Social Division with Endogenous Hierarchy*

James P. Choy

January 25, 2017

Abstract

Many societies are divided into multiple smaller groups. Certain kinds of interaction are more likely to take place within a group than across groups. I model a reputation effect that enforces these divisions. People who are observed to interact with members of different groups are believed to be less trustworthy by members of their own group. A hierarchical relationship between groups appears endogenously in equilibrium. The information requirements for my equilibrium to exist are much less demanding than the information requirements in related models. These different information requirements correspond to concrete differences between the institutions of different Indian castes.

Keywords: Cooperation, Caste, Social Institution
JEL Classification Numbers: C73, O12, O17

*Corresponding author: James P. Choy, Department of Economics, University of Warwick, Coventry CV4 7AL, United Kingdom. E-mail: j.choy@warwick.ac.uk. I am grateful to Mark Rosenzweig, Larry Samuelson, and Chris Udry for their advice and support throughout this project. Treb Allen, Priyanka Anand, David Berger, Gharad Bryan, Avinash Dixit, Tim Guinnane, Melanie Morten, Sharun Mukand, Motty Perry, Joe Vavra, and various seminar participants provided helpful comments. I acknowledge research funding from the Yale Economic Growth Center.

Many societies are divided into multiple smaller groups. These divisions are especially salient in many developing countries, where the groups have names such as castes, tribes, or clans, but developed countries are divided as well, for example by race and religion. One stylized fact about group divisions is that people are more likely to interact in certain ways with members of their own groups than with members of different groups. Interactions that take place primarily within groups include trade (Greif 1993, Anderson 2011), mutual insurance (Grimard 1997, Mazzocco and Saini 2012, Munshi and Rosenzweig 2016), and job referrals (Munshi and Rosenzweig 2006). At first glance the lack of interaction between groups is puzzling, since the argument from the gains from trade suggests that people should seek to interact with the most diverse possible range of partners. In this paper, I describe a mechanism that can generate such social divisions in equilibrium through a reputation effect, even when there are no fundamental barriers to interaction between members of different groups.

An example of social division of the type I wish to describe comes from Mayer's (1960) description of the caste system in the village of Ramkheri in central India. The central fact of the caste system, according to Mayer, is what he refers to as the commensal hierarchy, which prescribes who may eat with whom. There are five major caste groupings in the village, and members of higher ranked castes refuse to eat with or accept food from members of lower ranked castes, although members of lower ranking castes are willing to accept food from members of higher ranking castes. Mayer writes, 'Eating the food cooked or served by a member of another caste denotes equality with it, or inferiority, and not to eat denotes equality or superiority'. (Mayer, 1960, p. 34) As eating together is one of the main ways to develop friendships, friendships are less likely to form across caste lines than within castes.

Whether people follow the rules of the hierarchy depends to some extent on whether other members of their caste can observe them. Mayer describes a member of an upper caste who was born in the village but who is working in the city of Indore. On a visit to the village, he is offered tea by a member of a lower caste, but he refuses, saying 'I would willingly drink in Indore, but I must be careful not to offend anyone here'. (Mayer, 1960, p. 50) Similarly, Mayer describes a meal at a training camp for development workers held in the village, which is attended by delegates from many other villages. The delegates from other villages all eat together, while the delegates from Ramkheri sit separately in accordance with the caste rules. The Ramkheri delegates explain the situation, saying, 'We could not sit with them here; but they, being away from their villages, were able to sit next to Muslims and even Harijans [members of the lowest Hindu caste]'. (Mayer, 1960, p. 51) According to Mayer this phenomenon is due to the greater difficulty in observing violations of caste rules that take place outside the village. Mayer writes, 'The orthodox in Ramkheri know that the rules are being broken outside, [but] they are content not to investigate, so long as the matter is not given open recognition'. (Mayer, 1960, p. 50) Finally, after breaking the rules regarding caste contact, caste members are obliged to perform a ritual purification. However, whether the purification is in fact performed depends on whether the violation is observed. Mayer writes, ‘Touching a Tanner [one of the lowest castes] is a more generally acknowledged matter for purification..., though it is admitted that many people would not do anything if they were not seen to touch’. (Mayer, 1960 p. 58) Thus people seem to follow the rules of the hierarchy in part to preserve their reputations with members of their own castes.

Not all interactions between castes are penalized in Ramkheri. The Ramkheri caste system distinguishes between the sharing of different kinds of foods between castes. *Kacca* foods are foods cooked with water or salt. They include most daily staples. *Pakka* foods are foods cooked with butter. They are served at ceremonial occasions. The rules regarding *kacca* foods are much more stringent than the rules regarding *pakka* foods, and people are willing to accept *pakka* foods from members of lower castes from whom they would not be willing to accept *kacca* foods. My interpretation of this distinction is that sharing *kacca* food, which is eaten every day, is much more likely to lead to a deep, cooperative relationship than sharing *pakka* food, which is eaten only rarely.

To summarize, the Ramkheri caste system exhibits four important features. First, members of different castes do not interact in certain ways. Second, there is a hierarchy over castes, and members of higher ranking castes refuse to interact with members of lower ranking castes but not vice versa. Third, caste members follow the rules about non-interaction with other castes in part to preserve their reputations with members of their own castes. Fourth, the reputational penalties for interacting with members of other castes are more severe for those interactions which are most likely to lead to deep, cooperative relationships. I now outline a model that accounts for all of these features.

In the model, agents search over the community to find partners for cooperative relationships. If an agent cheats in any relationship, then the relationship breaks up and each partner to the relationship must search for a new partner. Search requires effort and hence is costly. Cooperation is maintained by the threat that any cheating agent will have to pay the cost of search, and the level of cooperation that any agent can support is inversely related to the search cost that the agent is expected to incur at the end of the relationship. Agents who expect to form matches with a larger fraction of potential partners pay lower search costs in expectation. Thus an agent who is expected to form matches with a larger proportion of the community can support a lower level of cooperation in any given relationship. Each agent is also a member of a payoff irrelevant group, and in equilibrium each agent interacts only with members of her own group. If an agent is observed to have formed a match with a member of a different group in the past, then it is believed that the agent will continue to accept matches both with members of her own group and with members of the other group in the future. Thus, agents who are observed to have interacted with members of different groups in the past are able to support lower levels of cooperation. This penalty for interacting with members of different groups is sufficient to prevent members of different groups from interacting in equilibrium. I refer to this state of affairs as group segregation.

Group segregation increases the level of cooperation that each agent can support compared to the situation without segregation, and if the benefits of cooperation are sufficiently important, then group segregation is welfare improving for the community as a whole.

The reputation mechanism yields novel theoretical insights. The first insight is that people may lose reputation with members of their own group by interacting with members of different groups. Specifically, people who interact with members of different groups are believed to be less trustworthy by members of their own group.

A second insight is that the reputation mechanism endogenously generates an asymmetry between different groups. Suppose that there are two groups, group 1 and group 2, and suppose that members of group 1 refuse to interact with members of group 2. Members of group 1 refuse to interact with members of group 2 because it is believed that a member of group 1 who has interacted with a member of group 2 in the past will continue to interact with members of group 2 in the future. However, this belief is rational only if members of group 2 are willing to interact with members of group 1. The groups are thus organized in a hierarchical structure, with higher ranking groups refusing to interact with lower ranking groups, but not vice versa.

