{t-1}. \quad (8)$$

The first implication of the preceding equation is that the business network will only grow over time if $\beta\Delta\omega_0 - \alpha > 0$; i.e. if the initial push, $\Delta\omega_0$, is sufficiently large. This may explain why a relatively small number of caste-communities have successfully managed the transition into business historically; it is not that they lack business potential, but that they were not fortunate enough to receive a large enough common shock that would have initiated subsequent group mobility.

Up to this point, the model has implicitly assumed that there is a single community (network) in the economy. In practice, multiple communities will co-exist, distinguished by their traditional occupations, which determine V , and their social connectedness, which determines β , the mapping from network size to business payoffs. The second implication of equation (8) is that once it does take off, the network will grow relatively fast in socially connected communities (with large β) and in communities with worse outside options (with small α and, hence, small V). This last result is especially interesting because it tells us that policies that attempt to jump-start business networks could be both growth enhancing and equalizing if they target historically disadvantaged communities.

Munshi (2011) tests the implications of the dynamic model with data from the Indian diamond

industry. This industry occupies an important position in the Indian economy, accounting for roughly 14% of India's total merchandise exports, and has competed with textiles, and more recently with computer software, as the country's top export industry over the past decades. For the purpose of our analysis, the most interesting feature of the diamond industry is its social heterogeneity. India does not produce rough diamonds and so the diamonds must be imported, for the most part from the Antwerp market, and then cut and polished in domestic factories before being sold in Bombay to foreign buyers or directly abroad. Two traditional business communities – the Marwaris and the Palanpuri Jains – controlled the business end of the industry from its inception in the mid-1960s, leaving the cutting and polishing to the Kanbi Patels, a community of lower caste agricultural laborers known informally in the industry as the Kathiawaris (Engelshoven 2002). The industry structure changed in the late 1970s with the discovery of massive diamond deposits (one-third of the world diamond supply by volume) in Australia's Argyle mines. This supply shock allowed Kathiawari firms to enter business, and today all three communities account for a substantial share of the industry. Our interest is in characterizing the precise conditions under which a sufficiently large number of Kathiawaris were able to enter initially and then to compare the subsequent trajectory of the three community networks.

Because of the difficulty in enforcing legal contracts, the diamond industry is associated with a high degree of community networking throughout the world (Coleman 1988, Richman 2006). Over 96% of Indian export-firms are drawn from just three caste-communities, indicating that networks must be active in the Indian industry as well. In the Indian industry, the role of the network appears to be to facilitate supplier credit. Most exporters visit the Antwerp market for a few days, once a month or every couple of months, to source rough diamonds. They tend to specialize in stones of a particular size, and while each exporter will have a small number of regular suppliers, he will also want to buy stones from other suppliers from one trip to the next. Given the high value of the diamonds, most Indian exporters must rely on supplier credit. What the network does is to allow its members to receive credit from suppliers with whom they do not have long-term connections; other members of the community who do have established relations with those suppliers stand guarantor for them, with the threat of social sanctions ensuring that the beneficiaries of the referrals do not renege on their commitment.⁸ An alternative strategy to source rough diamonds, typically followed by established (second-generation) entrepreneurs is to set up a branch in Antwerp. This allows the firm to build a reputation in the

⁸This cooperative arrangement has many features in common with Greif's (1993) description of the Maghribi traders' coalition a thousand years earlier.

Antwerp market and operate independently of its community network, both as an exporter and as a rough diamond supplier.

Retrospective information from a survey of 800 Bombay based diamond export firms conducted in 2004-2005 indicates that there was a sharp increase in the number of Palanpuri firms with branches in Antwerp just around the time that diamonds were discovered in the Argyle mines. The narrative in the industry is that these Palanpuris unloaded some of the increased supply of rough diamonds that came on the Antwerp market by encouraging their Kathiawari labor contractors, with whom they had established long-term bilateral relations, to enter business. A sufficiently large number of Kathiawaris must have entered at that time because hundreds of Kathiawaris entered business in the subsequent decades (presumably supported by an underlying network that was growing in strength over time). This discussion highlights the difficulty in jump-starting community networks; a large supply shock was needed, together with support from outsiders (who were no longer reliant on their community network). Such a confluence of favorable circumstances is unlikely to occur very often and if this is what is needed to jump-start a network, then it would explain why a relatively small number of castes are active in business today.

Once the dynamics have been initiated, the model tells us that networks will strengthen most rapidly in socially connected communities and in historically disadvantaged communities. Given the circumstances, we would expect the Kathiawari network to have strengthened relatively rapidly, with an accompanying decline in the ability of entrants from that community over time. We motivated a role for community-based business networks with the argument that these networks substitute for a family background in business. This implies that the family background of the Kathiawari entrants should be declining particularly steeply over time, and this is precisely what we observe with data from the survey (which includes the establishment year of the firm and the family background of the entrepreneur).⁹ Education, which is an alternative (and standard) measure of ability, also declines relatively steeply over time for the entering Kathiawaris, after accounting for age-cohort effects.

Our interpretation of the weakening ability of the entering entrepreneurs over time, especially the Kathiawari entrepreneurs, is that they are being supported by a strengthening network. Administrative data on firm-level exports over a ten-year period (1995-2004) allow us to validate this interpretation.

⁹While Marwari and Palanpuri entrants are slightly less likely to report that their father was a businessman over time, this decline is particularly steep for the Kathiawaris; 70% of the entrants in 1970 (before the supply shock and the establishment of the network) have a family background in business, but this statistic declines rapidly and drops below 20% for the entrants in 2000.

The Kathiawaris keep pace with their Marwari and Palanpuri rivals, despite the observed decline in their ability. Indeed, once this compositional change is accounted for with firm fixed effects, the Kathiawari export trajectory is significantly steeper than the corresponding trajectories for the other two communities (or its own trajectory without fixed effects). There is evidently an underlying community-level force that is giving a boost to the Kathiawari firms, resulting in a dramatic shift from farming to business over the course of a single generation.¹⁰ Similarly transitions have no doubt occurred before, but the diamond industry is not representative and so we know relatively little about the precise circumstances under which other transitions occurred. Given the importance of entrepreneurship in the development process and the potential role played by caste networks in stimulating entrepreneurship in India, further investigation along these lines could add much value. This could involve additional industry-level or caste-level case studies, or the use of comprehensive census or survey data to examine the evolution of private enterprise in its entirety (subject to the constraint that caste identifiers are available at the firm level).

4.3 Caste Networks and Static Inefficiency

Once a business network has formed, we have seen that it can transform the circumstances of an entire community in a single generation. From the perspective of the insiders, these informal institutions are entirely positive; they generate surplus for the community by efficiently allocating resources internally, and they are redistributive, with less able individuals benefiting from the support provided by more capable caste-members. However, we must also consider the consequences of these networks for outsiders (belonging to a different network or without a network of their own). One obvious inefficiency that arises in an economy where business is community-based is that outsiders are crowded out; capable individuals do not receive the appropriate level of resources or are excluded from business active entirely. The examples that follow document such inefficiencies in urban and rural India, respectively.

Banerjee and Munshi (2004) examine resource allocation within and across caste-communities in Tirupur's garment export cluster in a paper that pre-dates the misallocation literature in macroeconomics. Tirupur, a town in the South Indian state of Tamil Nadu, supplies 70% of India's knitted garment exports. The textile industry in Tirupur was initially controlled by a local trading community. However, after a prolonged period of labor unrest in the 1960's, it was taken over by the Vellala

¹⁰Closer inspection of the entrepreneurs' family background indicates that most of the inter-generational mobility was driven by a shift out of agriculture into business; 52% of the Kathiawaris report that their father was a farmer (and this figure is much higher for recent entrants).

Gounders, a wealthy farming community (Swaminathan and Jeyaranjan 1994). Details of the circumstances under which the Gounders entered industry are unavailable, but as with the diamond industry, there appears to have been a fortuitous element to the occupational transition.

Tirupur continued to be dominated by the Gounders for the next 20 years, with the industry producing exclusively for the domestic market. Starting from the mid-1980s, however, the export of knitted garments from Tirupur started to grow extremely rapidly (presumably as a consequence of another fortuitous shock) and by the early 1990s, the annual growth rate of exports exceeded 50%. This generated an influx of entrepreneurs from other regions of the country. In 1996, when Banerjee and Munshi conducted a survey of firms in the industry, about half the entrepreneurs were Gounders and the rest belonged to traditional business communities, including Gujarati Banias, Parsis, and Marwaris.

The survey collected retrospective information on investment and production over a four-year period for each firm. Banerjee and Munshi's objective is not to model the entry of firms over time, as we did with the diamond industry, but instead to identify the mismatch between entrepreneurial talent (ability) and capital that arises because credit fails to cross community lines. Their analysis is motivated by two stylized facts: First, the Gounders hold more capital stock than the Outsiders at all levels of experience. Adjusting for differences in production, the Gounders use roughly twice as much capital per unit of production than the Outsiders. Second, exports grow faster for the Outsiders than for the Gounders at all levels of experience. This is not mechanical convergence; the Outsiders do start with lower export levels, but they outstrip the Gounders after five years of experience.

These two stylized facts, taken together, can be shown to imply that there must be a misallocation of resources across communities in this industry. To see this, let the firm's exports, Y , be determined by entrepreneurial ability, ω , and capital stock, K . We make the assumption that the two inputs are complements,

$$Y = \omega K^\alpha, \tag{9}$$

where $\alpha < 1$ ensures that the production function has the standard neoclassical properties. Let r be the interest rate faced by the firm. The firm will choose K to maximize its profit, which yields the following expressions,

$$K = \frac{\omega^2}{4r^2} \quad Y = \frac{\omega^2}{2r}. \tag{10}$$

If all firms faced the same interest rate, as they would if markets functioned efficiently, then entrepreneurs with higher ability would hold more capital stock *and* have higher exports. This is evidently

not the case in Tirupur. Indeed, it can be shown that the two stylized facts necessarily imply that the Gounders face lower interest rates, $r_G < r_O$, and have lower ability, $\omega_G < \omega_O$. To see this, substitute the expressions for K and Y from equation (10),

$$K_G > K_O \Rightarrow \frac{\omega_G^2}{\omega_O^2} > \frac{r_G^2}{r_O^2} \quad (11)$$

$$Y_G < Y_O \Rightarrow \frac{\omega_G^2}{\omega_O^2} < \frac{r_G}{r_O}. \quad (12)$$

$\frac{r_G}{r_O} > \left(\frac{r_G}{r_O}\right)^2$ from inequalities (11) and (12), which implies, in turn, that $\frac{r_G}{r_O} < 1$. It follows, from inequality (12), that $\frac{\omega_G}{\omega_O} < 1$.

The Gounders are a wealthy landowning community and garment manufacturing was their first activity outside agriculture. Given that they have few alternative uses for their capital, it makes sense that (internal) capital is cheaper for them than for the Outsiders. Although we do not model selection into the industry, it also makes sense that they have lower ability on average. The inefficiency that Banerjee and Munshi uncover is that cheap capital fails to cross community boundaries, from the Gounders to the more productive Outsiders, presumably because commitment problems preclude such trade.¹¹

The inability to trade across castes is not restricted to industrial activity. Anderson (2011) provides evidence that caste is an impediment to trade in groundwater, a critical agricultural input, in North Indian villages, despite the fact that castes have been co-existing in these villages for centuries. Anderson's analysis exploits the fact that there are two types of villages in her sample: an upper caste is dominant; i.e. it owns the majority of the land, in 48% of Hindu villages, whereas a Backward Agricultural Caste (BAC) is dominant in the remaining villages. Upper castes are almost entirely absent from the BAC-dominated villages. A comparison of the two types of villages yields the key empirical finding, which is that lower castes – BAC, Other Backward Castes (OBC), and Scheduled Castes (SC) – have substantially higher incomes and crop yields in BAC-dominated villages. Anderson's explanation for this finding is that low caste water buyers in the BAC-dominated villages have better access to this input because they can trade more easily with water sellers of their own caste.

¹¹One alternative explanation for the two stylized facts is that ability and capital are substitutes, in which case higher ability Outsiders could grow faster even while holding less capital than the Gounders (with interest rates the same in both communities). But if this is true, then exports and capital stock should be negatively correlated *within* each community. However, within each community, where we expect entrepreneurs to face the same interest rates, firms holding more capital have higher exports (and export growth) consistent with the assumption that ability and capital are complements.

Identifying this misallocation is more challenging for Anderson because she does not have “conflicting” stylized facts, as do Banerjee and Munshi, that can be used to uncover an underlying market inefficiency. Anderson notes that the caste composition of the villages in her sample has likely been stable for centuries. What has changed is the pattern of land ownership in those villages where upper caste Brahmins were absentee landlords in the colonial period. Land reforms in the 1950’s, shortly after Independence, transferred land ownership to lower caste (BAC) cultivators and it is those villages that are BAC-dominated today. Anderson provides evidence that the two types of villages are the same with respect to cultivated area, population, overall access to irrigation, and the functioning of the water market. But if this is true, then there is a comparability problem because the characteristics of the lower castes and the quality of their land will necessarily vary across the two types of villages.

The lower castes were historically independent cultivators (with absentee landlords) in the BAC-dominated villages and we would expect this autonomy to have had long-term effects; indeed, they have significantly higher education than lower castes in the upper caste dominated villages. A second source of variation is that lower castes in the BAC-dominated villages will own land across the quality spectrum, whereas we would naturally expect the upper castes to have captured the best land in the villages where they dominate. The available quality measure, based on land value, does not reveal systematic differences for the lower castes in the two types of villages. However, there is reason to treat this finding with caution. If yields (and incomes) are higher for lower castes in BAC-dominated villages, then this should necessarily be capitalized in the value of their land. The absence of such variation in land prices indicates that either the land market is functioning imperfectly, in which case the prices are not informative, or that there is measurement error in the reported prices, which prevents a meaningful comparison.

The preceding discussion indicates that a comparison of irrigation, yields, or income for the lower castes across the two types of villages will not identify a misallocation unless the comparability issues are addressed. However, the direct information on the caste-based restrictions to trade that Anderson provides is more compelling. The strongest empirical finding is that lower caste water buyers have greater access to irrigation, with an accompanying increase in their yields and income, if they reside in a BAC-dominated village where water sellers are predominantly of the same caste. The issue of comparability remains, but it is reduced because we are now comparing cultivators on the same side of the water market across villages. Given that access to irrigation for water buyers as a whole does not vary across the two types of villages, it must be that upper caste water buyers have greater access to

irrigation in upper caste dominated villages (where water sellers are predominantly upper caste). This symmetry is reassuring, with the general finding that trade appears to be smoother when buyers and sellers belong to the same caste.

Taking this result as given, a natural question to ask is why potential buyers and sellers belonging to different castes cannot enter into long-term bilateral relationships. Social ties, and the social sanctions that go with them, are useful in supporting cooperation in a one-shot game. However, these farmers have been living together in the same village for centuries and, given the low mobility in rural India, would expect to live together long into the future. With a repeated game, social ties would seem to be less relevant, with the potential termination of the long-term relationship being sufficient to deter deviation from the cooperative equilibrium. One possible explanation for the failure of buyers and sellers from different castes to come together is that they start with negative priors about the long-term stability of the relationship. Just as a sufficiently large common shock is needed to jump-start community networks, a positive shock is needed to create a long-term inter-caste bilateral relationship. Anderson's results would suggest that such shocks occur fairly infrequently.