In the literature the paper most closely related to mine is Eeckhout (2006). Eeckhout also constructs a model in which agents refuse to cooperate at a high level with members of different groups, and in which group segregation increases the level of cooperation that can be sustained within each group. Eeckhout's model differs from mine in the reason why agents condition their actions on payoff-irrelevant group membership. In Eeckhout's model, there are multiple possible equilibria within each relationship. In relationships between partners who are members of the same group, the partners coordinate on a equilibrium with a high level of cooperation, while in relationships between partners who are members of different groups, the partners coordinate on little or no cooperation. A difficulty with this idea is that there is no reason why partners in mixed relationships could not jointly decide to coordinate on the (Pareto superior) high cooperation equilibrium instead of the low cooperation equilibrium required by the model. Formally, Eeckhout's model fails to satisfy an equilibrium refinement called bilateral rationality, first introduced by Ghosh and Ray (1996). Bilateral rationality is closely related to the various renegotiation-proofness concepts discussed in Farrell and Maskin (1989) and Bernheim and Ray (1989). The reputation effect introduced in my model solves this problem. The reputation effect makes renegotiation more difficult than it is in Eeckhout's model. In my model, an agent who interacts with a member of a different group changes the beliefs of every other member of the community about how the agent will behave in the future. In order to achieve a high level of cooperation in a mixed relationship, an agent in my model would have to renegotiate not only with her partner but also with all of the other members of the community. I assume that while agents may be able to renegotiate with their partners, they cannot renegotiate with the entire community, and the impossibility of this latter renegotiation allows for the reputation effect that characterizes my model.

My paper is also related to the larger literature on community enforcement. The central theme in this literature is that communities can be organized to help sustain higher levels of cooperation than can be sustained through isolated bilateral relationships. Two strategies that are commonly used in this literature to sustain cooperation are ostracism and contagion. Ostracism strategies, as discussed for example in Greif (1993), Dixit (2003), and Ali and Miller (2015), require that agents refuse to cooperate in the present with partners who have been observed to cheat in the past. Contagion strategies, as discussed for example by Kandori (1992), Ellison (1994), and Nava and Piccione (2014), require that agents who have been cheated in the past cheat their current partners, so that instances of cheating lead to a generalized breakdown of cooperation throughout the community. I consider an environment in which neither of these strategies are feasible. Ostracism is not feasible because I assume that agents cannot observe whether their current partners have cheated in their previous relationships, and because I assume that agents' strategy sets are sufficiently restricted that they cannot use complex behaviour to communicate their past actions to their present partners as in Deb (2008). Contagion is not feasible because I assume that the community contains a large (in fact infinite) number of agents. With a large number of agents, contagious waves of cheating remain confined to a small fraction of the community and cannot cause a generalized breakdown of cooperation. Of course, real communities may use multiple strategies including ostracism, contagion, and hierarchical segregation to sustain cooperation. I do not argue that ostracism and contagion do not exist, but by showing that hierarchical segregation can sustain cooperation in situations where ostracism and contagion are infeasible, I show that hierarchical segregation can be a useful complement to these other strategies.

Two other models with a structure closely related to mine are Akerlof (1976) and Pęski and Szentes (2013). In both of these papers, agents can observe not only with whom their current relationship partners have interacted in the past as in my model, but also with whom their partners' partners have interacted in the past, with whom their partners' partners' partners have interacted in the past, and so on to infinity. These papers then construct equilibria that enforce group segregation through a kind of infinite regress, in which there is a punishment for interacting with a member of a different group, a punishment for failing to inflict this punishment, a punishment for failure to punish failure to punish, and so on. This infinite regress solves the renegotiation-proofness problem in Eeckhout (2006), at the cost of introducing extremely demanding information requirements, since agents in the Akerlof and Pęski and Szentes models must have information about an unboundedly large number of past interactions. I argue that the very demanding information requirements of the Akerlof and Pęski and Szentes models can be satisfied in real societies only if there is a specialized institution dedicated to collecting the necessary information and broadcasting this information throughout the society. In contrast, the much less demanding information requirements of my model can be satisfied in a society in which information spreads through uncoordinated gossip. Using evidence from the anthropological literature, I argue that some Indian castes have in fact developed such a specialized information broadcasting institution, and that these castes have social punishments that correspond to the punishments in the Akerlof and Pęski and Szentes models. In other castes, in contrast, no information broadcasting institution exists, and in these castes social punishments correspond more closely to the punishments in my model.

The remainder of the paper is as follows. Section II presents the model. Section III discusses the relationship between my model and the models of Akerlof and Pęski and Szentes using evidence from the anthropological literature. Section IV concludes.

1 Model

1.1 Setup

Time is discrete, starts at period 0 and continues forever. There exists a continuum of agents with mass 1. Each agent is a member of one of G groups, and each group has mass $1/G$. Groups are payoff-irrelevant, but group membership is observable. Each agent has N relationship "slots", indexed $0, ..., N-1$. At the beginning of the game each agent from each group g is matched with a partner who is also from group g in each slot. Matches are sorted so that if agents i and j are matched, and agent j is in agent i's nth relationship slot, then agent i is also in agent j's nth relationship slot. Each agent has a fixed discount factor $\delta^{1/N}$. In period t, relationship slot $t \mod N$ becomes active for each agent (i.e. relationship slot 0 is active in periods 0, N, $2N$,..., relationship slot 1 is active in periods 1, $N+1$, $2N+1$,..., etc.). Then, the following things happen:

1. Each agent simultaneously proposes stage game stakes $a \in [0, \infty)$ for the active relationship slot[1]. If agents i and j are matched in the active slot, and they propose stakes a_i and a_j with $a_i \neq a_j$, then the relationship between i and j breaks up and agents i and j continue to step 3. If $a_i = a_j$, then agents i and j continue to step 2.

2. Each agent who still has a match in the active slot chooses whether to cooperate or to cheat. If either agent in a match chooses to cheat, then the match breaks up with certainty. If both agents choose to cooperate, then the match breaks up with exogenous probability p. If two agents are in a match that does not break up, then those agents receive no further payoff for the period. If an agent is in a match that breaks up, then that relationship slot becomes open and the agent continues to step 3.

3. Each agent with an open relationship slot pays a search cost c and is provisionally matched with another agent with an open relationship slot. Agents are provisionally matched according to a uniform probability distribution over the set of agents with open relationship slots in slot $t \mod N$.[2]

[1] This stage game is adapted from Ghosh and Ray (1996).

[2] More precisely, as will be seen below an agent can be completely characterized by her group g and a what I call the agent's

4. Provisionally matched agents observe their partners' groups and past match sets. Each agent may then choose to accept or reject the match. If either agent rejects the match, then the match is dissolved and both agents return to step 3 to search again. If both agents accept then the agents form a match in slot t mod N and neither agent receives any further payoff for the period.

Stage game payoffs are as follows. If both players cooperate at stakes a, then both players receive payoff $v(a)$. If one player cooperates and the other cheats, then the cheating player receives payoff $d(a)$ and the cooperating player receives payoff $\ell(a)$. I sometimes refer to $v(a)$ as the value of cooperating at level a, and $d(a)$ as the value of cheating at level a. If both players cheat then both players receive payoff 0. I make the following assumptions about the stage game payoffs:

Assumption 1. *1. For all $a > 0$, $d(a) > v(a)$ and $\ell(a) < 0$.*

2. $v(a)$ is bounded.

3. $v(0) = d(0) = 0$

4. $v(a)$ and $d(a)$ are continuous, twice differentiable, and strictly increasing in a.

5. $v(a)$ is strictly concave in a and $d(a)$ is strictly convex in a.

Part 1 of the assumption states that for all $a > 0$, cheating is the strictly dominant action in the stage game, which can be interpreted as a generalized prisoner's dilemma with variable stakes. Part 2 is required to rule out Ponzi schemes, in which any level of cooperation can be attained through the promise of ever higher levels of cooperation in the future. Parts 3 through 5 imply that the temptation to cheat is small for a small, and that the temptation to cheat grows large as a gets large. These assumptions ensure that the solution to each agent's maximization problem is interior.