5 Labor

5.1 Historical Background

Castes or *jatis* were originally occupational units and individuals remained within their caste's traditional vocation, from one generation to the next, prior to the arrival of the British. With colonial rule came overseas trade and, subsequently, industrial manufacturing. The accompanying growth of cities resulted in the demand for labor in non-traditional occupations, in the mills, docks, and factories. A second new source of labor demand came from mines and plantations, both in India and overseas. Historical accounts indicate that networks organized around the *jati* played an important role in supplying labor to cities and plantations.

We have seen that commercial activity under the British was centered around the cities of Bombay and Calcutta. Bombay's textile mill owners and factory owners appear to have adjusted their level of output and employment to demand shocks, periodically bearing the cost of idle capacity (Chandavarkar 1994). Similar fluctuations in the demand for labor have been documented on the docks, in the shipping industry, and in general laboring jobs. These are precisely the conditions under which labor networks will emerge, with incumbent workers providing referrals for new hires from their social group. Social

connections within the community provide the incumbent workers with information about the quality of new hires. The same connections ensure that new workers will not shirk even if they are hired temporarily, to maintain the community's reputation in the labor market (and the incumbent worker's reputation in his firm).

In Bombay, the labor networks appear to have been organized around the *mukadam* or jobber, a foreman who was in charge of a work gang in the mill, factory, dockyard, or construction site, and, more importantly, also in charge of labor recruitment.¹² In an environment with frequent fluctuations in labor demand, we would expect the jobber to have recruited from his social group, which in the Indian context would be the *jati*, and this is precisely what appears to have happened. As Morris (1965: 76) argues, "I am using the term [caste] to refer to what are precisely called subcastes (*jati*) ... for any analysis of labor recruitment and the study of the behavior of rural recruits in an industrial environment ... it is entirely inappropriate to lump into larger groups because of similarity of name, function, social status, or region-of-origin subcastes that are not endogamous."

The same pattern of caste-based recruitment appears to have emerged in Calcutta and in the mines and plantations. While business in Bombay was controlled by outsiders, labor was drawn mostly from the Maharashtrian hinterland. In contrast, labor in Calcutta's jute mills and factories was drawn from far away; from the western districts of Bihar, the adjoining eastern districts of Uttar Pradesh, and two Orissa districts, Cuttak and Balasore (Das Gupta 1981). Once again a foreman or *sardar* was in charge of labor recruitment, and once again he appears to have drawn on his caste back home to fill new jobs. The same source districts in Bihar and Uttar Pradesh supplied the bulk of the labor to overseas sugar colonies – Mauritius, the West Indies, and Fiji – during the latter part of the nineteenth century, and while the recruitment of indentured labor to these colonies was heavily controlled by the colonial government, the concentration of particular castes among the migrants indicates that there was, nevertheless, an endogenous aspect to the labor supply.

Within India, caste-based recruitment emerged wherever labor was needed in large numbers. For example, caste-based recruitment was also observed in southern tea plantations. A particular feature of the caste system in South India was the institution of agrestic slavery in which the members of certain castes (and their progeny) were owned by upper caste masters (Alexander 1989, Kooiman 1989). The abolition of slavery in India by the colonial government in 1861 coincided with the opening up of the plantations in South India and Sri Lanka. These plantations were located in previously uninhabited hills

¹²The jobber was known as the *maistry* in the construction industry and the *sarang* in the shipping industry.

and so labor had to be transported from the plains in Tamil Nadu. Initially, the British planters sent their own agents to recruit workers, but within a short time, workers began to migrate independently, in gangs drawn from the foreman or *kangany's jati*. Our dynamic model tells us that once a network has formed, it will strengthen most rapidly in community's with the weakest outside options, which in this context would have been the former slave castes. Luke and Munshi (2011) surveyed a sample of South Indian plantation workers three generations later and found that 68% of the workers belonged to the former slave castes (Pallars and Paraiyars) highlighting the role played by caste networks in supporting mobility, not only into business but also into non-agricultural labor occupations.

Luke and Munshi show how caste networks in the plantations were refreshed and reinforced by new arrivals from the plains, through marriage, from one generation to the next. Networks that formed in the city were similarly reinforced over the course of many generations by the arrival of new workers from the rural origin. Focusing on a single caste, Dandekar (1986) traces the evolution of a network of Jadhavs (a particular caste) from a single village in interior Maharashtra to one jobber who went to Bombay in the 1930's, working first on the docks and later in the textile mills. In 1942, 67% of the Jadhav migrants from that village were working in the textile mills and 4% in other factories. In 1977, their descendants and new arrivals from the village over the next generation were still in the same occupations, with 58% in the textile mills and 10% in other manufacturing industries. De Haan (1997) similarly traces the evolution of a network of Khandait migrants in Calcutta to a single *sardar*. This individual brought an initial group of 20 workers with him from Orissa at the beginning of the twentieth century, and this group was presumably large enough to jump-start a new network that persisted and grew in size over the next century.

The model of network formation and the preceding case studies emphasize the importance of initial conditions. When the model is extended to multiple destination locations or occupations, it generates the additional prediction that the community could get locked into those destinations where it (sometimes fortuitously) first established itself. Caste clusters have been documented among Bombay's dock workers (Cholia 1941), construction workers, and in the railway workshops (Burnett-Hurst 1925), the leather and dyeing industries, and in the Bombay Municipal Corporation and the Bombay Electric Supply and Transportation Company (Chandavarkar 1994). Castes sometimes found niches even within occupations; Gokhale's (1957) survey of textile workers in the 1950's found that particular castes were concentrated in particular departments within the mills, with different castes dominating a given department in different mills. De Haan (1997) documents similar caste segmentation in Cal-

cutta's jute mills. Urban jobs did not necessarily relate to traditional rural occupations; indeed, castes that were traditionally weavers ended up in the spinning section of the jute mills by historical accident, emphasizing the importance of initial conditions in determining long-term outcomes when networks are active.

5.2 Discrimination

One side of caste networks is that they harness social ties within the community to solve market imperfections and increase economic efficiency. The other side is that they exclude outsiders once they are established. This discrimination is not a consequence of social prejudice, but is simply the outcome of economic competition between groups; the foreman would always prefer to recruit from his own caste. In recent years, a number of studies have examined another aspect of caste discrimination in which upper caste employers prefer not to hire lower caste workers. One approach that has been used to identify caste discrimination is the Oaxaca-Blinder decomposition. In this method, the earnings differential between upper castes and lower castes is divided into two components: differences in the estimated returns to endowments such as education and occupation for the two caste-groups, and differences in the endowments themselves.

Banerjee and Knight (1985) apply the Oaxaca-Blinder decomposition to a sample of migrants that they surveyed in Delhi. Earnings are 17% lower for the Scheduled Castes (SC) migrants relative to the non-SC migrants. The two groups of migrants also differ in their endowments; education, the proportion who are salaried, and the proportion in unskilled manual jobs. The estimated earnings function coefficients reveal no difference between the two caste groups for education, age on arrival in Delhi, occupation, and the intercept. However, there is a greater premium for salaried workers versus daily wage earners among the SC's, a greater return to formal sector employment versus informal sector employment for the non-SC's, and a steeper earnings-experience profile for the non-SC's. Overall, differences in the estimated returns to individual characteristics favor the non-SC's, but are small. Depending on whether SC or non-SC earnings function estimates are used, half to two-thirds of the earnings differential can be explained by differences in endowments (mainly education and occupation).

Madheswaran and Attewell (2007) apply the Oaxaca-Blinder decomposition to test for caste discrimination with National Sample Survey (NSS) data. The advantage of their data is that it is nationally representative and available at multiple points in time. Scheduled Caste (SC) and Scheduled Tribe (ST) workers earn 5% less than other workers in 1983, 8.4% less in 1993-94, and 8.9% less in 1999-

2000. Corroborating Banerjee and Knight's finding, the endowment component explains the bulk of the earnings differential (70-87%) over the three time periods. If occupational variables are included in the earnings function, the endowment effect increases even further. Kijima (2006) applies a variant of the Oaxaca-Blinder decomposition to explain the consumption gap between SC/ST and non-SC/ST households over multiple NSS rounds (1983-1999). In contrast with the studies listed above, he finds that differential returns to endowments (education and occupation) explain half of the consumption gap, with the decomposition remaining stable over the sample period.

It is standard practice in this literature to interpret differences in the returns to endowments between groups as wage discrimination and differences in the endowments themselves as pre-market discrimination. This interpretation may, however, overstate the extent of current discrimination. The lower castes, particularly the SC's, were socially disadvantaged for centuries. They were originally assigned the lowest status occupations, requiring little investment in human capital, and even when they managed to achieve occupational mobility under the British, they typically ended up in low-skill industrial jobs. We have noted that caste networks remain active in the Indian labor market. Once the upper caste networks had established a presence in particular sections of the labor market, they would have ensured that outsiders were subsequently excluded from those jobs. Persistent differences in occupations between SC's and non-SC's could thus be determined by historical circumstances, rather than by discriminatory practices by upper caste employers today.

Given historically low levels of human capital, and the importance of parental human capital in determining pre-school human capital, the education deficit for the SC's could be similarly determined by historical circumstances that are slow to change rather than by current discrimination in education. Differences in the returns to endowments, particularly education, are also difficult to interpret as discrimination.¹³ For example, well qualified non-SC parents could transmit non-cognitive skills and other traits that are complementary to education to their children. If these parental inputs are observed by the market but not the researcher, then the estimated returns to education would be lower for the SC's. Caste-based reservation in institutions of higher education would also affect the estimated returns to education. If affirmative action serves its intended purpose, then SC students enrolled in a given course in a given college will be less prepared than their non-SC classmates when they join. If they fail to catch up over the course of the program, then they will graduate with the same credentials but

¹³Banerjee and Knight detect no differences in the returns to education for SC's and non-SC's, whereas Madheswaran and Attewell's estimates indicate that the returns are consistently lower for the SC's.

will be less competent. If this difference in competence is accounted for by employers, but unobserved by the researcher, then the estimated returns to education will, once again, be lower for the SC's. In general, it is difficult to disentangle discrimination from other channels, based on unobserved individual characteristics, through which inter-group earnings differentials are generated.¹⁴

Following Bertrand and Mullainathan's (2004) research design, which was used to document racial discrimination in the United States, a field experiment in which caste affiliation is randomly assigned to otherwise identical job applications is the cleanest way to identify caste discrimination in the Indian labor market. Thorat and Attewell (2007) conducted such an experiment by replying to job advertisements in major English daily newspapers, sending three applications to each call: an applicant with an upper caste name, an applicant with a Scheduled Caste (SC) name, and an applicant with a Muslim name. The applicants had identical educational qualifications and experience. No explicit mention of caste or religious background was made in any application; the only distinguishing feature was the name of the applicant. In addition, a "discordant" application was submitted to each advertisement; for jobs requiring an undergraduate degree, an applicant with an SC name and a master's degree was submitted, and for jobs requiring a postgraduate degree, an applicant with an upper caste name and an undergraduate degree was submitted.¹⁵

The main finding of the study is that equally qualified applicants with SC or Muslim names were significantly less likely to be called for an interview or for testing. In the first discordant experimental treatment, the probability of success for the over-qualified SC applicant was greater than (less than) the corresponding probability for an appropriately qualified SC (upper caste) applicant. However, these differences are not significant at conventional levels. The second discordant treatment had greater statistical power, with under-qualified upper caste applicants being significantly less likely (as likely) to be successful as appropriately qualified upper caste (SC) applicants.

These experimental results indicate that unobserved individual characteristics that are correlated with caste identifiers may not entirely explain the findings from the Oaxaca-Blinder decompositions. It appears that private firms do select applicants explicitly by caste. The next question to ask is whether this discrimination is statistical; i.e. firms use caste identifiers to proxy for applicants' unobserved

¹⁴Deshpande and Newman (2007) compared the job-market expectations of SC and non-SC students in their final year of university in Delhi. The SC students expect a lower wage than the non-SC students, but this could be due to differences in endowments or perceived discrimination. The SC students perform worse than the non-SC students on the final university exam and have weaker family backgrounds (measured by parental occupation and education).

¹⁵Initially, one set of four applications was submitted for each advertisement. However, this was later increased to three sets to increase statistical power.

characteristics, or taste-based; i.e. based on preferences for a particular group. One way to dig deeper and answer this question is through qualitative research. Jodhka and Newman (2007) conducted in-depth interviews with managers responsible for hiring workers in 25 large private firms in the New Delhi area. The central finding was the view, expressed by virtually every respondent, that workers should be selected strictly on merit. However, merit was defined broadly enough to include family background, and while the respondents did not explicitly mention caste as a consideration, they did endorse regional stereotypes. In a situation where information on family background is missing, as in the field experiment above, they would very likely have used caste and regional characteristics as a proxy for family characteristics when making their hiring decision. Complementing this finding, Banerjee et al. (2009) conducted a resume audit study in Delhi that was restricted to software and call center jobs. They find no evidence of discrimination against Scheduled Castes (SC's), Scheduled Tribes (ST's), or Other Backward Castes (OBC's) for software jobs. However, they do find large and significant differences in callback rates between upper castes and OBC's (and to a lesser extent SC's) for call-center jobs. These results are consistent with statistical discrimination, since unobserved family characteristics that are correlated with caste status are more likely to matter for the call-center jobs. Statistical discrimination is not necessarily economically inefficient, but it does punish high-achieving members of under-performing groups, and it can lead to self-fulfilling stereotypical behaviors in equilibrium.¹⁶

5.3 Caste-Based Labor Networks and Dynamic Inefficiency

Although caste networks may reduce imperfections in the labor market, they can also generate inefficiencies of their own. We saw earlier that the failure of capital to cross caste lines results in a mismatch between entrepreneurial talent and investment in capital. Caste-based networks can similarly misallocate labor inputs by excluding competent outsiders from the market. Even the gain in efficiency within the caste through internal cooperation is not without cost. Munshi and Rosenzweig (2006) show that the same caste networks that supported the movement of their members from the village to the city historically, discouraged the occupational mobility of their members in subsequent generations. This phenomenon is not unique to India. Heavily networked blue-collar communities, originally established

¹⁶There is no direct evidence that I am aware of that shows (conditional on observed credentials) that lower caste workers are less productive. Deshpande and Weisskopf (2011) use data from the Indian Railways to show that an increase in the fraction of reserved workers does not appear to lower productivity. These results must be treated with caution, however, given the endogeneity of the inputs and the fraction of jobs that are reserved.

by dynamic migrants, have been observed to be especially resistant to change in other economies; e.g. Gans (1962), Kornblum (1974).