I assume that a relationship breaks up automatically if either partner to the relationship cheats. This assumption simplifies the analysis by allowing me to disregard the possibility of complex punishments in which partners choose asymmetric actions or in which cheaters are forgiven after some limited punishment period. A similar assumption appears in related models including Ghosh and Ray (1996) and Jackson, Rodriguez-Barraquer and Tan (2011). The assumption could be justified as a reduced form for one of two mechanisms. First, the assumption could be justified by supposing that agents put a negative value on interacting with partners who have cheated them in the past, perhaps due to some kind of preference for reciprocity (e.g. Fehr, Gächter and Kirchsteiger 1997). Second, the assumption could be justified by supposing that each agent gets a privately known match-specific value from each relationship, and that cheating indicates that the relationship

past match set, $\mathcal{H}$. There are a finite number of possible tuples $(g, \mathcal{H})$. The probability that an agent is provisionally matched with a partner with group and past match set $(g, \mathcal{H})$ is just the proportion of agents with group and past match set $(g, \mathcal{H})$ within the population of all agents with open relationship slots.

has low value for the agent who cheats (e.g. Halac 2012). Either of these assumptions would imply that at least one partner to a relationship in which cheating has occurred would prefer to search for a new partner rather than continuing the existing relationship. I hypothesize that the intuition for my model extends to cases in which more complex punishments for cheating are allowed, but I leave the details of this exploration for future research.

Each agent can observe her group and the group of any other agent with whom she is matched. Each agent can also observe the history of play within each current match, but she cannot observe the history of play in any match in which she does not participate. However, each agent can observe something about with whom each of her partners has matched in the past. Specifically, for each group g and each partner j, an agent can observe whether partner j has ever been matched with any agent in group g. Let $\mathcal{H}_i \subseteq \{1, ...G\}$ be the set of groups g such that agent i has been matched with a member of group g in the past. I refer to the set $\mathcal{H}_i$ as agent i's past match set. Note that for all groups g, if agent i is in group g then $g \in \mathcal{H}_i$, since at the beginning of the game each agent is matched with N members of her own group.

A (pure) strategy s_i for an agent i is a choice of whether to accept a match with any given partner, a choice of stakes to propose in the stage game, and a choice of whether to cooperate or to cheat, conditional player i's information. I consider strategies in which the agent conditions her actions only on her own group and past match set and the groups and past match sets of her current partners. Let μ_i be a vector containing the groups and past match sets of the agents who are currently matched with agent i. I sometimes refer to μ_i as agent i's match profile. Then a strategy for agent i takes the form

$$\begin{aligned}
s_i &= s(g_i, \mathcal{H}_i, g_j, \mathcal{H}_j, \mu_i) \\
&= (m(g_i, \mathcal{H}_i, g_j, \mathcal{H}_j, \mu_i), a(g_i, \mathcal{H}_i, g_j, \mathcal{H}_j, \mu_i), w(a, g_i, \mathcal{H}_i, g_j, \mathcal{H}_j, \mu_i))
\end{aligned}$$

where $m \in \{A, R\}$ is the decision to accept or reject a match with a partner with group and past match set $(g_j, \mathcal{H}_j)$, $a \in [0, \infty)$ is the stage game action chosen in a match with that partner, and $w \in \{O, H\}$ is the decision to cooperate or to cheat depending on the stakes a chosen in for the stage game. Since the function s mapping groups and past match sets to strategies is the same for all agents, I also use s to denote the strategy profile consisting of each agent's strategy. Finally, when considering possible deviations it is necessary to distinguish the strategies that agents choose with each of their partners. I write s_{ij} to denote that strategy that agent i follows with partner j.

Given a strategy profile, an agent's expected continuation payoff in a given period depends on her group, her past match set, the groups and past match sets of the agents with whom she is matched in the period, and the agent's beliefs about her partners' matches and about the overall distribution of groups and past match sets in the population. I now introduce notation to describe an agent's beliefs. Suppose that player i is matched with agent j. Agent j has match profile μ_j, and the composition of μ_j may affect agent j's choice of action and hence agent i's continuation payoff. Let ν_{ij} be a vector containing agent i's belief about the probability that agent j's partners are μ_j for each possible μ_j. Let ν_i be a vector containing ν_{ij} for all agents j matched with agent i. Similarly, let Γ_i contain agent i's belief about the distribution of groups and past match sets in the pool of unmatched agents in a given period. An agent's beliefs Γ_i may affect her payoff by affecting her expectation about the payoff from future matches if any current matches break up.

An agent i's belief about her partners' match profiles could in principle depend on any information that agent i has observed. I denote agent i's information in a given period by $\mathcal{I}_i$, and to emphasize this dependence I write agent i's belief as $\nu(\mathcal{I}_i)$. I can then write an agent i's expected continuation payoff in a given period as:

$$EU_i(s, g_i, \mathcal{H}_i, \mu_i | \nu(\mathcal{I}_i), \Gamma_i).$$

An agent i who is matched with a partner j may deviate from a strategy profile s by choosing to follow strategy s'_{ij} in her relationship with partner j, while continuing to follow strategy s_i in all of her other relationships. I denote the strategy of following s'_{ij} with agent j and s_i with all other agents by $\frac{s'_{ij}}{s_i}$. Letting s_{-i} denote the strategy profile followed by all agents other than agent i, I denote agent i's utility from this deviation by:

$$EU_i(\frac{s'_{ij}}{s_i}, s_{-i}, g_i, \mathcal{H}_i | \nu(\mathcal{I}_i), \Gamma_i).$$

A pair of matched agents i and j may also choose to deviate simultaneously from a strategy profile s to follow an alternative strategy in their relationship with each other, while continuing to follow strategy profile s in their other relationships. Suppose that agent i deviates to strategy s'_{ij} in her relationship with agent j, and agent j deviates to strategy s'_{ji} in her relationship with agent i. Letting s_{-ij} denote the strategy profile followed by all agents other than agents i and j, I denote agent i's utility from this simultaneous deviation by:

$$EU_i(\frac{s'_{ij}}{s_i}, \frac{s'_{ji}}{s_j}, s_{-ij}, g_i, \mathcal{H}_i | \nu(\mathcal{I}_i), \Gamma_i).$$

1.2 Equilibrium Concept

A strategy profile s is stationary if, when every agent follows the strategy profile, the distribution Γ of groups and past match sets in the pool of unmatched agents is the same in every period. A stationary strategy profile is an equilibrium of my model if it satisfies two requirements. First, a strategy profile must satisfy an individual rationality requirement. Informally, a strategy profile is individually rational if there is no unilateral deviation by any agent that provides the agent with higher expected utility than the agent would expect to receive if she followed the strategy profile. In order to define an agent's expected utility, I must first define the agent's beliefs. On the equilibrium path, I assume that beliefs are determined according to Bayes' rule. Off the equilibrium path, I face the familiar problem that agents must assign posterior beliefs to events with prior probability zero. If beliefs were unrestricted, then an agent who observed a deviation could come to believe that a positive mass of other unobserved agents had also deviated. This kind of updating seems unreasonable, however, since each agent observes only a finite number of other agents' actions and so observing a deviation by one agent can only convey information about the actions taken by a finite number of other agents. In the words of Fudenberg and Tirole (1991), it seems reasonable to suppose instead that there is "no signalling what you don't know", and so agents never believe that more than a finite number of unobserved agents have deviated.[3] Following this slogan, I make two assumptions on beliefs. First, I assume that agents always believe that the distribution of groups and past match sets in the pool of unmatched agents is Γ, the equilibrium path distribution, since a finite number of deviations does not affect the distribution in a continuum of agents. Second, I assume that each agent i's belief $\nu(\mathcal{I}_i)$ assigns probability zero to the event that any of her partners' partners have deviated from the equilibrium strategy profile in the past, regardless of the information $\mathcal{I}_i$ possessed by the agent. Again, this is because each agent believes with probability 1 that any given unobserved agent has not deviated from the strategy profile, regardless of how many deviations the agent has observed. I say that beliefs are consistent if they satisfy these conditions. I then define individual rationality formally as follows:

Individual rationality: A stationary strategy profile s, the associated distribution Γ, and beliefs $\nu(\mathcal{I}_i)$ satisfy individual rationality if $\nu(\mathcal{I}_i)$ is consistent, and if, for all s'_{ij}, g_i, $\mathcal{H}_i$, and μ_i,

$$EU_i(s_i, s_{-i}, g_i, \mathcal{H}_i, \mu_i | \nu(\mathcal{I}_i), \Gamma) \geq EU_i(\frac{s'_{ij}}{s_i}, s_{-i}, g_i, \mathcal{H}_i, \mu_i | \nu(\mathcal{I}_i), \Gamma)$$

Individual rationality implies that no agent has a profitable individual deviation after any history. This condition does not rule out all possible deviations that might be considered. If two matched agents have the ability to communicate before choosing their actions, then they might consider deviating simultaneously in a way that is profitable for both of them. Not all such joint deviations are credible, however. A proposed joint deviation is credible if neither deviating partner has a further profitable individual deviation from the proposed joint deviation. As before, I assume that expectations regarding payoffs are taken with respect to consistent beliefs. My second requirement for an equilibrium is that an equilibrium strategy profile must not allow for any mutually profitable and credible joint deviations by any pair of matched agents. Following Ghosh and Ray (1996), I call this requirement bilateral rationality. Bilateral rationality is defined as follows:

[3] Agents in my model are not subject to random shocks. However, if there were correlated shocks that could affect the profitability of deviations for positive masses of agents, then an agent who observed a deviation could come to believe that a positive mass of other agents had also deviated. Such beliefs would make the segregated equilibrium described below more difficult to sustain. This suggests that hierarchical segregation may be more difficult to sustain in environments in which agents are subject to large, correlated shocks.

Bilateral rationality: A stationary strategy profile s, the associated distribution Γ, and beliefs $\nu(\mathcal{I}_i)$ satisfy bilateral rationality if $\nu(\mathcal{I}_i)$ is consistent and if, for any g_i, $\mathcal{H}_i$, g_j, $\mathcal{H}_j$, μ_i, and μ_j, there does not exist a deviant strategy profile s' such that

$$EU_i(\frac{s'_{ij}}{s_i}, \frac{s'_{ji}}{s_j}, s_{-ij}, g_i, \mathcal{H}_i, \mu_i | \nu(\mathcal{I}_i), \Gamma) \geq EU_i(s_i, s_j, s_{-ij}, g_i, \mathcal{H}_i, \mu_i | \nu(\mathcal{I}_i), \Gamma) \text{ and}$$
$$EU_j(\frac{s'_{ji}}{s_j}, \frac{s'_{ij}}{s_i}, s_{-ij}, g_j, \mathcal{H}_j, \mu_j | \nu(\mathcal{I}_i), \Gamma) \geq EU_j(s_j, s_i, s_{-ij}, g_j, \mathcal{H}_j, \mu_j | \nu(\mathcal{I}_i), \Gamma)$$

with at least one of the two previous inequalities strict, and such that for all s'',

$$EU_i(\frac{s'_{ij}}{s_i}, \frac{s'_{ji}}{s_j}, s_{-ij}, g_i, \mathcal{H}_i, \mu_i | \nu(\mathcal{I}_i), \Gamma) \geq EU_i(\frac{s''_{ij}}{s_i}, \frac{s'_{ji}}{s_j}, s_{-ij}, g_i, \mathcal{H}_i, \mu_i | \nu(\mathcal{I}_i), \Gamma) \text{ and}$$
$$EU_j(\frac{s'_{ji}}{s_j}, \frac{s'_{ij}}{s_i}, s_{-ij}, g_j, \mathcal{H}_j, \mu_j | \nu(\mathcal{I}_i), \Gamma) \geq EU_j(\frac{s''_{ji}}{s_i}, \frac{s'_{ij}}{s_i}, s_{-ij}, g_j, \mathcal{H}_j, \mu_j | \nu(\mathcal{I}_i), \Gamma).$$

The first set of inequalities in the definition state that the joint deviation from s to s' is preferred to s by both agents i and j, and strictly preferred by at least one of them. The second set of inequalities state that there is no further deviation to s'' from s' that is individually profitable for either agent i or j. A strategy profile is bilaterally rational if there is no such mutually profitable and credible joint deviation. A strategy profile and the associated beliefs are an equilibrium if they satisfy both individual rationality and bilateral rationality. As will be seen, the requirement that equilibrium satisfy bilateral rationality is the main innovation of this paper, and the need to satisfy the bilateral rationality requirement drives essentially all of my results.

One problem with my equilibrium concept is that, as will be seen, an equilibrium does not necessarily exist for all parameter values. This is a problem common to many of the renegotiation-proofness concepts in the literature. For example, equilibrium also fails to exist for some parameter values in Ghosh and Ray (1996). The fact that equilibrium fails to exist for some parameter values suggests that the equilibrium concept is too strong in some way, but it is unclear what is the right weakening of the equilibrium concept to ensure existence for all parameter values.

1.3 A Benchmark Equilibrium

I will begin my analysis by discussing a benchmark strategy profile in which agents do not condition their actions on their own or their partner's group membership or past match set. If the benchmark strategy profile is part of an equilibrium, I will refer to that equilibrium as a benchmark equilibrium.

A benchmark strategy profile is as follows:

Benchmark strategy profile All agents accept all matches, and propose a level of cooperation $\bar{a}_B$ within each match. Agents cooperate if the agreed level of cooperation is less than or equal to $\bar{a}_B$ and cheat otherwise.

A benchmark strategy profile is an equilibrium if there are no profitable individual or joint deviations. I must check that no agent can profit individually by cheating in any relationship, and also that no pair of matched agents can jointly profit by deviating to a higher level of cooperation that is individually rational for both agents. In principle, I also need to check that it is optimal for all agents to accept matches with all other members of the community. However, this last condition is trivial in the benchmark equilibrium, since all match partners are identical. Moreover, since all agents behave the same way regardless of groups or past match sets, agents' beliefs about their partners' match profiles and the overall distribution of groups and past match sets in the pool of unmatched agents are irrelevant.

Because actions taken in one relationship do not affect any other relationship under the benchmark strategy profile, it is possible to analyse each relationship slot separately. Let V_B^m be the value that an agent expects to receive from a filled relationship slot at the beginning of any period. Let V_B^u be the value that an agent expects to receive from an open relationship slot. I also define V_B^f to be the expected value to each agent of having a filled relationship slot at the beginning of any future period. In the proof of proposition 1 it is helpful to distinguish V_B^f from V_B^m because agents may be able to affect V_B^m through renegotiation, but they cannot affect V_B^f. Bilateral rationality dictates that each pair of matched agents chooses the level of cooperation that maximizes their joint utility, subject to the constraint that no agent can profit individually by choosing to cheat. That is, V_B^m must satisfy:

$$V_B^m = \max_a (1-\delta)v(a) + pV_B^u + (1-p)\delta V_B^m \tag{1}$$

subject to the constraint

$$(1-\delta)v(a) + pV_B^u + (1-p)\delta V_B^m \geq (1-\delta)d(a) + V_B^u \tag{2}$$

Equation (1) says that an agent gets $v(a)$ from a match in a given period, and then becomes unmatched with probability p and remains matched with the existing parter with probability $(1-p)$. Note that the current relationship slot will next be active N periods in the future, which means that the discount factor applied to future interactions with the current partner is $(\delta^{1/N})^N = \delta$. The constraint (2) is the individual rationality constraint. It states that the payoff from cooperating must be greater than the payoff from cheating. If the agent cheats she receives $d(a)$ in the current period and then has an empty relationship slot that she must fill through search. The payoff to having an empty relationship slot V_B^u is defined by:

$$V_B^u = -(1-\delta)c + \delta V_B^f \tag{3}$$

Equation (3) says that an agent with an empty relationship slot must pay the search cost before being matched with a new partner and receiving the payoff to that match.