To provide economic foundations for this inertia, we return to the dynamic model of migration that was described in the previous section. There was a continuum of ability levels in that model, which sorted individuals into business and the traditional non-business occupation. We now assume that there are two levels of ability – high and low – and two occupations – skilled and unskilled – in the labor market. Education plays an important role in the empirical application that follows by determining the type of job that the individual receives. We incorporate this in the model by assuming that high (H) education leads to the high-skill occupation, while individuals with low (L) education are channelled into the low-skill occupation. Ability sorts individuals into occupations once again, but it is now via the educational choices that they make. It costs \bar{C} for low ability individuals to attain high education, whereas the corresponding cost for high ability individuals is $\underline{C} < \bar{C}$. We normalize so that the cost of attaining low education is zero.

Absent any other forces, high ability individuals would select into the high-skill occupation if $W_H - \underline{C} > W_L$ and low ability individuals would select into the low-skill occupation if $W_H - \bar{C} < W_L$. Once we introduce community networks, however, all individuals in a community could choose the same occupation, regardless of their ability, with these community-level choices persisting over many generations. A notable feature of the historical descriptions of urban caste-based networks in India is that they focus on working class jobs. This is a feature of labor market networks worldwide, possibly because the information and commitment problems that give rise to the networks are most acute in those jobs. In an early study set in Chicago, Rees (1966) found that informal sources accounted for 80% of all hires in eight blue-collar occupations and 50% of all hires in four white-collar occupations. Matching these patterns, 68% of blue-collar workers and 38% of white-collar workers reported having received help finding a job in Gore’s (1970) study of migrants in Bombay. Thirty years later, Munshi and Rosenzweig (2006) document similar referral patterns in an area close to Gore’s study, with 70% of blue-collar workers and 35% of white-collar workers reporting that they received help from a member of their caste in finding their first job. The model incorporates this feature of the labor market by assuming that the skilled wage, W_H , is exogenously determined, whereas the unskilled wage, W_L , is increasing in the (lagged) size of the community network in that occupation. In particular, members of the community from the previous generation in the low-skill occupation provide referrals, and increase wages, for the next generation.

Let $W_L(0)$ be the unskilled wage when no one from the individual's community selected that occupation in the previous period. Let $W_L(N)$ be the corresponding wage when all N members of the previous generation selected the unskilled occupation. If communities were concentrated entirely in either the skilled or the unskilled occupation in the initial period, then the following conditions ensure that there will be occupational persistence at the level of the community in all subsequent generations:

$$\mathbf{C1.} \quad W_H - \bar{C} > W_L(0)$$

$$\mathbf{C2.} \quad W_H - \underline{C} < W_L(N).$$

The first condition says that if the low-skill network was not active in the previous generation, then individuals of both types would invest in high education in the current generation and end up in skilled jobs. The second condition says that if everyone in the community selected into the network in the previous generation, then individuals of both types would select low education and end up in the unskilled occupation. If conditions C1 and C2 are satisfied, it follows that communities will stay in the initial equilibrium from one generation to the next, with everyone either investing or not investing in education.

Economies will, however, restructure over the course of the development process and one consequence of development is that the returns to education, and the skilled wage, will start to grow. When the inequality in condition C2 is just reversed, high ability individuals from communities that traditionally exclusively selected into the low skill occupation will invest in high education and earn ϵ more in the skilled occupation than they would have if the entire community remained in the traditional occupation. If the fraction of high ability individuals in these communities is non-negligible, then their exit from the network will result in a substantial (first-order) decline in the wage received by low ability individuals. From a utilitarian perspective, welfare will decline, providing an economic rationale for community-based restrictions on mobility.

Individuals who select out of the traditional low-skill occupation will often move to a new location. The conventional punishment mechanisms that maintain cooperation within communities will then no longer be effective. Munshi and Rosenzweig argue that an alternative strategy to maintain cooperation in that case would be to instil a strong sense of community identity in childhood, which ensures that individuals remain tied to their community in adulthood.¹⁷ The idea that identity, and values more

¹⁷Another possibility would be to provide compensatory transfers to those individuals to induce them to stay. However, the level of these transfers would need to be agreed upon long before individuals entered the labor market because

generally, are purposefully instilled to further community objectives is in line with previous work on this topic in economics; e.g. Bisin and Verdier 2000, Tabellini 2008. It also explains why heavily networked blue-collar communities tend to instill an especially strong sense of identity among their members.

The community identity described above aligns individual choices more closely with the social optimum. While this identity may thus be welfare enhancing when it is first put in place, it can result in a dynamic inefficiency if it persists in subsequent generations past the point where W_H has grown large enough that it is socially optimal for the high ability individuals to exit. Cultural norms and social values are persistent by design, which explains why blue-collar communities often appear to stubbornly resist change (despite the fact that the same communities were extremely dynamic when they formed). Munshi and Rosenzweig build on this idea to document a dynamic inefficiency in Bombay's labor market, which was historically organized around caste-based networks. The starting point for Munshi and Rosenzweig's analysis is the economic and financial liberalization of the Indian economy in the 1990s. This increased white-collar wages (W_H in the model) substantially. The question is whether the occupational persistence associated with the blue-collar networks prevented the next generation of workers from taking advantage of these new opportunities.

When the Indian economy restructured in the early 1990's, the next generation of workers would have still been in school. Munshi and Rosenzweig's strategy to test for occupational persistence exploits the fact that schooling choice is often a good predictor of future occupational outcomes. Schooling in Bombay, the setting for their research, can be either in English or Marathi (the local language). Marathi schooling channels the child into networked blue-collar jobs, whereas more expensive English schooling substantially increases the probability that the child will obtain a high-wage white-collar job in the future. The economic restructuring thus effectively increased the returns to English schooling. If networks discourage individual mobility, then parents in some castes could, nevertheless, have continued to make schooling choices for their children that channelled them into the traditional blue-collar occupation.

Munshi and Rosenzweig test for this inter-generational occupational persistence by examining schooling choices for 20 cohorts of Maharashtrian children who entered first grade between 1982 and 2001, in one area of Bombay city consisting of 30 schools (20 Marathi and 10 English). The key result is that the probability that a boy will be schooled in English is decreasing in the proportion of

investment in education is necessary for entry into the skilled sector. Reaching such an agreement *ex ante* is difficult in an economy where the (steeply increasing) wage trajectory is uncertain. The community would also need to credibly commit to making these transfers *ex post*, thereby creating potentially substantial income inequality among individuals doing the same job.

men in his caste employed in blue-collar (working class) jobs. This result is robust to the inclusion of parental education and household income as regressors. One concern is that children from lower castes, who were historically concentrated in the blue-collar occupations, have lower pre-school human capital, which leads their parents to favor Marathi schooling. The parental and household variables control for this to some extent, but Munshi and Rosenzweig also implement an alternative approach that relies on the fact that the blue-collar networks were largely restricted to men. Male occupational choice in the previous generation should not affect girls' schooling choices, and this is indeed what they find (once parental and household controls are included in the regression). Pooling boys and girls, and including caste fixed effects, the second (complementary) result is that the gender gap in schooling choice is increasing in male occupational choice in the previous generation. More importantly, this relationship remains strong and stable over the course of the 1990's, even as the returns to English grew, providing micro-level evidence of the lock-in into specific traditional occupations.

At the aggregate level, this lock-in is reflected in the absence of convergence in schooling choice between upper caste and lower caste boys. Upper caste boys who entered school in the 1980's are much more likely to be schooled in English than lower caste boys, reflecting the fact that upper castes historically had access to administrative and professional jobs, while lower castes were concentrated in blue-collar occupations. While there is a steep increase in the fraction of boys enrolled in English schools in the 1990's, there is no convergence between the castes. This can be explained by the stable inter-generational correlation in occupational choice over the course of the 1990's that we noted above. In contrast, while there is also a caste-gap in schooling choice for the girls in the 1980's, this gap narrows dramatically in the subsequent decade. Although lower caste networks may have historically supported the movement of their members from the village to the city, with accompanying occupational mobility, many generations later they appear to hold their descendants back.

As described in the previous section, caste-based business networks are continually forming in the dynamic Indian economy. We would expect labor market networks to be similarly forming as new opportunities arise, with high ability individuals leading the way into new occupations. At the same time, the evidence presented in this section indicates that other caste networks will simultaneously be decaying, although social frictions could substantially delay this process. Munshi and Rosenzweig show that relatively high ability children (where ability is proxied by the father's years of schooling) are the first to exit the traditional blue-collar occupation, providing a useful counter-point to the patterns of selection that are observed with network formation. A fruitful area for future theoretical and empirical

research would be to bring the twin processes of network formation and network decay together.

6 Education

6.1 Aggregate Trends

Given their historically disadvantaged circumstances, the lower castes lagged far behind the upper castes with respect to education and other socio-economic indicators at independence in 1947. Article 46 of the Indian Constitution put into place one of the largest and most aggressive affirmative action programs in the world to rectify this historical disadvantage. Seats in universities and central government jobs are reserved for Scheduled Castes (SC's) and Scheduled Tribes (ST's) to match their share of the population; 15% for the SC's and 7.5% for the ST's. More recently, reservation has been extended to Other Backward Castes (OBC's), bringing the overall share of reserved seats close to 50%. In an economy where university seats are heavily oversubscribed and unemployment is high, the affirmative action program has naturally been highly controversial. Much of the public debate has centered on whether the SC's and ST's have actually benefitted materially from the reservations and, if so, whether privileged individuals within those groups have captured the bulk of the targeted resources.

Despite the importance of these questions, there have been few broad-based evaluations of the affirmative action program and, more generally, examinations of the relative circumstances of the SC's and ST's in the decades after independence. In the most comprehensive study to date, Hnatkovska, Lahiri, and Paul (2012) use data from five successive rounds of the National Sample Survey (NSS), covering the 1983-2005 period, to measure differences in education, occupations, wages, and consumption between SC/ST and non-SC/ST households. The key finding is that there is significant convergence in all of these outcomes over the period under consideration, which extends well past the point in the early 1990's when the Indian economy started to liberalize. Decomposition analysis indicates that much of the convergence in wages and consumption is driven by convergence in education between the SC/ST's and non-SC/ST's. Moreover, the convergence is observed at all points in the respective distributions.

In a follow-up paper, Hnatkovska, Lahiri, and Paul (2013) compare the inter-generational mobility of SC/ST's and non-SC/ST's using the same data. Their analysis takes advantage of the fact that 60% of households in the NSS sample include members from different generations. Complementing their previous findings, inter-generational mobility of SC/ST's was lower than that of non-SC/ST's in 1983, but rises faster over the subsequent decades and is roughly equal by 2005. For example, the

probability that an SC/ST child's education differs from his parent's education increases from 0.42 to 0.67 over the 1983-2005 period, whereas the corresponding probabilities for the non-SC/ST's are 0.57 and 0.67. Similar convergence in inter-generational mobility is observed for income and occupation. Once again, these gains in mobility are observed at all points in the distribution. In particular, the largest inter-generational changes in education for SC/ST's are at the bottom of the distribution; out of illiteracy into primary and secondary school.

While these results are heartening, there are a couple of important caveats. First, the convergence cannot be entirely attributed to affirmative action (which targets higher education). Some of this catch-up could be mechanical, since the SC/ST's started behind the non-SC/ST's. Moreover, the government has taken steps to expand basic education, as described below, and caste networks, which are active in many economic domains could also have played a role in this convergence. Second, while there may have been convergence, substantial gaps in the *level* of education between SC/ST's and non-SC/ST's remain. Using representative NSS data, Hnatkovska, Lahiri, and Paul show that this gap is as large as two years of schooling. The discussion below provides micro-foundations for the observed convergence and the persistence of the caste-gap, describing the challenges faced by SC/ST students in school, the benefits they receive through affirmative action in university, and the role played by caste networks in determining the demand for education.

6.2 Access to Schooling

It is well known that school participation is particularly low among the SC/ST's. Drèze and Kingdon (2001) use data from the PROBE survey, covering randomly selected villages in the northern states of Bihar, Madhya Pradesh, Rajasthan, Uttar Pradesh, and Himachal Pradesh to compare school attendance and grade attainment for SC/ST and non-SC/ST children. Controlling for parental education, household wealth, measures of school quality, and teacher inputs, SC/ST children (and OBC children) are significantly less likely to attend school.

One explanation for this finding is that the school-quality and teacher-input variables do not fully capture the (lower) quality of education that lower caste children receive. The Government of India has made it a priority to provide a school within walking distance of each rural household. Given the segregation by caste within villages, this policy has resulted in lower caste students being separated from upper caste students, denying them the positive peer effects they would otherwise have enjoyed (Kochhar 2008). The additional effect of this policy is that the average lower caste student resides

in a neighborhood of smaller size and will, therefore, be enrolled in a smaller school (with worse infrastructure) than the average upper caste student (Rawal and Kingdon 2010).

Even if upper caste and lower caste students were mixed in school, the available evidence indicates they would not be treated the same. Rawal and Kingdon use data on primary school children in rural Bihar and Uttar Pradesh to show that student performance increases substantially when the teacher belongs to the same caste. Hanna and Linden (2012) conducted an experiment in which teachers were recruited to grade examinations. Candidate characteristics – caste, gender, and age – were randomly assigned to each cover sheet to ensure that there was no relationship between these observed characteristics and exam quality. The main finding is that exams assigned to lower castes are scored lower by examiners, although the effects are extremely small in magnitude. Interestingly, this bias is accounted for entirely by low caste graders. It appears to be generated by statistical discrimination – a prior belief in that group of graders that lower caste students are worse on average – because the bias disappears over time as the link between demographic characteristics and test performance is revealed (this would not be the case with taste-based discrimination).

In addition to differential treatment by teachers, lower caste students must also confront negative stereotypes when they interact with higher caste students in the classroom. A large literature in social psychology finds that individuals belonging to a group associated with a particular stereotype perform worse on tasks associated with that stereotype when observed by others. In the Indian context, this implies that lower castes, who were historically perceived to be intellectually inferior, would perform worse on cognitive tasks if the results of their efforts were made public. Hoff and Pandey (2006) test this hypothesis by conducting an experiment in which low-caste and high-caste sixth and seventh graders were assigned the task of solving mazes in six-person groups. When the participants' caste status was kept anonymous, low-castes and high-castes performed equally well on the task. However, when caste identity was revealed, in homogeneous or mixed-caste groups, there was a sharp decline in low-caste performance. This result is obtained from a stylized lab-in-the-field experiment, which may be subject to its own biases. For example, the high-castes perform equally well when their caste status is kept private or is revealed in mixed-caste groups. However, their performance declines substantially when their caste status is revealed in high-caste only groups. This curious result is difficult to explain and it suggests that more evidence is needed, with more comprehensive data, in support of the experimental findings. Indeed, the available evidence indicates that conditional on attendance, grade attainment does not vary by caste (Kingdon 1998, Drèze and Kingdon 2001). Where caste differences in performance do

appear is in higher education, where affirmative action has substantially increased college attendance for (less qualified) lower castes in recent decades (Darity, Deshpande, and Weisskopf 2011).

6.3 Affirmative Action

As discussed in Frisanco Robles and Krishna (2012), there are three issues associated with affirmative action in higher education: (i) Targeting – whether the beneficiaries are truly disadvantaged or whether the resources are captured by wealthy elites within the historically disadvantaged groups.¹⁸ (ii) Catch up – whether there is convergence between reserved groups, who start behind by design, and unreserved groups over the course of the university degree. (iii) Mismatch – whether the beneficiaries of affirmative action would have actually done better by not taking advantage of their legally mandated opportunity, perhaps because the college or the major they chose turned out to be too challenging.

Bertrand, Hanna, and Mullainathan (2010) address the targeting and mismatch issues with data on all individuals who took the engineering entrance exam in one Indian state in 1996.¹⁹ The minimum score for admission, out of a total of 900, is 480 for the upper castes, 419 for the OBC's, and 182 for the SC's, confirming that affirmative action does indeed bring in weaker students at the margin from historically disadvantaged castes. To gather more information on the socio-economic background of the candidates and their post-college labor market outcomes, the research team conducted a household survey between 2004 and 2006 in which the sample was based on candidates on either side of the cutoff test score for their caste group.

Combining the survey data with NSS data, which is representative of the state population, the first finding of the study, related to targeting, is that although lower caste applicants are positively selected from their population sub-group, affirmative action does redistribute educational resources to less wealthy households. Lower-caste applicants to the engineering colleges have household incomes that are four times larger than the average income of their caste-group in the state. Nevertheless, the lower

¹⁸A related concern is that individuals belonging to the upper castes could misreport their caste in order to take advantage of affirmative action. It is commonly believed that such identity manipulation does occur. For example, Cassan (2015) documents caste identity manipulation in the colonial Punjab by using changes in the caste composition across census rounds. However, there is no quantitative evidence on the prevalence of this practice. Moreover, while it may be possible for individuals to misreport their caste affiliation to the government and take advantage of affirmative action, it is unlikely that they will succeed in changing their identity on the ground; i.e. in the village. Indeed, Christian converts continue to marry within their original castes despite having changed their religion generations ago.

¹⁹As discussed below, Bagde, Epple, and Taylor (2016) use similar data with a different identification strategy to assess whether reservation improves college performance for beneficiaries. In contrast, Kochar (2010) and Frisanco Robles and Krishna (2012) focus on catch-up by examining student progress through college. A limitation of both studies is that they are restricted to a single college and lack information about the entire pool of candidates (differential selection by caste group into college is thus a major concern).

caste applicants who are just above the test score threshold that secures them a position in engineering college have household incomes that are just over half the income of the upper caste candidates who they displace (with test scores just below the threshold, for their caste-group, that would have secured them admission).

The second finding of the study, related to mismatch, is that despite the fact that lower caste entrants have much weaker skills than upper caste entrants (by virtue of their substantially lower scores on the entrance exam), they still gain by attending engineering college. However, this gain is more than offset by the loss in income for the high caste candidates they displace. Thus, even if there is catch-up over the course of the engineering program, the two groups do not converge with respect to their earnings potential by the time the program is completed.

Bertrand, Hanna, and Mullainathan estimate the gain from affirmative action by comparing the income of lower caste students above and below the entrance exam cutoff for their group. They estimate the location of the cutoff using a standard structural break approach; i.e. they select as the cutoff the score that minimizes the sum of squared residuals in a regression of engineering college attendance on a dummy equal to one if an individual's test score is above a particular score threshold, zero otherwise. Lower caste students just above and just below the cutoff can be assumed to have effectively the same intrinsic characteristics and so a comparison of the two groups of students provides an estimate of the gain from affirmative action for the marginal student who benefits from that policy. In practice, Bertrand, Hanna, and Mullainathan sample applicants with scores in a relatively wide range around the threshold. While the two groups – above and below the threshold – do have the same observed characteristics on average, differences in unobserved characteristics are now a concern. Their point estimates are also now interpreted as the effect of affirmative action on a group (of students) that is less clearly defined.

A more recent examination of affirmative action in higher education by Bagde, Epple, and Taylor (2016) uses similar data but a different identification strategy to estimate the effect of reservation in engineering colleges in one Indian state on the performance of lower caste (SC, ST, and OBC) students one year into the program. Their approach takes advantage of the fact that all entrants to the 215 colleges in the state must take a common entrance exam and, once admitted, sit for another common exam at the end of the first year. Bagde, Epple, and Taylor also obtain high school test scores for most (89 percent) of the candidates, which provide additional information on intrinsic ability.

Instead of focussing on the marginal entrant, Bagde, Epple, and Taylor estimate the effect of

affirmative action on admitted students over the entire ability distribution. They do this by constructing two ranks; the absolute rank of each individual, based on the performance on the entrance exam, and the effective rank, which adjusts for the caste reservation. Conditional on the absolute rank or, equivalently, the score on the entrance exam, which measures intrinsic individual ability, the effective rank determines the effect of reservation on both entry into college and first-year performance (conditional on entry). High school performance provides an additional measure of ability and, in practice, latent ability is specified to be a polynomial function of the score on the college entrance exam and the high school exam.

The results of the analysis are as follows. First, the probability of college attendance is (not surprisingly) increasing in effective rank. Second, conditional on latent ability as measured above, performance on the first-year exam is increasing in effective rank. Note that the effective rank contains no information about individual ability beyond what is contained in the entrance exam score; i.e. the absolute rank. This result thus provides an unbiased estimate of the effect of affirmative action on college performance. An augmented specification interacts effective rank with caste-gender dummies. The coefficients on the interaction terms are insignificant, indicating that all eligible caste groups benefit equally from affirmative action.²⁰

The preceding results tell us that affirmative action increases the college attendance of reserved groups and improves their performance (conditional on attendance). However, they do not provide a mechanism through which these benefits are generated. One natural mechanism is that affirmative action improves the quality of colleges to which reserved students have access and this improves their performance. College quality, measured by the mean entrance exam rank of men in the unreserved (Open) category, is increasing in effective rank. Instrumenting for college quality with effective rank, first-year performance is found to be increasing in college quality.²¹ Contrary to the mismatch hypothesis, access to better colleges by less prepared lower caste students significantly improves their outcomes.

Bagde, Epple, and Taylor conclude their analysis by quantifying these improvements. For this exercise they use the estimated coefficients from the analysis described above, replacing effective rank with absolute rank. The counter-factual analysis indicates that affirmative action increases engineering

²⁰With a more parsimonious specification that includes effective rank and its interaction with the gender dummy, effective rank has a positive and significant effect for both men and women, although the effect is slightly weaker for women.

²¹Conditional on the latent ability measure, which subsumes the absolute rank, the effective rank can only affect first-year performance by changing the college that a student attends. The instrument thus satisfies the exclusion restriction.

college attendance for men and women from reserved castes at all points in the entrance exam distribution, although the effects are strongest in the middle of the distribution.²² Despite these gains, the lower castes are under-represented in the engineering colleges relative to their share of the population. With regard to the impact of reservation on test scores, average gains for SC and ST men are 0.12 and 0.18 standard deviations, respectively, with similarly sized effects for SC and ST women. These magnitudes are roughly in line with the estimated effect from large-scale education interventions in the U.S. and other developing countries, including India.

6.4 Demand for Education

The preceding studies describe how different caste groups fare in the Indian educational system. While the convergence in education levels documented by Hnatkovska, Lahiri, and Paul may be generated, in part, by universal access to primary schooling and affirmative action in higher education, the differential *demand* for education, by caste, may also have played a role in this convergence. Munshi and Rosenzweig (2006) document differences in schooling and occupational choice, by caste and gender, in Bombay city. In their urban setting, lower caste men, who have been engaged in heavily networked blue-collar jobs for generations, are least responsive to the new opportunities that become available in the globalizing Indian economy. This inertia is reflected in the schooling choices made by lower caste boys. Globalization will also affect schooling choices in rural India, where the majority of the population resides, but the new opportunities that arise will mostly be low-skill (urban) occupations that require a different type of education. Luke and Munshi (2011) exploit a unique setting – a group of tea plantations in the South Indian High Range – to assess how economic development will affect schooling choices, by caste, in the future. The key requirement when estimating differential effects by caste on schooling choices is to control for differences in household wealth across castes. As discussed below, Luke and Munshi’s research setting satisfies this requirement.

The High Range, a mountainous area straddling the modern Indian states of Kerala and Tamil Nadu, was virgin forest until it was acquired by British planters and converted into tea plantations in the last quarter of the nineteenth century. Because the plantation land was previously uninhabited, workers were brought to the High Range from the plains in Tamil Nadu. The workers on the tea

²²Unlike most studies on affirmative action in higher education, which focus on broad caste groups (SC, ST, OBC), Cassan (2015) examines the effect of reservation within caste groups. He does this by exploiting the re-drawing of state borders in 1956, which resulted in the same caste having differential reservation status in different parts of a state. These discrepancies were resolved in 1976. Newly available access to reservation is found to increase education for SC boys at the expense of SC girls.

plantations today are the third-generation descendants of those migrants, whose population has been supplemented by a fresh influx of new workers from the plains in each subsequent generation through marriage.

The workers in the High Range continue to be tied to their ancestral communities in rural Tamil Nadu, despite having lived in the tea plantations for many generations. Loans and transfers flow back and forth between the tea plantations and the origin communities, the children of the workers are often sent home to study, and many workers will buy land, build a house, and return to their ancestral villages when they retire. Perhaps the most distinctive feature of South Indian kinship structure is marriage among close relatives. Reinforcing existing network ties, many workers continue to marry their children to relatives from the ancestral location in the traditional fashion. The key parental decision is the extent to which children will be tied to their ancestral communities and the traditional economy. This decision will, in turn, determine the probability that children will marry among their relatives as well as their (optimal) educational attainment.

As in Munshi and Rosenzweig (2006), attachment to the traditional network lowers parents' incentives to invest in their children's education, but for different reasons. First, education does not help the child (particularly the daughter) match with a better partner if she is married in the customary fashion to a relative. Second, attachment to the network reduces spatial mobility and accompanying job opportunities, as in Munshi and Rosenzweig (2016), with an accompanying decline in the returns to education. The heterogeneous caste composition in the tea estates allows Luke and Munshi to test this hypothesis. Recall that a particular feature of the caste system in South India was the institution of agrestic slavery. Abolition of slavery coincided with the establishment of the tea plantations and, not surprisingly, workers belonging to the former slave castes – the Pallars and the Paraiyars – are disproportionately represented in the High Range. The slave castes were historically spread over a very wide area, with their members having minimal social interactions with each other. We would expect their rural networks to be relatively weak and, thus, the expectation is that workers from those castes would have the greatest incentive to invest in their children's education.

Using data from a survey of plantation workers, augmented with administrative data on income over multiple years, Luke and Munshi show that workers belonging to the former slave castes have the same income and the same jobs in the tea plantations, but are less attached to their ancestral communities and have higher levels of education than the non-slave castes.²³ It is worth noting that these statistics

²³The fact that Pallar and Paraiyar workers have the same jobs and earn the same income but have higher education

are very similar across the non-slave castes, despite the fact that they are drawn from very different positions in the social hierarchy.

Looking within the household, Luke and Munshi show that part of the difference in education across castes in the tea plantations can be explained by variation in the incentives of men and women to remain attached to their ancestral communities. Interactions with the ancestral community and the prospect of permanently settling there after retirement are especially burdensome for women belonging to the former slave castes. Thus, while both parents might benefit from mutual insurance, old age support, and other forms of assistance organized around their ancestral communities, the cost of maintaining the social ties that are necessary to obtain this social support may be substantially larger for the women than the men in those castes.

The wider gender preference-gap among the former slave castes is likely reinforced by the greater ability among the women in those castes to act on their preferences. Lower caste women historically worked outside the home, and although they were typically employed in low-paying menial occupations, it has been argued that the income and the independence that this work provided gave rise to a culture in which these women were relatively assertive within their households (Kapadia 1995, Chakravarti 1993). Given that culturally determined beliefs and practices are often slow to change, we expect these caste differences to persist in the High Range even though all women work in the tea estates, irrespective of their caste affiliation.

Luke and Munshi test for these gender-differences, by caste, by estimating the effect of an increase in female income, net of total household income, on important decisions and outcomes for the children. Women have permanent full-time jobs as tea-leaf pluckers in the plantations and they actually earn more than the men on average. Exploiting exogenous variation in male and female incomes generated by differences in rainfall across plantations and using a variety of independent estimation strategies, Luke and Munshi find that a relative increase in female income weakens the family's ties to the ancestral community as the woman gains bargaining power within the household. The children are less likely to marry a relative, to be schooled in the ancestral location, and to ultimately settle there. At the same time, an exogenous increase in relative female income increases the educational attainment of the children. Most importantly, significant female income effects are obtained for the former slave castes alone, accounting for part of the caste differences in schooling and marriage choice observed in both

is not necessarily inconsistent. Higher education for someone growing up in the tea plantations increases the probability of exit, with access to better job opportunities, in adulthood, but this outcome is not certain. What we observe are those individuals who stayed behind.

the parental and the current generation.

The apparent willingness of the Pallar and Paraiyar women to act on their preferences does not, however, come without a cost. Previous evaluations of female income-generation programs in developing countries have documented a male backlash, as measured by the increased incidence of marital violence, when recently empowered women shift household decisions away from their husbands' preferences (Schuler et al. 1996). Similar patterns of marital violence have been documented in the international migration literature as well (Min 2001). Luke and Munshi provide evidence that a relative increase in female income among the former slave castes also leads to an increase in marital violence. In contrast, female income has no effect on household decisions and outcomes among the other castes in the tea estates and has no effect on marital violence as well. Apart from providing independent support for the cross-caste results described above, this result also links marital violence, a serious social problem in India, to underlying caste-based institutions in a way that has not been done before.

Munshi and Rosenzweig's (2006) analysis of educational choices in urban India and Luke and Munshi's (2011) analysis of these choices in the tea plantations both indicate that historically disadvantaged caste groups may be especially responsive to new economic opportunities, leading them to invest in their children's education, because they have fewer ties to the traditional economy to hold them back. Demand-side effects thus explain part of the convergence across castes that was documented by Hnatkovska, Lahiri, and Paul in recent decades. Luke and Munshi's analysis indicates, in addition, that disadvantaged individuals within those groups, notably the women, may be at the forefront of these changes, providing an interesting and relatively unexplored gender component to the role played by caste in the changing Indian economy.²⁴

7 Politics and Access to Public Resources

7.1 State and National Politics

Clientilism, a phenomenon by which politicians deliver benefits to specific individuals or groups of individuals in return for their votes, is pervasive in many developing countries. For a reciprocal arrangement of this nature to be sustained, each side must credibly commit to cooperate with the other. Politicians

²⁴Given that lower caste women face fewer restrictions on mobility and labor force participation, we would expect them to be more responsive to new economic opportunities than upper caste women. However, an experiment providing basic financial literacy and business skills to women in Ahmedabad generated the curious result that upper caste women but not lower caste women responded to the intervention (Field, Jayachandran, and Pande 2010). The argument that the returns to training were for some reason higher from more restrictive groups in this setting does not square with the additional finding that there was no effect for the most restricted Muslim women.

represent political parties, which have a long-term reputation to maintain with the electorate, and so will fulfill their promises even if they are elected for a single term and have personal preferences that do not align with the promises they made *ex ante*. While it is difficult to keep track of the voting behavior of individuals, well organized groups can similarly credibly commit to honoring their reciprocal obligation to a political party (because they too have a long-term reputation to maintain). In India a natural social unit around which “vote banks” are organized is the caste or *jati*. We have seen in many different settings that social connectedness within the caste facilitates private economic activity by solving information and commitment problems. The same social ties can be used to support reciprocal relationships between castes and parties, and it is thus not surprising that caste has played a central role in Indian democratic politics.