A benchmark equilibrium is a benchmark strategy profile such that V_B^m, V_B^u, and V_B^f satisfy equation (1) subject to (2) and equation (3), such that $\bar{a}_B$ maximizes (1) subject to (2), and such that $V_B^m = V_B^f$.

Define $\hat{a}$ to be the value of a that solves

$$\max_a v(a) - (1 - \delta(1-p))d(a).$$

The following proposition provides conditions under which a benchmark equilibrium exists, and derives the level of cooperation in a benchmark equilibrium:

Proposition 1. *A benchmark equilibrium exists if and only if c satisfies*

$$c \geq \frac{1}{\delta(1-p)}[d(\hat{a}) - v(\hat{a})]. \tag{4}$$

If a benchmark equilibrium exists, then the equilibrium level of cooperation $\bar{a}_B$ solves

$$d(\bar{a}_B) - v(\bar{a}_B) = (1-p)c \tag{5}$$

Omitted proofs are in appendix A.

The interpretation of the expression for the level of cooperation in the benchmark equilibrium is straightforward. If an agent cheats in the current period, her net gain in the period is the difference between the value of cheating $d(\bar{a}_B)$ and the value of cooperating $v(\bar{a}_B)$. The cost of cheating is that the cheating agent's match will break up and the agent will have to pay the search cost with certainty rather than with probability p, for a net cost of $(1-p)c$. The maximum level of cooperation that can be sustained is the level of cooperation such that the net cost of cheating is equal to the net benefit. The bilateral rationality condition ensures that all agents will renegotiate up to the highest possible level of cooperation, so only the maximum sustainable level of cooperation is consistent with equilibrium.

I briefly discuss the intuition for the fact that no bilaterally rational equilibrium exists unless c is sufficiently large. I consider strategy profiles in which all agents choose the same level of cooperation every period. Since all agents accept all matches, any agent can cheat in her current relationship, break up the relationship at the end of the period, pay the search cost c, and find a new partner in the next period. Since all agents choose the same level of cooperation, the deviating agent will be able to cooperate at the same level in her new relationship as she did in the old relationship. Thus, if c is low, then the penalty for cheating in any given relationship is low, and so the common sustainable level of cooperation is low. However, if all agents are cooperating at some common low level, then any two matched agents can jointly deviate to a higher level of cooperation. This higher level of cooperation does not violate the individual rationality constraint, so long as only two agents are cooperating at the high level, because the penalty for breaking up this deviant relationship is high: if either agent breaks the relationship, both agents must go back to cooperating at the low common level of cooperation. Thus the individual rationality requirement rules out all strategy profiles except those strategy profiles with a low common level of cooperation, and the bilateral rationality requirement rules out strategy profiles with a low common level of cooperation, so that there are no remaining equilibrium strategy profiles. As c gets larger, higher levels of cooperation become compatible with the individual rationality constraint, and for c sufficiently large there exist levels of cooperation that are high enough to satisfy the bilateral rationality requirement while still satisfying the individual rationality constraint.[4]

It is useful to have a closed form expression for the value of a relationship V_B^m in the benchmark equilibrium. Rearranging the definitions of V_B^m and V_B^u yields:

$$V_B^m = v(\bar{a}_B) - pc$$

That is, an agent's utility from each relationship slot under the benchmark equilibrium is equal to the value the agent gets from cooperating in each period, minus the search cost that the agent pays each period with probability p.

1.4 *Motivating the Segregated Equilibrium*

My goal is to construct an equilibrium that supports higher levels of cooperation and provides agents with higher welfare than the benchmark equilibrium. I do this by constructing an equilibrium in which agents reject some matches, instead of accepting all matches as in the benchmark equilibrium. If agents reject some matches, then the expected cost of search for an unmatched agent is higher than in the benchmark equilibrium, and so the penalty for cheating and the level of cooperation that can be supported in each match are also higher. If the exogenous probability p that a match breaks up is sufficiently low, then the benefit of a higher level of cooperation outweighs the cost of having to pay greater search costs, and so welfare is greater than under the benchmark equilibrium.

The main barrier to constructing an equilibrium in which agents reject some potential matches is the bilateral rationality requirement. To build intuition for why bilateral rationality makes it difficult to construct such an equilibrium, consider the following strategy profile, which is a simplified version of the strategy profile considered by Eeckhout (2006), and which I will refer to as strategy profile E:

Strategy profile E: Agents accept matches with members of their own group, and reject matches with

[4] A similar issue arises in Ghosh and Ray (1996), and the proof of proposition 1 draws on ideas from the proofs in that paper.

members of any other group, regardless of past match histories. Within each match all agents propose a level of cooperation $\bar{a}_E$. Agents cooperate if the proposed level of cooperation is less than or equal to $\bar{a}_E$ and cheat otherwise.

As in the benchmark equilibrium, under strategy profile E actions taken in one relationship slot do not affect the optimal action in any other relationship slot. Thus it is possible to analyse each relationship slot separately. Let V_E^m be the value to an agent from having a filled relationship slot in a period, and let V_E^u be the value to an agent from having an empty relationship slot in a period. Under strategy profile E the composition of the pool of unmatched agents is strategically relevant. Fortunately the composition is easy to describe. If all agents follow strategy profile E, then in each period the proportion of agents from each group in the pool of unmatched agents is $\frac{1}{G}$. Thus a searching agent meets a partner from her own group with probability $\frac{1}{G}$. Using this probability I can write expressions for V_E^m and V_E^u as follows:

$$V_E^m = (1-\delta)v(\bar{a}_E) + pV_E^u + (1-p)\delta V_E^m$$
$$V_E^u = -(1-\delta)c + \frac{1}{G}\delta V_E^m + \frac{G-1}{G}V_E^u$$

The first equation says that an agent who is matched cooperates at level $\bar{a}_E$ in the current period. Her match then breaks up with probability p, and she continues to the next period with the same partner with probability $(1-p)$. The second equation says that an unmatched agent pays the search cost and is matched with a partner with probability $\frac{1}{G}$, and otherwise remains unmatched and must pay the search cost again.

Strategy profile E satisfies individual rationality if

$$V_E^m \geq (1-\delta)d(\bar{a}_E) + V_E^u$$

Substituting in the definitions of V_E^m and V_E^u and rearranging yields that strategy profile E satisfies individual rationality if

$$d(\bar{a}_E) - v(\bar{a}_E) \leq (1-p)Gc$$

Comparing this expression to the expression defining the benchmark level of cooperation $\bar{a}_B$ yields the following:

LEMMA 1. *If $G > 1$, then there exist values of $\bar{a}_E$ such that $\bar{a}_E > \bar{a}_B$ and such that strategy profile E satisfies individual rationality.*

Proof. Since $d(\cdot)$ is strictly convex and $v(\cdot)$ is strictly concave, the difference $d(a) - v(a)$ is increasing in a. Thus for $G > 1$ there exist values of $\bar{a}_E$ that solve

$$d(\bar{a}_E) - v(\bar{a}_E) \leq (1-p)Gc$$

and such that $\bar{a}_E > \bar{a}_B$, where $\bar{a}_B$ is defined as in proposition 1. □

Higher levels of cooperation are individually rational under strategy profile E than in the benchmark equilibrium because agents expect to form matches with only $1/G$ of their potential partners under strategy profile E, while they expect to form matches with all of their potential partners in the benchmark equilibrium. Thus, the expected cost of breaking up a relationship is higher under strategy profile E than in the benchmark equilibrium, and so the individually rational level of cooperation is higher under strategy profile E than in the benchmark equilibrium.

Although strategy profile E is individually rational and may allow agents to achieve higher levels of cooperation than the benchmark equilibrium, we also have the following:

LEMMA 2. *Strategy profile E is not an equilibrium, because it does not satisfy the bilateral rationality requirement.*

Proof. Let $\bar{a}_E$ be an action such that $d(\bar{a}_E) - v(\bar{a}_E) \leq (1-p)Gc$, which implies that strategy profile E satisfies the individual rationality constraint. Consider two agents from different groups who are provisionally matched. By rejecting the match both agents get utility V_E^u, while by jointly deviating to accept the match both agents get $V_E^m > V_E^u$. Moreover, the incentives in this deviant relationship are exactly the same as the incentives in relationships entered into by following strategy profile E, and so since strategy profile E satisfies the individual rationality condition so does the joint deviation to accepting this match. Thus the joint deviation to accepting the match is individual rational and makes both partners to the match strictly better off, and so strategy profile E does not satisfy the bilateral rationality condition and is not an equilibrium. □

The problem with strategy profile E is that under the strategy profile relationships between members of different groups are just as profitable as relationships between members of the same group, and yet members of different groups do not interact. Intuitively it seems implausible that people would consistently fail to seize opportunities for profitable interaction in this way. The bilateral rationality requirement formalizes this intuition. A more plausible theory of group segregation would provide a reason why relationships between members of different groups are less profitable than relationships between members of the same group. In the next section I construct a strategy profile that contains just such a reason, and which therefore does satisfy the bilateral rationality requirement.

1.5 Segregated Equilibrium

In this section I propose what I will call a segregated strategy profile. As before, if a segregated strategy profile is part of an equilibrium, I refer to the equilibrium as a segregated equilibrium. Although the segregated strategy profile is similar to strategy profile E, it includes some additional features that ensure that unlike strategy profile E, a segregated strategy profile can satisfy bilateral rationality and hence can be an equilibrium.

My discussion of the segregated equilibrium proceeds in four steps. First, I define a segregated strategy profile. Second, I define value functions for various states under the segregated strategy profile, and discuss some properties of these value functions. Third, I describe the levels of cooperation chosen under a segregated equilibrium. Finally, I prove that a segregated equilibrium does in fact exist for some parameter values.

1.5.1 Definition

Under a segregated strategy profile, groups are ranked in a hierarchy. I label the groups so that group 1 is ranked highest in the hierarchy and group G is ranked lowest. Thus $g < g'$ means that g is ranked above g'. Informally, a segregated strategy profile is defined as follows. Agents who have not deviated in the past accept matches with members of the same or higher ranking groups but not with members of lower ranking groups. Thus, on the equilibrium path, new matches form only between members of the same group, since the higher ranking agent rejects the match when members of different groups are provisionally matched. If an agent has accepted a match with a member of a lower ranking group in the past, then the agent continues to accept matches with members of that group in the future, allowing matches to form between members of different groups. Within each match, agents choose a level of cooperation that depends on the groups and past match sets of both partners to the match.

A complication arises from this informal definition. Since each agent's behaviour depends on her past match history, an agent's optimal action in a given period may depend on the past match histories of her partners. It turns out that the informally defined segregated strategy profile described above may not be optimal for agents who have multiple partners who have deviated from the segregated strategy profile in the past.

I deal with this issue by defining the actions taken under a segregated strategy profile only for agents who have at most one partner who has deviated from the strategy profile in the past. More formally, I say that an agent i follows the informal definition of a segregated strategy profile above if the agent's match profile μ_i is such that the agent has at most one partner or provisionally matched partner j whose past match history $\mathcal{H}_j$ includes any groups ranked lower than j's group g_j. Notice that an agent j can have a past match history that includes groups ranked below g_j if and only if agent j has previously deviated from the segregated strategy profile. Consistency of beliefs implies that each agent believes that the probability of meeting a future partner who has previously deviated from the segregated strategy profile is 0. Therefore, the event that an agent has more than one partner who has deviated from the segregated strategy profile can be ignored when calculating continuation values, and so it is not necessary to specify continuation strategies at these events to show that a segregated equilibrium exists.

Formally, a segregated strategy profile is any strategy profile that satisfies the following conditions:

Segregated strategy profile: For each agent i, let $\mathcal{M}_i$ be the set of match profiles μ_i such that if $g_k \in \mathcal{H}_j$, then $g_k \leq g_j$ for all but at most one of i's partners j. A segregated strategy profile is any strategy profile that satisfies the following:

1. An agent i with group and past match set $(g_i, \mathcal{H}_i)$ and match profile μ_i accepts a match with a partner j with group and past match set $(g_j, \mathcal{H}_j)$ if $\mu_i \in \mathcal{M}_i$, and if either $g_j \in \mathcal{H}_i$ or $g_j \leq g_i$.
2. An agent i with group and past match set $(g_i, \mathcal{H}_i)$ and match profile μ_i rejects a match with a partner j with group and past match set $(g_j, \mathcal{H}_j)$ if $\mu_i \in \mathcal{M}_i$, and if both $g_j \notin \mathcal{H}_i$ and $g_j > g_i$.
3. An agent i with group and past match set $g_i, \mathcal{H}_i$ matched with a partner j with group and past match set $(g_j, \mathcal{H}_j)$ proposes a level of cooperation $\bar{a}_S(g_i, \mathcal{H}_i, g_j, \mathcal{H}_j)$.
4. If an agent i with match profile $\mu_i \in \mathcal{M}_i$ is in a partnership with proposed level of cooperation $a \leq \bar{a}_S(g_i, \mathcal{H}_i, g_j, \mathcal{H}_j)$ as described in the previous step, then the agent chooses to cooperate. If an agent i with match profile $\mu_i \in \mathcal{M}_i$ is in a partnership with proposed level of cooperation $a > \bar{a}_S(g_i, \mathcal{H}_i, g_j, \mathcal{H}_j)$, then the agent chooses to cheat.

1.5.2 Value functions under the segregated strategy profile

It is useful to define the values of certain states under a segregated strategy profile. I begin by defining $\gamma(g, \mathcal{H})$ by

$\gamma(g_i, \mathcal{H}_i)$=The number of groups g_j such that, under a segregated strategy profile, $m(g_i, \mathcal{H}_i, \mu_i) = A$ and $m(g_j, \{g_j\}, \mu_j) = A$ for all μ_i and μ_j such that $\mu_i \in \mathcal{M}_i$ and $\mu_j \in \mathcal{M}_j$

That is, $\gamma(g, \mathcal{H})$ is the number of groups with whom an agent with group and past match set $(g, \mathcal{H})$ expects to form matches if all other agents follow the segregated strategy profile. For example, a member of group 1 with past match set $\{1, 2\}$ expects to form matches with members of groups 1 and 2 under the segregated strategy profile, so $\gamma(1, \{1, 2\}) = 2$. On the other hand, a member of group 2 with past match set $\{1, 2\}$ expects to form matches only with other members of group 2, so $\gamma(2, \{1, 2\}) = 1$.

I will be particularly interested in the maximum individually rational value that an agent who expects to form matches with γ groups can achieve from a filled and unfilled relationship slot, if all other agents follow a segregated strategy profile. Implicitly define $V^m(\gamma)$, $V^u(\gamma)$, and $\bar{a}(\gamma)$ as the solutions to the following set of equations:

$$
\begin{aligned}
V^m(\gamma) &= (1-\delta)v(\bar{a}(\gamma)) + pV^u(\gamma) + (1-p)\delta V^m(\gamma) \\
&= (1-\delta)d(\bar{a}(\gamma)) + V^u(\gamma) \\
V^u(\gamma) &= -(1-\delta)c + \frac{\gamma}{G}\delta V^m(\gamma) + \frac{G-\gamma}{G}V^u(\gamma)
\end{aligned}
$$

The level of cooperation $\bar{a}(\gamma)$ is defined to be the maximum level of cooperation that is individually rational for an agent who forms matches with γ groups, and who expects to cooperate at level $\bar{a}(\gamma)$ in all of her matches. The value $V^m(\gamma)$ is the value of a filled relationship slot for such an agent at the beginning of a period, and the value $V^u(\gamma)$ is the value of an empty relationship slot. Rearranging the definitions of $V^m(\gamma)$ and $V^u(\gamma)$ yields:

$$
d(\bar{a}(\gamma)) - v(\bar{a}(\gamma)) = (1-p)\frac{G}{\gamma}c \tag{6}
$$

Comparing this equation to (5) shows that $\bar{a}(\gamma) > \bar{a}_B$ for all $\gamma < G$. This is because an agent who expects to form matches with $\gamma < G$ groups expects to pay a higher search cost upon breaking up the relationship and so can support a higher level of cooperation than an agent who expects to form matches with all possible partners.