Yadav (1999) divides Indian politics after independence into three electoral systems. The first electoral system, covering the first four general elections from 1952 to 1967, was dominated by the Congress Party. The opposition was regionally fragmented and often came from within the Congress. “The voter of course did not vote as an individual but rather as a member of a politicized *jati*” (Yadav 1999: 2394). For the caste to vote as a block, voters spread over a wide area must coordinate on a candidate. One way for the decentralized caste to coordinate in a given election would be with the help of a caste association. Rudolph and Rudolph (1960) note that caste associations did play a role in Indian electoral politics immediately following independence. However, relatively few Indians are members of formal religious or caste associations. For the caste to vote as a block, other coordinating mechanisms are needed.

A common feature of clientelist politics is the role played by brokers or middlemen who link voting blocks to parties. Srinivas (1960) describes how higher caste landowners and moneylenders served this function in the Mysore village that he studied. In other areas, representatives of the caste itself served this vital function; this is similar to the role played by foremen in linking workers from their caste to firms in the urban economy. The success of the Congress Party in the decades after independence has been attributed to its ability to establish a network of such middlemen (Kothari 1970, Kohli 1990). Although most Congress leaders came from the upper castes, the middlemen generated popular support for the Party that was much broader than its leadership (Brass 1990).

Although the Congress Party may have been successful in organizing caste-based voting banks, many castes were left out. Competition within the Party for scarce resources also resulted in some castes benefiting more than others. According to Yadav, the second electoral system, starting in

1971, was characterized by the entry of middle-caste (OBC) politicians and the end of the complete domination of national politics by the Congress Party. As politics became more competitive, there was a consolidation of vote banks into state-wide multi-*jati* alliances, such as KHAM in Gujarat and AJGAR in Uttar Pradesh.²⁵ Nevertheless, the basic strategy of creating a multi-caste coalition of voters at the local level by delivering public resources to them once elected continued to be followed by all political parties (Chhibber 1999).

This changed in the late 1980's with the onset of Yadav's third electoral system, characterized by the emergence of parties organized explicitly on (broad) caste lines. As discussed above, caste has always played an important role in Indian politics. But while the *jati* in its local setting constituted the voting block in the first electoral system and statewide alliances of *jatis* served this role in the second electoral system, the third electoral system witnessed the formation of a new caste identity based on *jati* groupings such as upper caste, OBC, or SC.

In Chhibber's view, the event that precipitated this change in the political equilibrium was the Mandal Commission report, which expanded reservation in government jobs and educational institutions to the backward castes (OBCs). The Congress Party was forced to distance itself from the Report because forward, backward, and scheduled castes were part of its coalition. The social polarization created by the Report thus opened the door for other parties to step in and represent the interest of broad caste groupings that were large enough to influence electoral outcomes at the state level. This reconfiguration of party politics explicitly on caste lines is most visible in Uttar Pradesh, India's largest state with a population of 200 million, where the Bharatiya Janata Party (BJP) represented the upper castes, the Samajwadi Party (SP) represented the backward castes, and the Bahujan Samaj Party (BSP) represented the scheduled castes. Similar social cleavages in party politics also emerged in other states such as Gujarat, Maharashtra, and Bihar.

The advantage of clientelist politics is that it allows reciprocal relations between political parties and population groups to be sustained. The well known disadvantage is that there will be an under-provision of public goods in favor of transfers targeted at the elected leader's vote bank. In the Indian variant of clientelist politics, as described above, the emergence of explicitly caste-based parties was accompanied by the formation of a new caste-group identity (Yadav 1999). Banerjee and Pande (2010) examine the (negative) consequences of this identity for the quality of elected leaders. Their basic idea is that if voters care about their leader's group identity, then the party associated with the numerically

²⁵KHAM stands for Kshatriya, Harijan, and Muslim and AJGAR stands for Ahir, Jat, Gurjar, and Rajput.

dominant caste-group in a political constituency enjoys an electoral advantage. Its candidates will win even if they are less competent than their rivals. This implies that in constituencies with a numerically dominant caste-group, an increase in caste-group identity will result in a decline in leadership quality if the elected candidate belongs to the party associated with that caste-group. Conversely, if the elected candidate belongs to a party associated with the minority group in that constituency, he will be of relatively high quality.

Banerjee and Pande test this prediction with data from 100 constituencies in Uttar Pradesh at two points in time; 1980 and 1996. Caste identity is believed to have increased substantially over this period in Uttar Pradesh with the emergence of parties such as the SP, favoring the backward castes, and the BSP, favoring the scheduled castes. This implies that in a constituency dominated by low castes, elected politicians representing upper caste parties should have relatively high quality in 1996. Conversely, in constituencies dominated by upper castes, elected politicians from the same upper caste parties should be of relatively low quality. Politician quality is based on measures of corruption collected retrospectively from key informants (journalists and former politicians) in each constituency. The assumption here is presumably that more corrupt politicians capture more resources for themselves at the expense of the electorate.

The first result is that the probability that a constituency dominated by the low castes has a leader from an upper caste party; i.e. the Congress or the BJP, fell by 38% from 1980 to 1996. The second result is that those leaders are substantially less likely to be corrupt in 1996 versus 1980, whereas the converse is true in constituencies dominated by the upper castes (where they are more likely to be corrupt in 1996). An additional prediction of Banerjee and Pande's model is that there should be no change in the quality of elected politicians over time in competitive constituencies where no caste-group is dominant. This implication, which is especially useful in ruling out alternative explanation for the differential change in leadership quality across constituencies, is also borne out in the data.²⁶

7.2 Affirmative Action

The change in Indian party politics over time would have served the interest of particular caste-groups that succeeded in mobilizing politically. However, other caste-groups could have been left worse off. Apart from these distributional effects via the competitive democratic process, the reservation for

²⁶Banerjee and Pande's model also has implications for the difference in quality of the winning candidate and the runner-up that are borne out in the data.

SC's and ST's in state and national legislatures could also have shifted public resources towards these historically disadvantaged caste-groups. In parallel with affirmative action in higher education and central government jobs, the Indian constitution also requires that reservation for SC's and ST's in the national legislature reflect their share of the population in each state. If political parties can credibly commit to particular policies regardless of the preferences of their elected representatives, as assumed in the discussion on clientelist politics above, then reservation of seats in specific constituencies (13% on average for the SC's and 7% on average for the ST's) will have no effect. However, as Pande (2003) notes, if political parties cannot exercise complete control over their representatives once they are elected and if these representatives can influence the allocation of resources by the legislature, then reservation will improve outcomes for the beneficiaries of affirmative action.

Pande examines the effect of caste reservation on access to public resources using state level data from 1960 to 1992. Political reservation is measured by the fraction of jurisdictions reserved (separately) for SC's and ST's in a state in a given year. Targeted public resources are measured by (i) the fraction of state government spending devoted to SC and ST welfare programs (this is 6-8% of total spending on average), and (ii) the fraction of state government jobs reserved for SC/ST's. One potential concern when estimating the relationship between targeted government spending and reservation is that these expenditures could have been made in the absence of reservation. A state with a greater share of reserved seats is, after all, a state with a greater share of SC/ST's in the population and they would naturally have a say in resource allocation.

Pande's solution to this identification problem, which is similar to Bagde, Epple, and Taylor's (2016) research design, takes advantage of the fact that the underlying variable upon which the affirmative action is based is not perfectly correlated with reservation. In Bagde, Epple, and Taylor's application, the effective rank in the list of engineering college applicants is positively (but not perfectly) correlated with the absolute rank; i.e. the score on the college entrance exam. The latter variable can thus be used to control for individual ability when estimating the effect of reservation, measured by the effective rank, on college performance. In Pande's application, the slippage between the population share of the SC/ST's and the fraction of reserved seats is generated by the time-lag in changing quotas. These quotas only change in an election year and are based on the preceding census, whereas the population share changes continuously over time.²⁷ The main finding of the analysis is that conditional on the

²⁷The population share, which is not directly measured, is derived by interpolating between census rounds. Because the preceding census is used for this interpolation, the reservation variable, which is constructed from the preceding census alone, provides no additional information about the actual population share.

SC/ST population share, an increase in SC reservation has a strong positive effect on job quotas but not on welfare programs, whereas an increase in ST reservation results in a substantial increase in ST welfare programs (without a significant effect on jobs or SC welfare programs).

In a related study, Banerjee and Somnathan (2007) examine the location of public goods between 1971 and 1991 in approximately 500 parliamentary constituencies in rural India to assess the changing relationship between the caste composition of the constituency and access to public resources. They consider 15 public goods, covering the areas of education, health, water, electricity, and communication, with access to each good measured by the fraction of villages in the constituency where it is available. The 1971 data indicate that an increase in the share of SC/ST's in the constituency is associated with a deficit in almost all public goods. Once the 1971 and 1991 data are pooled, however, the evidence indicates that there has been substantial convergence in access to public goods over time.²⁸ When this convergence is decomposed by caste group, it turns out that the gains are mostly in the SC-dominated areas, while the ST's appear to have been left behind. The convergence documented by Banerjee and Somnathan can be explained, in part, by the political reservation policy and its redistributive effect that Pande finds. The reconfiguration of electoral politics on caste lines over time, coupled with the broad anti-poverty programs that were pushed by the Indian government, could also have played a role in this convergence. The main message from the preceding analyses, complementing the findings of Hnatkovskaya, Lahiri, and Paul in the private sphere, is that there has been substantial equalization in access to public resources over time.

Did the targeting of public resources to specific caste-groups have negative consequences for overall public good provision? Although the counter-factual experiment, without affirmative action, is unavailable, the evidence indicates that overall public good provision has, if anything, increased over time. Banerjee and Somnathan document a dramatic expansion in rural infrastructure between 1971 and 1991. By 1991, primary schools had been established in 75% of all villages and the share of villages with electricity increased from 18% to 70%. There was a nine-fold increase in access to piped water over this period. Although there are many contributory factors, part of this dramatic increase can be directly attributed to targeted spending through affirmative action. Pande (2003) shows that ST (but not SC) reservation significantly increased per capita state government spending and Chin and Prakash (2010) using a similar identification strategy show that ST (but not SC) reservation significantly re-

²⁸This convergence does not arise mechanically because access to public goods was close to complete in advantaged constituencies in 1971. Half the public goods that Banerjee and Somnathan consider were available in less than 5% of Indian villages in 1971 and in less than 10% of villages in 1991.

duced poverty between 1960 and 2000. There are evidently important differences in the effect of SC and ST reservation, and these two population groups have fared very differently in the post-independence period. Further investigation of these differences would have great value for the design of development policy in India.

7.3 Local Politics

At the state and national level, the literature has devoted much attention to the role that caste plays in party politics. At the local level, parties are less relevant; in principle, candidates are expected to stand independently, without party affiliation. Caste matters now through the caste affiliation of the elected representative, and the focus of much of the literature on local governments in India has been to understand the relationship between the characteristics of the leader (including caste and gender) and the supply of public resources.

The 73rd Amendment of the Indian Constitution, passed in 1991, established a three-tier system of local governments or *panchayats* – at the village, block, and district level – with all seats to be filled by direct election. The village *panchayats*, which often cover multiple villages, were divided into 10-15 wards. *Panchayats* were given substantial power and resources, and regular elections for the position of *panchayat* president and for each ward representative have been held every five years in most states. At the *panchayat* level, seats are reserved in each election for SC's, ST's, and OBC's, but now in proportion to their share of the population in the district. Seats are reserved randomly across wards and, for the position of the president, randomly across *panchayats*, from one election to the next in each district. The only restriction is that no seat can be reserved for the same group across consecutive election within a constituency.²⁹

Without the discipline of a party affiliation, and with limited possibility of re-election (given the frequent exogenous turnover due to randomized reservation), candidates in *panchayat* elections cannot credibly commit to implementing policies once elected that diverge from their own preferences. Nevertheless, if individuals care about the welfare of members of their caste-group or if side-transfers can flow within the group, then elected representatives will target public resources to their own group. This is, after all, the rationale of the reservation policy.

Numerous studies have examined the effectiveness of the reservation policy with micro-data from

²⁹One-third of the seats are reserved randomly for women, independently of caste reservation. Gender reservation can thus be ignored in the discussion on caste reservation that follows.

different parts of the country. The major responsibilities of the *panchayat* are to construct and maintain local infrastructure (public buildings, water supply and sanitation, roads) and to identify welfare recipients. Data from the 2006 Rural Economic Development Survey (REDS), which covers all the major states, indicate that although public goods account for the bulk of *panchayat* expenditures, publicly funded transfers to individual households, including programs for households below the poverty line (BPL) and employment schemes, account for 15% of total expenditures. Focusing on the transfers, Besley, Pande, and Rao (2007) find that SC/ST households residing in the southern states of Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu are seven percent more likely to benefit from a welfare scheme when the position of *panchayat* president is reserved for a member of their own caste-group. Expanding the set of public resources to include housing and sanitation, Bardhan, Mookherjee, and Torrado (2010) find that SC/ST reservation for the *panchayat* president's position in the state of Bengal significantly increases access to public resources, not just for the SC/ST's but for the *panchayat* as a whole.

At the level of the village, the caste or *jati*, rather than caste-groupings such as SC, ST, or OBC, is the social unit around which economic and political activity is organized (Brass 1990). This argument has run as a common thread through all the preceding sections of this paper. Thus, while it may make sense from a policy perspective to assess whether political reservation for particular caste-groups benefits the members of those groups on average, the institutional background indicates that targeting will be at a narrower caste level. The 2006 REDS includes retrospective information on the receipt of public transfers by a sample of households residing in the major Indian states over three election terms, as well as detailed information on the caste affiliation of the household and its ward representative. Using these data over multiple election terms, Munshi and Rosenzweig (2016) show that a household is more likely to benefit from programs targeted at economically disadvantaged households when it shifts from being an outsider to an insider; i.e. when its ward representative belongs to its own caste.³⁰ The additional result is that belonging to the representative's caste-group, but not his caste, does not confer the same advantage.

Having established that the caste rather than the caste-group is relevant for targeting, at least at the most local (ward) level, Munshi and Rosenzweig proceed to examine other roles that caste might

³⁰Although most studies on targeting at the local level focus on welfare transfers, Gille (2016) uses the 2006 REDS to show that individuals are more likely to apply for reserved public sector jobs when the *panchayat* president belongs to the same caste. Her interpretation of this result is that the president uses his political connections at the state level to help the applicant from his caste get a job.

play in local governments. The assumption in the preceding sections has been that a high degree of social connectedness within the caste supports economic cooperation. A natural question to ask is whether such cooperation in private economic activity could have been extended to the public sphere. We have already seen some evidence along these lines with the targeting within castes. But the intra-caste cooperation could also be used, in principle, to solve other problems that commonly arise in local governments.