It is also useful to know how $V^m(\gamma)$ and $V^u(\gamma)$ depend on γ. Rearranging the definitions of $V^m(\gamma)$ and $V^u(\gamma)$ yields the following:

$$
V^m(\gamma) = v(\bar{a}(\gamma)) - p\frac{G}{\gamma}c \tag{7}
$$

Using this expression I can describe how $V^u(\gamma)$ and $V^m(\gamma)$ depend on γ when the search cost c is sufficiently large and the exogenous probability of match breakup p is sufficiently small:

Lemma 3. *There exists $\bar{p}$ such that if $p \leq \bar{p}$ and*

$$
c \geq \frac{1}{\delta(1-p)}[d(\hat{a}) - v(\hat{a})].
$$

Then

1. *$V^m(\gamma)$ is strictly decreasing in γ.*

2. *$V^u(\gamma)$ is strictly increasing in γ.*

3. *$\delta V^m(G) > V^u(G)$.*

The first part of lemma 3 is true for any search cost c. From (6) and the fact that $v(\cdot)$ is concave and $d(\cdot)$ is convex, it follows that $\bar{a}(\gamma)$ is decreasing in γ. From (7) it follows in turn that $V^m(\gamma)$ is decreasing in γ if p is sufficiently small. The third part of lemma 3 is also true for any search cost, and follows directly from the definition of $V^u(G)$. The second part of lemma 3 is more subtle. To see why it is true, suppose to the contrary that for some $\gamma > \gamma'$, $V^u(\gamma) \leq V^u(\gamma')$. From the definition of $\bar{a}(\cdot)$, we have that

$$V^m(\gamma') = (1-\delta)d(\bar{a}(\gamma')) + V^u(\gamma')$$

which implies that

$$V^m(\gamma') \geq (1-\delta)d(\bar{a}(\gamma')) + V^u(\gamma)$$

So, given two matched agents who expect to receive $V^u(\gamma)$ when unmatched, it is individually rational to cooperate at level $\bar{a}(\gamma') > \bar{a}(\gamma)$ when matched. In other words, given a strategy profile in which the members of γ groups choose to accept matches only with each other and to cooperate at level $\bar{a}(\gamma)$ in each period, any two matched agents have a jointly profitable and individually rational joint deviation to a higher level of cooperation. However, using logic similar to the logic in proposition 1, it can be shown that if the search cost c is sufficiently large then there is no such joint deviation. So when the search cost c is sufficiently large, it must be the case that $V^u(\gamma) > V^u(\gamma')$ for all $\gamma > \gamma'$.

Putting together the pieces of lemma 3 yields the following corollary:

Corollary 1. *There exists $\bar{p}$ such that if $p \leq \bar{p}$ and*

$$c \geq \frac{1}{\delta(1-p)}[d(\hat{a}) - v(\hat{a})].$$

Then

$$V^u(1) < ... < V^u(G) < \delta V^m(G) < ... < \delta V^m(1)$$

1.5.3 Levels of cooperation in the segregated equilibrium

Using the value functions defined in the previous section, I can describe the levels of cooperation chosen in a segregated equilibrium if an equilibrium exists, as follows:

Lemma 4. *If a segregated equilibrium exists, then the equilibrium level of cooperation chosen in a match between an agent i with group and past match set $(g_i, \mathcal{H}_i)$ and an agent j with group and past match set $(g_j, \mathcal{H}_j)$ is*

$$\bar{a}_S(g_i, \mathcal{H}_i, g_j, \mathcal{H}_j) = \min\{\bar{a}(\gamma(g_i, \mathcal{H}_i)), \bar{a}(\gamma(g_j, \mathcal{H}_j))\}$$

Lemma 4 states that in a segregated equilibrium each pair of matched agents chooses the highest level of cooperation that is individually rational for both agents. Intuitively, if the level of cooperation were higher, then at least one agent would be able to profit by cheating, while if the level of cooperation were lower than both agents would have a profitable joint deviation to a higher level of cooperation that would still be individually rational for each of them.

Using lemma 4, I can describe the expected utility that an agent gets from a filled and an empty relationship slot in a segregated equilibrium, if an equilibrium exists:

LEMMA 5. *If a segregated equilibrium exists, then in a segregated equilibrium the expected utility that an agent i with group and past match set $(g_i, \mathcal{H}_i)$ and match profile $\mu_i \in \mathcal{M}_i$ receives from an empty relationship slot is $V^u(\gamma(g_i, \mathcal{H}_i))$. Suppose that for some agent j with group and past match set $(g_j, \mathcal{H}_j)$, $\mathcal{H}_j$ does not contain any groups ranked lower than g_j and $\mathcal{H}_i$ contains g_j. If a segregated equilibrium exists, then in a segregated equilibrium the expected utility that agent i receives from a relationship slot in which she is matched with agent j at the beginning of a period is $V^m(\gamma(g_i, \mathcal{H}_i))$.*

Lemmas 4 and 5 both depend on the fact that in a segregated equilibrium, an agent i with group and past match set $(g_i, \mathcal{H}_i)$ and match profile $\mu_i \in \mathcal{M}_i$ expects to cooperate at level $\bar{a}(\gamma(g_i, \mathcal{H}_i))$ in all of her matches with future partners, regardless of the level of cooperation that she achieves in her matches with her present partners. Consistency of beliefs ensures that agents do in fact have this expectation in a segregated equilibrium. First, consistency of beliefs implies that each agent expects that each of her future provisional matches j will have group and past match set $(g_j, \{g_j\})$ for some group g_j. In addition, the match will form if and only if $g_j \geq g_i$ and $g_j \in \mathcal{H}_i$. This implies that if future partner j has match profile $\mu_j \in \mathcal{M}_j$, then agents i and j will choose level of cooperation $\bar{a}(\gamma(g_i, \mathcal{H}_i))$ and will choose to cooperate at that level for the duration of their match. Second, consistency of beliefs implies that agent i believes that all of her future partners j will have match profiles $\mu_j \in \mathcal{M}_j$. The value to an agent of an unfilled relationship slot in which the agent expects to cooperate at level $\bar{a}(\gamma(g_i, \mathcal{H}_i))$ in all future matches is $V^u(\gamma(g_i, \mathcal{H}_i))$, which implies the first part of lemma 5. The value to an agent of a filled relationship slot in which an agent expects to cooperate at level $\bar{a}(\gamma(g_i, \mathcal{H}_i))$ both in her current match and in all future matches is $V^m(\gamma(g_i, \mathcal{H}_i))$, which implies the second part of lemma 5.