A well known reason for the under-supply of public goods, going back to Olson (1965), is that individual volunteers will not internalize the benefits that other residents of their constituency derive from non-excludable public goods, resulting in a sub-optimal level of effort and under-provision of those goods. This incentive problem is perhaps most acute in local governments, including the *panchayats*, where non-professional politicians must volunteer for important tasks that involve the supply and distribution of substantial public resources. A related problem is that competent individuals with superior outside options in the private economy will be deterred from volunteering to represent the constituency (Ferraz and Finan 2011). The simple model sketched out below shows how intra-group cooperation can be used to partly solve both of these problems.

N individuals belonging to K castes reside in a ward. Each caste k consists of N_k individuals, with the k subscript sorting castes by size; $N_{k-1} < N_k$. Each caste can put forward a candidate to represent it in the ward election. If elected, the unpaid representative's responsibility is to supply non-excludable public goods to the ward, with their level increasing in the effort he exerts. The key assumption of the model is that each caste can credibly commit to compensate its representative *ex post* for his effort, conditional on being elected, even if he holds office for a single term. Given these side-transfers, the individual who is selected to represent the caste in the ward election, and the effort he exerts if elected, will be optimal from the perspective of the caste.

When choosing the level of effort, the representative will take account of the cost of effort, which is decreasing in his ability, and (because of the side-transfers) the benefit derived from public goods by all his co-ethnics residing in the ward. A representative with ability, ω , belonging to group k of size N_k will thus choose effort, e , to maximize

$$N_k e^\beta - \frac{e}{\omega}, \quad (13)$$

where e^β is the level of public goods ($\beta < 1$) and the implicit assumption is that the level of public goods maps one-for-one into the utility derived from their consumption by each resident. The representative's

optimal level of effort can thus be derived as a function of N_k and ω .

Given the level of effort that it expects *ex post*, each caste will choose as its representative the individual who maximizes the total utility derived from the public good by its members, net of the representative's time and effort cost. We assume that utility is transferable and so the compensatory side-transfers from the caste to the representative can be ignored. ω will thus be chosen to maximize

$$N_k [e(\omega, N_k)]^\beta - \frac{e(\omega, N_k)}{\omega} - \alpha\omega, \quad (14)$$

where $\alpha\omega$ measures the opportunity cost to an individual from volunteering as the representative, with this cost increasing in ability ($\alpha > 0$). Substituting the expression for $e(\omega, N_k)$ from the first maximization problem and then maximizing with respect to ability, both ω and e can ultimately be derived as increasing functions of N_k . Intuitively, a representative from a larger group internalizes the utility derived from public goods by a larger number of individuals, resulting in the selection of more competent representatives who exert more effort. Cooperation within the caste results in a higher level of public good provision relative to the benchmark model where the representative only cares about his own utility (N_k would be replaced by one in the maximization problems above). However, public good provision is less than first-best (where N_k would be replaced by N).

This model of public good provision shares an important feature of previous models of insurance, migration, and occupational choice, which is that outcomes are increasing in the size of the relevant group; the caste network in previous applications and the size of the caste in the ward in the current model. We model migration and occupational choice non-cooperatively, in the sense that individuals benefit from their caste networks and take account of these benefits when they make their decisions, but ignore the benefits they confer, in turn, on other members of the caste (by increasing the size of the network). In contrast, the models of mutual insurance and local public good provision allow for side-transfers, which implies that individual decisions are optimal from the perspective of the caste. Whether the caste should be modelled cooperatively or non-cooperatively will, in general, depend on whether or not side-transfers are feasible.

Munshi and Rosenzweig (2016) test the predictions of the cooperative model of public good provision with data from the 2006 REDS. This (latest) round of the REDS includes information on whether new construction or maintenance of specific public goods *actually* took place on each street in the village over the past three election terms. The analysis focuses on six goods – drinking water, sanitation, roads, electricity, street lights, and public telephones – which account for 67.5% of the *panchayat's*

discretionary spending.³¹ The supply of each public good is measured by the fraction of streets in the ward that received that good in a given election term, with overall public good provision measured by the average across all goods. The caste composition of the ward is obtained from the household census conducted in all REDS villages and the caste of the elected ward representative is available for 35% of ward-terms. Using these data, Munshi and Rosenzweig document a strong positive relationship between the numerical size of the representative's caste in the ward and both his education (a reasonable measure of competence in this context, which is correlated with wealth and occupation) and public good provision.

Counter-factual simulations of the estimated model indicate that cooperation within castes substantially increases the supply of public goods at the ward level, relative to the benchmark where the representative only cares about the utility he derives by consuming the public goods. Although the assumption that castes are able to cooperate internally, with transfers flowing from the caste to its representative, follows naturally from our description of caste-based economic activity in many different settings and time periods, it contrasts with the standard view of clientelist politics. In that view, it is the politician who makes transfers to the electorate in return for their votes, resulting in the over-supply of targeted public transfers and an accompanying under-supply of public goods. Whether transfers flow from or to the representative will depend on the context. For example, Anderson, Francois, and Kotwal (2015) study clientelist politics in rural Maharashtra, where a single caste, the Marathas, unusually dominates both the population and landownership. Although *panchayat* elections in this region are competitive and fair, leadership is controlled by landowning Maratha elites. Closer inspection of the economic and social relations within these communities reveals that Maratha landowners provide insurance against income shocks for economically disadvantaged members of their caste in return for their votes. Political power allows them to allocate government resources to public goods that they (but not the broad electorate) prefer.

The reciprocal relationship described above is not necessarily inefficient. When there is heterogeneity in the preferences for public goods, side-transfers allow sub-groups that do not receive their preferred public goods to be suitably compensated. The inefficiency that arises in Anderson, Francois, and Kotwal's setting is that when government safety nets, such as employment guarantee schemes, become

³¹The assumption is that while the supply of these goods is endogenously determined by the ward representative's effort, they cannot be targeted more finely within the ward; i.e. at the street level, so as to favor the representative's own caste. Munshi and Rosenzweig provide empirical support for this assumption. Note that education and health are not included in the list of public goods because they fall outside the purview of the *panchayats*.

available, the political elites actively discourage the implementation of these programs to maintain their bargaining power with the electorate and to keep wages low. Anderson, Francois, and Kotwal's model generates additional predictions for the relationship between clientilism and both the fraction of Marathas in the *panchayat* as well as their share of landownership, which are borne out in the data they collect from 320 villages.

While the preceding characterization of local caste politics would appear to be at odds with Munshi and Rosenzweig's description of intra-caste cooperation and its positive effect on public goods provision, the context is somewhat unusual. A single caste will rarely dominate both the population and landownership, as the Marathas do in Maharashtra, and even when they do, caste could affect the supply of public resources through other channels. If a single caste dominates local politics, then it will continue to do so across election terms, with different reservation schemes, by installing proxy representatives. Munshi and Rosenzweig (2016) allow for this possibility by including ward fixed effects in their analysis, which covers all the major Indian states. Net of ward fixed effects, they continue to find that the size of the representative's caste has a strong and positive effect on the representative's education and the level of public good provision.³²

Munshi and Rosenzweig's analysis provides a new positive perspective on local caste politics. At the same time, it uncovers unintended efficiency and equity consequences of caste-based reservation. Caste reservation mechanically lowers the size of the representative's caste by restricting the set of castes that are eligible to stand in a given election term. This size-effect reduces the elected representative's competence and public good provision. Munshi and Rosenzweig's estimates indicate that rolling back the reservation system at the ward level would substantially increase the supply of public goods. However, this policy would have adverse distributional consequences within the ward. As noted, ward representatives target welfare transfers to their own caste, with the crowding-out of outsiders increasing in the size of the representative's group. A potential solution to the resulting inequity would be to remove the distribution of welfare transfers from the representative's list of responsibilities. But even if that policy solution were feasible, de-reservation could still have adverse distribution consequences across wards, to

³²Which caste's representative gets elected is an outcome of the competitive democratic process. Munshi and Rosenzweig take a step back to model the relationship between the size of the largest eligible caste in the ward, which changes exogenously across election terms with reservation, and the size of the elected representative's caste. The main result is that there will be a discontinuous increase in the size of the representative's caste, with an accompanying increase in his competence and the supply of public goods, when the population share of the largest caste crosses a threshold. Net of ward fixed effects, which subsume the time-invariant demand for public goods by the electorate and the political control exerted by a dominant caste if it is active, there is indeed a discrete increase in these outcomes at a threshold.

the extent that the supply of public goods is fixed at the *panchayat* level, by persistently favoring wards with larger castes.³³ Caste-based affirmative action will, in general, have unintended consequences in different economic spheres and at different levels of government that must be examined and quantified to objectively assess this important policy.

8 Conclusion

The caste system is associated with terrible prejudice and exploitation and is, for many Indians, their collective shame. While we may want to wish it away, this system continues to exist in modern India. Both from a research and from a policy perspective, it is therefore important to understand the causes of this persistence, some of which go beyond prejudice. For example, we have seen that caste networks continue to play a role across a wide range of sectors in the Indian economy. Affirmative action in higher education and in India's representative democracy has, if anything, made caste more salient in society and in the public discourse. Given these trends, what is the prognosis for the future?

Newly available evidence with nationally representative data indicates that there has been convergence in education, income, occupations, and consumption across caste groups over time. If this trend continues, then there will no longer be a rationale for affirmative action. Caste-based reservation is now highly politicized and there will always be special interest groups that stand to gain by maintaining the status quo. At the same time, there are systems in place that allow for constitutional amendments, and the expectation is that affirmative action will ultimately be rolled back once it is no longer necessary. It is important to realize, however, that such policy changes will not occur in the near future. The current gap in educational attainment between SC/ST's and upper castes is very wide; over two years of schooling. Given the strong intergenerational persistence in human capital, the key variable driving convergence, it will be many generations before income and consumption are equalized across caste groups.

The caste-based economic networks, which currently serve many functions, will also ultimately disappear. These networks continue to be active in the globalizing Indian economy because information and commitment problems are exacerbated during a period of economic change. In the long-run, however, the markets will settle into place and the community networks will lose their purpose. This has certainly been the experience in many developed countries. In the United States, for example,

³³This is less of a concern if more competent representatives are able to bring back more resources to the *panchayat* from the center or use existing resources more effectively.

ethnic networks based on a European country (region) of origin supported their members through the nineteenth century into the middle of the twentieth century. Ultimately, however, these networks no longer served a useful role and today, outside of a few pockets, European ethnic identity in the United States is largely symbolic. We might expect caste to similarly lose its salience as India develops into a modern market economy, and there is some evidence that this process may have already begun.

Munshi and Rosenzweig (2006) found that lower caste boys in Bombay were held back by their working class networks, whereas lower caste girls (without that historical baggage) were apparently willing and able to take advantage of the new opportunities that became available with economic liberalization in the 1990's. This is reflected not only in their schooling choices, which will presumably map into their labor market outcomes, but also in their marriage decisions; specifically, whether they married within their caste. Figure 3 reports the marriage choices of female family members of the students who were sampled by Munshi and Rosenzweig. The marriage choices are reported in five-year intervals over a time period (1970-2000) that substantially overlaps with Figure 1, which documented corresponding marriage patterns in rural India in ten-year intervals from 1950 to 2000. While out-marriage remained extremely stable at 5% over the entire 50-year period in rural India, out-marriage increases over time in the urban data. It is even lower than 5%, somewhere between 2% and 4%, prior to economic liberalization, but then starts to increase steeply in the 1990's and is as high as 12% in the 1995-2000 period. Munshi and Rosenzweig's study is set in Bombay city, which is at the vanguard of economic development in India. These results are thus almost certainly unrepresentative, but they do tell us what we might expect to see in the future.³⁴ Marriage within the caste is key to the efficient functioning of caste-based networks and the formation of a caste identity. The changes in marriage patterns uncovered in Bombay will be replicated elsewhere in the future as the economy develops and this will signal the end of the caste system as we know it.

These social changes, just like the end of affirmative action, will take many decades (even generations) to unfold. In the interim, both research and policy would benefit from a recognition of the informal caste-based institutions that are active in the economy. A counter-factual policy experiment described in Munshi and Rosenzweig (2016) highlights the potential cost of ignoring these institutions. Recall that Munshi and Rosenzweig's explanation for India's low mobility is based on a combination

³⁴Similar results are obtained by Banerjee et al. (2013) who examine the marriage advertisements that were placed in an English language newspaper in Calcutta and the matches that subsequently took place. They find that 72% of the prospective grooms and 68% of the prospective brides who had married after a year found a partner within their caste. These statistics are substantially lower than the standard figure of 95% obtained from nationally representative surveys.

of well-functioning (caste-based) rural insurance networks and the absence of formal insurance, which includes government safety nets and private credit. Having empirically validated this explanation and having estimated the model of insurance and migration, they proceed to implement a counter-factual experiment in which private credit is made available to wealthy households (with the resources to put up collateral). We saw that these households subsidize less wealthy members of their caste and, thus, are more likely to leave in the current equilibrium. While the absence of alternative consumption-smoothing mechanisms keeps their migration currently at a low level, this level increases dramatically in the counter-factual experiment when they get access to credit and are no longer reliant on the rural insurance network.

The increased migration by wealthy households leaves less wealthy households in their caste worse off in two ways. First, the size and scope of the insurance network is reduced by the exit of the wealthy households, making it function less effectively and increasing the consumption-variance of the households that remain. Second, the level of redistribution will decline as a way of getting some wealthy households to stay in the network, reducing the consumption level of the poor households. A well intentioned policy that succeeded in its objective of increasing migration could thus have unintended distributional consequences that would go unnoticed if researchers evaluating the programme and the policymakers who designed it were unaware of the underlying links within the caste. There is an increasing awareness that spillovers must be accounted for when designing and evaluating development programs. In the Indian context, these spillovers will often extend beyond the village to cover the entire caste, spread over a wide area.

References

- [1] Akerlof, George. 1976. The Economics of Caste and of the Rat Race and Other Woeful Tales, *Quarterly Journal of Economics* 90(4): 599-617.
- [2] Alexander, K.C. 1989. Caste Mobilization and Class Consciousness: The Emergence of Agrarian Movements in Kerala and Tamil Nadu, in: F.R. Frankel and M.S.A. Rao (eds), *Dominance and State Power in Modern India; Decline of a Social Order*, vol.1, Delhi: Oxford University Press, p. 362-413.
- [3] Anderson, Siwan. 2011. Caste as an Impediment to Trade, *American Economic Journal: Applied* 3(1):239-263.
- [4] Anderson, Siwan, Patrick Francois, and Ashok Kotwal. 2015. Clientilism in Indian Villages, *American Economic Review* 105(6): 1780-1816.
- [5] Angelucci, Manuela, Giacomo De Giorgi, and Imran Rasul. 2015. Resource Pooling within Family Networks: Insurance and Investment, University of Michigan, typescript.
- [6] Bagde, Surendrakumar, Dennis Epple, and Lowell Taylor. 2016. Does Affirmative Action Work? Caste, Gender, College Quality, and Academic Success in India, *American Economic Review* 106(6): 1495-1521.
- [7] Banerjee, Abhijit, Marianne Bertrand, Saugato Datta, and Sendhil Mullainathan. 2009. Labor market discrimination in Delhi: Evidence from a field experiment, *Journal of Comparative Economics* 37: 14-27.
- [8] Banerjee, Abhijit, Esther Duflo, Maitreesh Ghatak, and Jeanne Lafortune. 2013. Marry for What? Caste and Mate Selection in Modern India, *American Economic Journal: Microeconomics* 5(2): 33-72.
- [9] Banerjee, Abhijit and Kaivan Munshi. 2004. How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur, *Review of Economic Studies* 71(1):19-42.
- [10] Banerjee, Abhijit and Rohini Pande. 2010. Parochial Politics: Ethnic Preferences and Politician Corruption, mimeo, Harvard University, Kennedy School of Government.

- [11] Banerjee, Abhijit and Rohini Somnathan. 2007. The political economy of public goods: some evidence from India, *Journal of Development Economics* 82(2): 287-314.
- [12] Banerjee, Biswajit and J.B. Knight. 1985. Caste Discrimination in the Indian Urban Labour Market, *Journal of Development Economics* 17: 277-307.
- [13] Bardhan, Pranab, Dilip Mookherjee, and Monica Parra Torrado. 2010. Impact of Political Reservations in West Bengal Local Governments on Anti-Poverty Targeting, *Journal of Globalization and Development* 1(1): 1-36.
- [14] Bayly, Susan. 2001. *Caste, Society and Politics in India from the Eighteenth Century to the Modern Age*. Cambridge: Cambridge University Press.
- [15] Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan. 2010. Affirmative action in education: Evidence from engineering college admissions, *Journal of Public Economics* 94: 16-29.
- [16] Bertrand, Marianne and Sendhil Mullainathan. 2004. Are Emily and Greg more employable than Lakisha or Jamal, *American Economic Review* 94(4): 991-1013.
- [17] Besley, Timothy, Rohini Pande, and Vijayendra Rao. 2007. Political Economy of Panchayats in South India, *Economic and Political Weekly* 42(8): 661-667.
- [18] Bidner, Chris and Mukesh Eswaran. 2012. A Gender-Based Theory of the Origin of the Caste System in India, mimeo, University of British Columbia.
- [19] Bisin, Alberto and Thierry Verdier. 2000. "Beyond the Melting Pot": Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits, *Quarterly Journal of Economics* 115(August): 955-988.
- [20] Brass, Paul. *The politics of India Since Independence*. Cambridge: Cambridge University Press.
- [21] Burnett-Hurst, A.R. 1925. *Labour and Housing in Bombay: A Study in the Economic Conditions of the Wage-Earning Classes of Bombay*. London: P.S. King and Son.
- [22] Caldwell, John C., P.H. Reddy, Pat Caldwell. 1986. Periodic High Risk as a Cause of Fertility Decline in a Changing Rural Environment: Survival Strategies in the 1980-1983 South Indian Drought, *Economic Development and Cultural Change* pp. 677-701.

- [23] Cassan, Guilhem. 2015. Identity-Based Policies and Identity Manipulation: Evidence from Colonial Punjab, *American Economic Journal: Economic Policy*, 7(4): 103131.
- [24] Cassan, Guilhem. 2016. Affirmative Action, Education and Gender: Evidence from India, mimeo, University of Namur.
- [25] Chakravarti, Uma. 1993. Conceptualizing Brahmanical Patriarchy in Early India: Gender, Caste, Class and State, *Economic and Political Weekly*, 3 April.
- [26] Chandavarkar, Rajnarayan. 1994. *The Origins of Industrial Capitalism in India: Business strategies and the working classes in Bombay, 1900-1940*. Cambridge: Cambridge University Press.
- [27] Chhibber, Pradeep K. 2002. *Democracy without Associations: Transformation of the Party System and Social Cleavages in India*. Ann Arbor: University of Michigan Press.
- [28] Chin, Aimee and Nishith Prakash. 2011. The redistributive effects of political reservation for minorities: Evidence from India, *Journal of Development Economics* 96: 265-277.
- [29] Cholia, R.P. 1941. *Dock Labourers in Bombay*. Bombay.
- [30] Choy, James P. 2016. Social Divisions with Endogenous Hierarchy, mimeo, University of Warwick.
- [31] Coleman, James S. 1988. Social Capital in the Creation of Human Capital, *American Journal of Sociology* 94: S95-S120.
- [32] Damodaran, Harish. 2008. *India's New Capitalists: Caste, Business, and Industry in Modern India*. Ranikhet: Permanent Black.
- [33] Dandekar, Hemalata C. 1986. *Men to Bombay, Women at Home: Urban Influence on Sugao Village, Deccan Maharashtra, India, 1942-1982*. Ann Arbor, Michigan.
- [34] Darity, William Jr., Ashwini Deshpande, and Thomas Weisskopf. 2011. Who is Eligible? Should Affirmative Action be Group- or Class-Based? *American Journal of Economics and Sociology* 70(1): 238-268.
- [35] Das Gupta, Ranajit. 1981. Structure of the Labour Market in Colonial India, *Economic and Political Weekly* 16(44/46): 1781-1806.

- [36] De Haan, Arjan. 1997. Unsettled Settlers: Migrant Workers and Industrial Capitalism in Calcutta, *Modern Asian Studies* 31(4): 919-949.
- [37] De Zwart, Frank. 2000. The logic of affirmative action: Caste, class and quotas in India, *Acta Sociologica* 43(3): 235-249.
- [38] Deshpande, Ashwini and Katherine Newman. 2007. Where the Path Leads: The Role of Caste in Post-University Employment Expectations, *Economic and Political Weekly* 42(41):4133-4140.
- [39] Deshpande, Ashwini and Thomas E. Weisskopf. 2011. Does Affirmative Action Affect Productivity in the Indian Railways? Centre for Development Studies, Department of Economics, Delhi School of Economics Working Paper No. 185.
- [40] Dhillon, Amrita, Vegard Iversen, and Gaute Torsvik. 2013. Employee Referral, Social Proximity and Worker Discipline: Theory and Evidence from India, CESifo Working Paper No. 4309.
- [41] Dirks, Nicholas B. 2001. *Castes of Mind: Colonialism and the Making of Modern India*. Princeton: Princeton University Press.
- [42] Drèze, Jean and Geeta Gandhi Kingdon. 2001. School Participation in Rural India, *Review of Development Economics* 5(1): 1-24.
- [43] Dumont, Louis. 1970. *Homo Hierarchicus: The Caste System and Its Implications*. Chicago: University of Chicago Press.
- [44] Engelshoven, Miranda. 2002. Rural to Urban Migration and the Significance of Caste: The Case of the Saurashtra Patels of Surat, in *Development and Deprivation in Gujarat*, eds. Ghanshyam Shah, Mario Rutten, and Hein Streefkerk, New Delhi: Sage Publications.
- [45] Fafchamps, Marcel and Susan Lund. 2003. Risk Sharing Networks in Rural Philippines, *Journal of Development Economics* 71: 261-287.
- [46] Ferraz, Claudio and Frederico Finan. 2011. Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance, mimeo, University of California at Berkeley.
- [47] Field, Erica, Seema Jayachandran, and Rohini Pande. 2010. Do Traditional Institutions Constrain Female Entrepreneurship? A Field Experiment on Business Training in India, *American Economic Review Papers and Proceedings* 100(2): 125-129.

- [48] Frisancho Robles, Verónica C. and Kala Krishna. 2012. Affirmative Action in Higher Education in India: Targeting, Catch Up, and Mismatch, National Bureau of Economic Research Working Paper 17727.
- [49] Gadgil, D.R. 1959. *Origins of the Modern Indian Business Class*, Institute of Pacific Relations, New York.
- [50] Gans, Herbert. 1962. *The Urban Villagers: Group and Class in the Life of Italian-Americans*. New York: The Free Press of Glencoe.
- [51] Ghurye, Govind S. 1932. *Caste and Race in India*. London: Routledge and Kegan Paul.
- [52] Gille, Véronique. 2016. The role of local politics and caste networks in the take-up of affirmative action in India, mimeo, DIAL.
- [53] Gokhale, R.G. 1957. *The Bombay Cotton Mill Worker*. Bombay.
- [54] Gore, M.S. 1970. *Immigrants and Neighborhoods: Two Aspects of Life in a Metropolitan City*. Tata Institute of Social Sciences, Bombay.
- [55] Greif, Avner. 1993. Contract Enforceability and Economic Institutions in Early Trade: the Maghribi Traders' Coalition, *American Economic Review* 83(3):525-48.
- [56] Grimard, Franque. 1997. Household Consumption Smoothing through Ethnic Ties: Evidence from Cote d'Ivoire, *Journal of Development Economics* 53(3): 319-422.
- [57] Gupta, Bishnupriya. 2014. Discrimination or Social Networks? Industrial Investment in Colonial India, *Journal of Economic History* 74(1): 141-168.
- [58] Gupta, Dipankar. 2000. *Interrogating Caste: Understanding Hierarchy and Difference in Indian Society*. Penguin Books, India.
- [59] Hanna, Rema N. and Leigh L. Linden. 2012. Discrimination in Grading, *American Economic Journal: Economic Policy* 4(4): 146-168.
- [60] Hnatkovska, Viktoria, Amartya Lahiri, and Sourabh Paul. 2012. Castes and Labor Mobility, *American Economic Journal: Applied Economics* 4(2): 274-307.

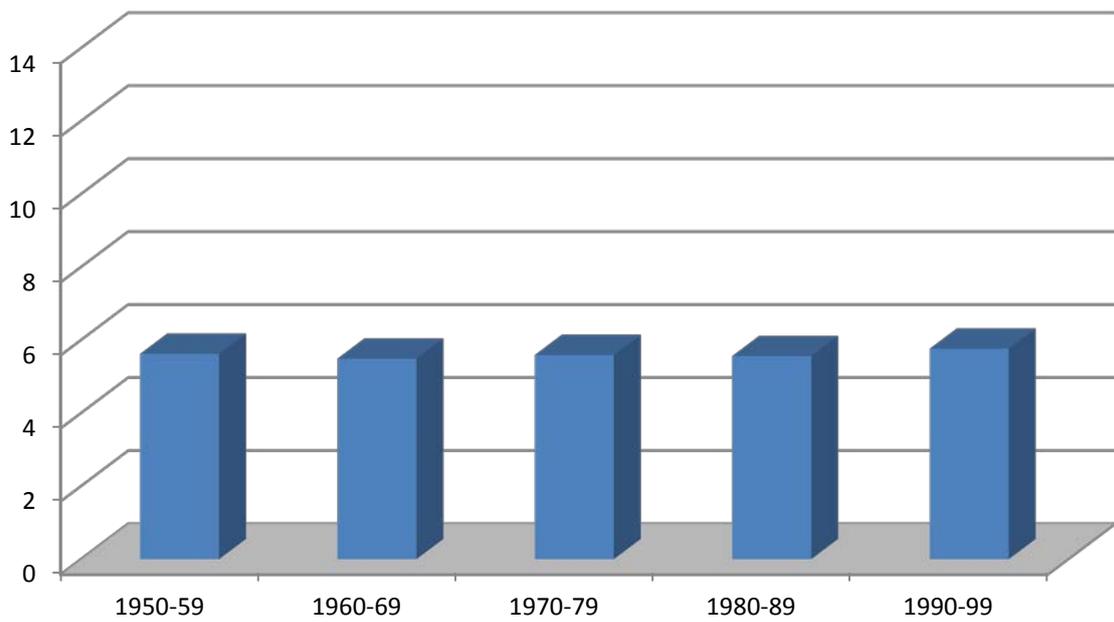
- [61] Hnatkowska, Viktoria, Amartya Lahiri, and Sourabh Paul. 2013. Breaking the Caste Barrier: Intergenerational Mobility in India, *Journal of Human Resources* 48(2): 435-473.
- [62] Hoff, Karla and Priyanka Pandey. 2006. Discrimination, Social Identity, and Durable Inequalities, *American Economic Review* 96(2): 206-211.
- [63] Jodhka, Surinder S. and Katherine Newman. 2007. In the Name of Globalization: Meritocracy, Productivity and the Hidden Language of Caste, *Economic and Political Weekly* 42(41): 4125-4132.
- [64] Kapadia, Karin. 1995. *Siva and Her Sisters; Gender, Caste, and Class in Rural South India*. Boulder: Westview Press.
- [65] Kijima, Yoko. 2006. Caste and Tribe Inequality: Evidence from India, 1983-1999, *Economic Development and Cultural Change* 54(2): 369-404.
- [66] Kingdon, Geeta Gandhi. 1998. Education of Females in India: Determinants and Consequences: A Case Study of Urban Uttar Pradesh, mimeo, McNamara Fellowships Office, Economic Development Institute of the World Bank, Washington D.C.
- [67] Kochar, Anjini. 2008. Can schooling policies affect schooling inequality? An empirical evaluation of school location policies in India, in eds S. Bery, B. Bosworth and A. Panagariya, *India Policy Forum 2007/8* 4: 53-59, Sage Publications Inc.
- [68] Kochar, Anjini. 2010. Affirmative Action Through Quotas: The Effect on Learning in India, Stanford Center for International Development Working Paper 426.
- [69] Kohli, Atul. 1990. *Democracy and Discontent: India's Growing Crisis of Governability*. Cambridge: Cambridge University Press.
- [70] Kooiman, Dick. 1989. *Conversion and Social Equality in India; the London Missionary Society in South-Travancore in the 19th Century*, Delhi: Manohar Press.
- [71] Kornblum, William. 1974. *Blue Collar Community*. Chicago and London: University of Chicago Press.
- [72] Kothari, Rajni. 1970. *Politics in India*. Little, Brown.

- [73] Lamb, Helen B. 1955. The Indian Business Communities and the Evolution of an Industrialist Class, *Pacific Affairs*, 28(2):101-116.
- [74] Luke, Nancy and Kaivan Munshi. 2011. Women as Agents of Change: Female Income and Mobility in India, *Journal of Development Economics* 94(1):1-17.
- [75] Madheswaran, S. and Paul Attewell. 2007. Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey, *Economic and Political Weekly* 42(41): 4146-4153.
- [76] Mazzocco, Maurizio and Shiv Saini. 2012. Testing Efficient Risk-Sharing with Heterogeneous Risk Preferences, *American Economic Review* 102(1): 428-468.
- [77] Mehta, Makrand. 2001. Gujarati Business Communities in East African Diaspora: Major Historical Trends, *Economic and Political Weekly* 36(20): 1738-1747.
- [78] Min, Pyong Gap. 2001. Changes in Korean Immigrants' Gender Role and Social Status, and Their Marital Conflicts, *Sociological Forum* 16(2): 301-320.
- [79] Moorjani, Priya, Kumarasamy Thangaraj, Nick Patterson, Mark Lipson, Po-Ru Loh, Periyasamy Govindaraj, Bonnie Berger, David Reich, and Laiji Singh. 2013. "Genetic Evidence for Recent Population Mixture in India," *American Journal of Human Genetics* 93(3):422-438.
- [80] Morris, Morris David. 1965. *The Emergence of an Industrial Labor Force in India: A Study of the Bombay Cotton Mills, 1854-1947*. Berkeley: University of California Press.
- [81] Morten, Melanie. 2012. Temporary Migration and Endogenous Risk Sharing in Village India, Stanford University typescript.
- [82] Munshi, Kaivan. 2011. Strength in Numbers: Networks as a Solution to Occupational Traps, *Review of Economic Studies* 78:1069-1101.
- [83] Munshi, Kaivan and Mark Rosenzweig. 2006. Traditional Institutions Meet the Modern World: Caste, Gender and Schooling Choice in a Globalizing Economy, *American Economic Review*, 96(4):1225-1252.
- [84] Munshi, Kaivan and Mark Rosenzweig. 2016. Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap, *American Economic Review* 106(1): 46-98.

- [85] Nafziger, Wayne. 1978. *Class, Caste, and Entrepreneurship: A Study of Indian Industrialists*. Honolulu: The University Press of Hawaii.
- [86] Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*, Cambridge, Massachusetts: Harvard University Press.
- [87] Pande, Rohini. 2003. Can mandated political representation increase policy influence for disadvantaged minorities? Theory and evidence from India, *American Economic Review* 93(4): 1132-1151.
- [88] Rawal, Shenila and Geeta Kingdon. 2010. Akin to my teacher: Does caste, religious, or gender distance between student and teacher matter? Some evidence from India, Department of Quantitative Social Science, Institute of Education, University of London Working Paper No. 10-18.
- [89] Rees, Albert. 1966. Information Networks in Labor Markets, *American Economic Review (Papers and Proceedings)* 56: 559-66.
- [90] Richman, Barak. 2006. How Community Institutions Create Economic Advantage: Jewish Diamond Merchants in New York, *Law and Social Inquiry* 31(2): 383-420.
- [91] Rudner, David West. 1994. *Caste and Capitalism in Colonial India: The Nattukottai Chettiars*, Berkeley: University of California Press.
- [92] Rudolph, Lloyd I. and Susanne H. Rudolph. 1960. The political role of India's caste associations, *Pacific Affairs* 33: 5-22.
- [93] Schuler, Sidney Ruth, Syed M. Hashemi, Ann P. Riley, and Shireen Akhter. 1996. Credit Programs, Patriarchy and Men's Violence Against Women in Rural Bangladesh, *Social Science and Medicine* 43(12):1729-1742.
- [94] Srinivas, Mysore N. 1960. *The Social Structure of a Mysore Village*. Bobbs-Merrill.
- [95] Swaminathan, Padmini and J. Jeyaranjan. 1994. The Knitwear Cluster in Tiruppur: An Indian Industrial District in the Making? Madras Institute of Development Studies Working Paper No. 126.
- [96] Tabellini, Guido. 2008. The Scope of Cooperation: Values and Incentives, *Quarterly Journal of Economics* 123(3): 905-950.

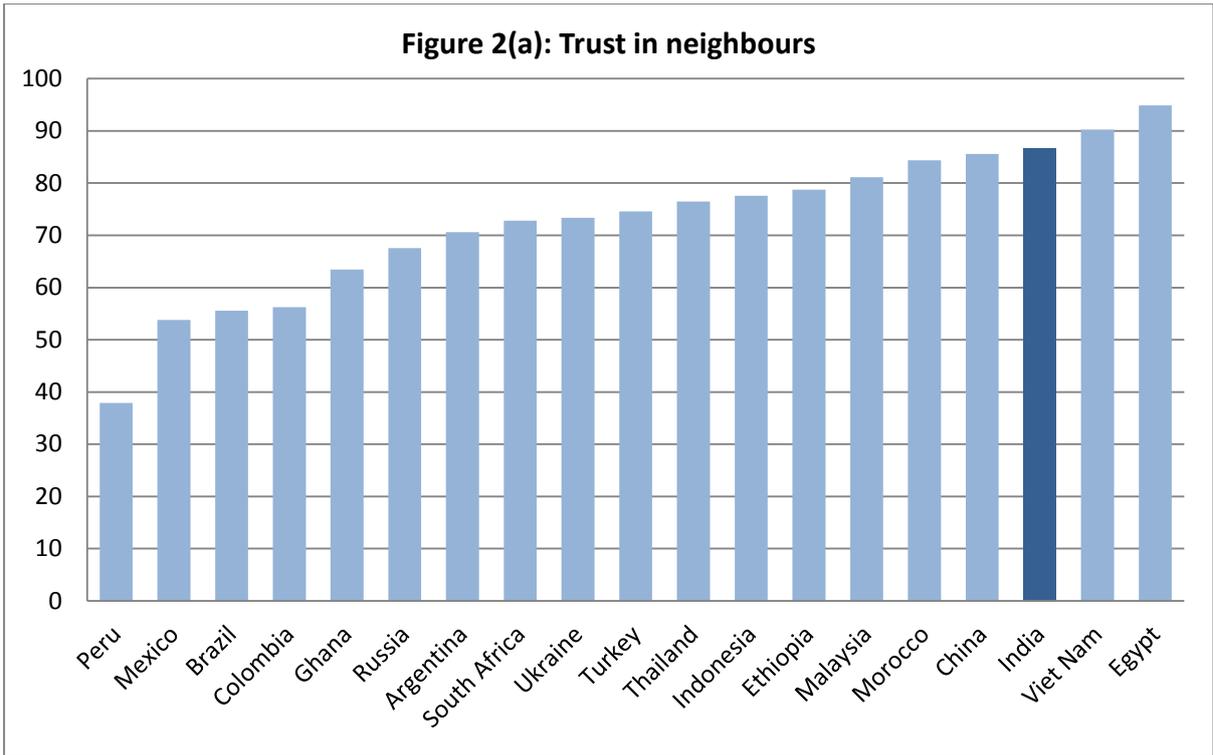
- [97] Thorat, Sukhadeo and Paul Attewell. 2007. The Legacy of Social Exclusion: A Correspondence Study of Job Discrimination in India, *Economic and Political Weekly* 42(41): 4141-4145.
- [98] Timberg, Thomas. 1978. *The Marwaris: From Traders to Industrialists*, New Delhi: Vikas Publishing House Pvt. Ltd.
- [99] Townsend, Robert. 1994. Risk and Insurance in Village India, *Econometrica* 62(3): 171-84.
- [100] Yadav, Yogendra. 1999. Electoral Politics in the Time of Change: India's Third Electoral System, 1989-99, *Economic and Political Weekly* 34(34/35): 2393-2399.

Figure 1: Marriage outside the caste in rural India

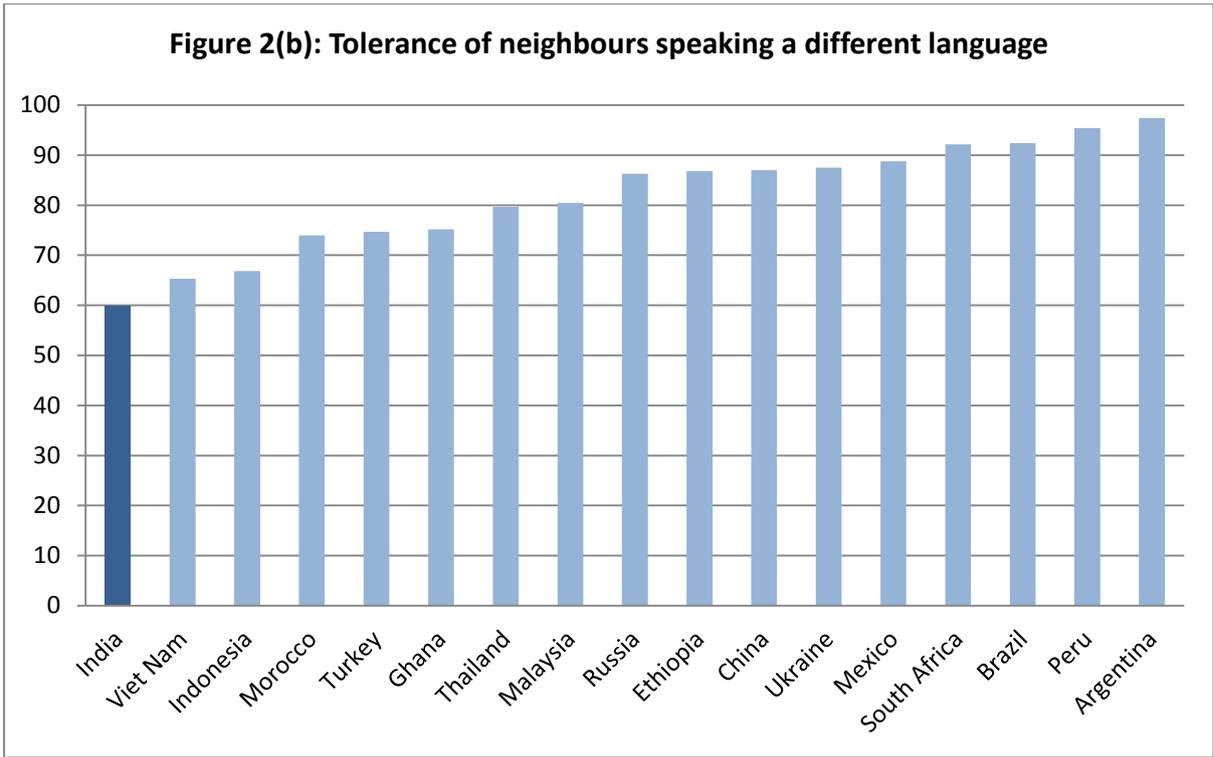


Source: Rural Economic Development Survey, 1999.

Percentage of marriages by women in the household head's family (sisters and daughters) outside their caste.

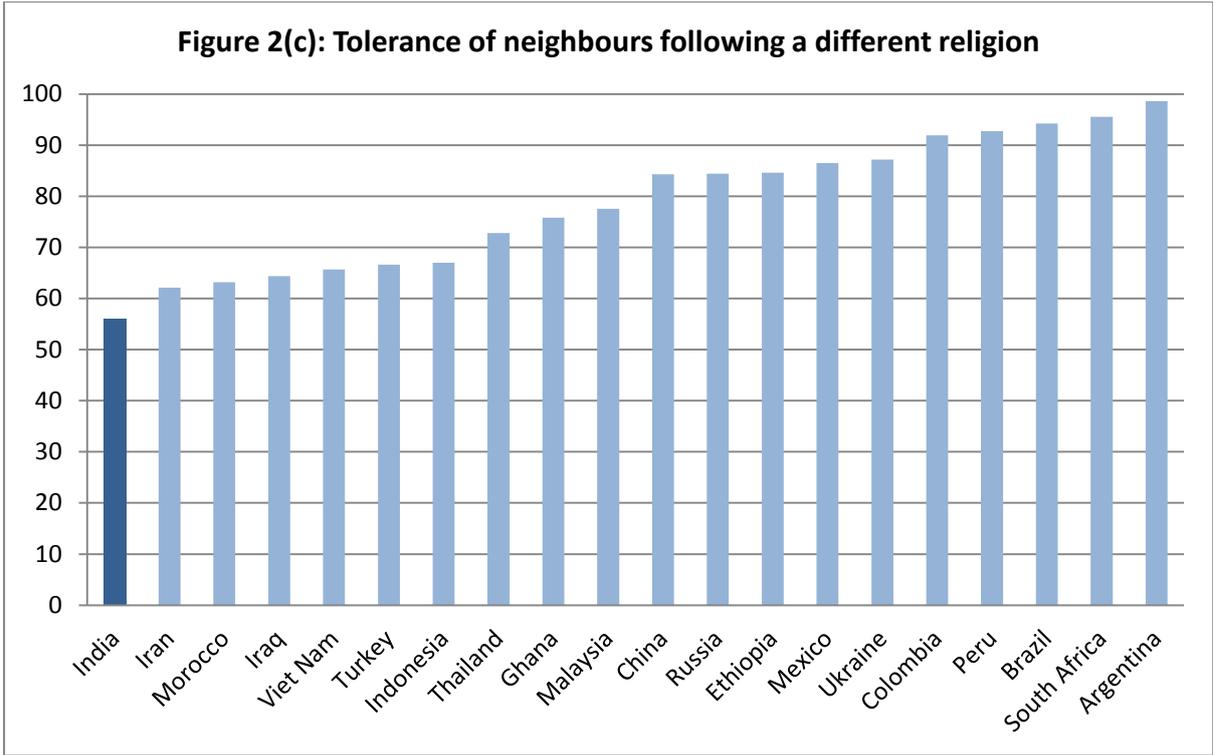


Source: World Values Survey, Wave 5.
 Percentage of respondents who trust people in their neighbourhood.



Source: World Values Survey, Wave 5.

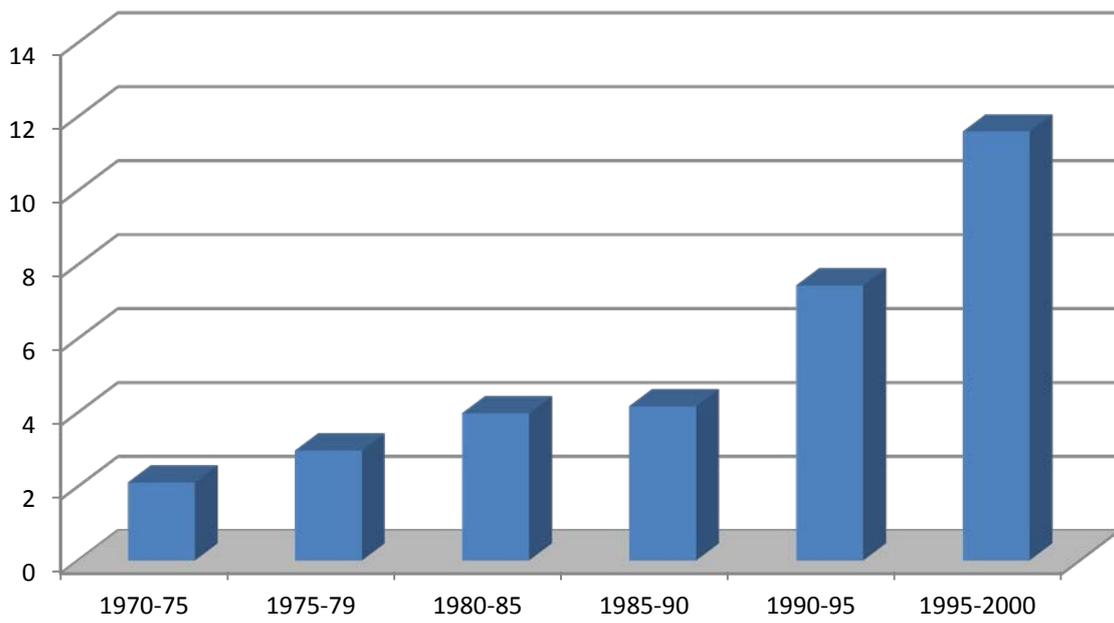
Percentage of respondents who do not have a problem with their neighbours speaking a different language.



Source: World Values Survey, Wave 5.

Percentage of respondents who do not have a problem with their neighbours following a different religion.

Figure 3: Marriage outside the caste in urban India



Source: Survey of school children and their families in Bombay, 2001.

Percentage of marriages by women in the student's family (sisters and mothers) outside their caste.