1.5.4 Existence of a segregated equilibrium

Using lemmas 3, 4 and 5, I can prove the following proposition:

PROPOSITION 2. *Fix $v(\cdot)$, $d(\cdot)$, c, and δ. There exists $\bar{p}$ such that if $p \leq \bar{p}$ and*

$$c \geq \frac{1}{\delta(1-p)}[d(\hat{a}) - v(\hat{a})] \tag{8}$$

then there exists $\bar{N}$ such that for all $N > \bar{N}$, a segregated equilibrium exists. If a segregated equilibrium exists, then the equilibrium level of cooperation chosen by an agent i with group and past match set $(g_i, \mathcal{H}_i)$ matched *with a partner j with group and past match set $(g_j, \mathcal{H}_j)$ is*

$$\bar{a}_S(g_i, \mathcal{H}_i, g_j, \mathcal{H}_j) = \min\{\bar{a}(\gamma(g_i, \mathcal{H}_i)), \bar{a}(\gamma(g_j, \mathcal{H}_j))\} \tag{9}$$

The complete proof of proposition 2 is in the appendix, but I outline the main idea of the proof here. A segregated equilibrium exists if four conditions are satisfied. First, it must not be possible for any agent to profit by cheating in any relationship. Second, no pair of matched agents can have a mutually profitable and credible joint deviation to a higher level of cooperation. Third, agents must prefer to accept matches with members of the same or higher ranked groups, and with members of groups with whom they have matched in the past. Fourth, agents must prefer to reject matches with members of lower ranked groups, if they have not accepted matches with members of those groups in the past. These conditions must hold for all agents i with match profiles $\mu_i \in \mathcal{M}_i$.

Lemma 4 implies that the first condition holds, and an argument similar to the argument in the proof of proposition 1 implies that the second condition holds for c sufficiently large, so it remains to show only that the third and fourth conditions hold. First, consider whether an agent prefers to accept a match with a member of the same or higher ranking group, or with a member of groups with whom she has matched in the past. Consider an agent i with group and past match set $(g_i, \mathcal{H}_i)$ who is matched with $N-1$ partners j in inactive relationship slots, and who is provisionally matched in the active relationship slot with a partner k with group g_k such that either $g_k \leq g_i$ or $g_k \in \mathcal{H}_i$. Suppose that agent i's match profile μ_i satisfies $\mu_i \in \mathcal{M}_i$. The maximum value that agent i can get from the active relationship slot if she rejects the provisional match is $V^u(G)$. The minimum value that agent i can get from having a filled relationship slot at the beginning of a period is $V^m(G)$, and so, applying the discount factor, the minimum value that agent i can get from the active relationship slot by accepting the provisional match with partner k is $\delta V^m(G)$. The value that the agent receives from her inactive relationship slots is unaffected by the action she takes in the active relationship slot. Since $\delta V^m(G) > V^u(G)$ by lemma 3, agent i always prefers to accept the match with partner k.

Now consider whether an agent prefers to reject a match with a member of a lower ranking group with who the agent has not interacted in the past. That is, consider again an agent i with group and past match set $(g_i, \mathcal{H}_i)$ who is matched with $N-1$ partners j in inactive relationship slots. This time, suppose that the agent is provisionally matched in the active relationship slot with a partner k with group g_k such that $g_k > g_j$ and $g_k \notin \mathcal{H}_i$, and suppose again that $\mu_i \in \mathcal{M}_i$. Now, accepting a match with partner k affects the value that agent i expects to receive from her matches with her other partners j. If agent i rejects the match with partner k, then player i will get value $V^m(\gamma(g, \mathcal{H}))$ from all but at most one of her currently inactive relationship slots at the beginning of the next period in which each relationship slot is active. On the other hand, if she accepts the match with partner k, then player i will get value $V^m(\gamma(g, \mathcal{H}) + 1)$ from all but at most one of her other relationship slots at the beginning of the next period in which each relationship slot is active. The minimum value that agent i can get from an empty relationship slot is $V^u(1)$. The minimum value that agent i can get from her relationship with her one partner j whose past match set may include groups ranked lower than j's group g_j, at the beginning of the next period in which that relationship slot is active, is $V^m(G)$. Thus, the minimum total utility that the agent can get from rejecting the match with partner k is:

$$V^u(1) + \delta^{1/N} V^m(G) + \sum_{t=2}^{N-1} \delta^{t/N} V^m(\gamma(g, \mathcal{H})).$$

The maximum value that agent i can get from a relationship slot in which she forms a match with partner k is $V^m(\gamma(g,\mathcal{H})+1)$, which is also the maximum value that the agent can get from all of her other relationship slots if she accepts the match with partner k. Thus the maximum total utility that the agent can get from accepting the match with partner k is

$$\sum_{t=1}^{N} \delta^{t/N} V^m(\gamma(g, \mathcal{H}) + 1). \tag{10}$$

Putting these expressions together, agent i prefers to reject the match with partner k if

$$V^u(1) + \delta^{1/N} V^m(G) + \sum_{t=2}^{N-1} \delta^{t/N} V^m(\gamma(g, \mathcal{H})). \geq \sum_{t=1}^{N} \delta^{t/N} V^m(\gamma(g, \mathcal{H}) + 1) \tag{11}$$

Since $V^m(\gamma(g,\mathcal{H}) + 1) < V^m(\gamma(g,\mathcal{H}))$ for all $(g,\mathcal{H})$ by lemma 3, (11) is satisfied if N is sufficiently large. Therefore, for sufficiently large N each agent prefers to reject matches with members of lower ranking groups with whom the agent has not interacted in the past.

The intuition behind proposition 2 is as follows. An agent who accepts a match with a member of a higher ranking group achieves a lower level of cooperation in that relationship slot than she could achieve if she were matched with a member of her own group. However, accepting a match with a member of a higher ranking group does not affect the level of cooperation that an agent can achieve in any of her other relationship slots. In contrast, accepting a match with a member of a lower ranking group reduces the level of cooperation that the agent can achieve in all of her relationship slots. Thus the penalty for accepting a match with a member of a lower ranking group is larger than the penalty for accepting a match with a member of a higher ranking group. As the number of relationship slots N grows, the penalty for accepting a match with a member of a lower ranking group increases while the penalty for accepting a match with a member of a higher ranking group stays the same, and so for N sufficiently large the conditions for the existence of the segregated equilibrium are satisfied.[5]

[5] The threshold value $\bar{N}$ above which a segregated equilibrium exists depends (among other things) on the curvature of the $v(\cdot)$ and $d(\cdot)$ functions. If $v(\cdot)$ and $d(\cdot)$ are close to linear, then the level of cooperation that an agent can sustain falls quickly as her past match set grows. In this case it is optimal for an agent to refuse to interact with members of lower ranking groups even if N is small. In contrast, if $v(\cdot)$ and $d(\cdot)$ have high curvature, then the level of cooperation that an agent can sustain does not fall much

1.6 Welfare

The following proposition describes welfare under the segregated equilibrium:

Proposition 3. *If the exogenous probability that a relationship breaks up p is sufficiently small, then an agent's expected utility at the beginning of the game is higher under the segregated equilibrium than under the benchmark equilibrium.*

Proof. On the segregated equilibrium path, $\gamma = 1$ for all agents in all periods. Plugging $\gamma = 1$ into (7) yields that the expected value of each relationship slot for each agent at the beginning of the game is $V^m(1) = v(\bar{a}(1)) - pGc$. The difference between the expected value of a relationship slot under the segregated equilibrium and under the benchmark equilibrium is $V^m(1) - V^m_B = v(\bar{a}(1)) - v(\bar{a}(G)) - p(G-1)c$. As p goes to zero, this expression goes to $v(\bar{a}(1)) - v(\bar{a}(G))$ which is positive since $\bar{a}(\gamma)$ is decreasing in γ from the proof of lemma 3. Thus every relationship slot has higher value under the segregated equilibrium than under the benchmark equilibrium, and so each agent's total utility at the beginning of the game is higher under the segregated equilibrium. □

The segregated equilibrium allows agents to cooperate at a higher level than the benchmark equilibrium, at the cost of requiring agents to pay a higher search cost when matches break up exogenously. As the probability of exogenous relationship break up goes to zero, expected utility under the segregated equilibrium becomes larger than expected utility under the benchmark equilibrium. The fact that welfare is higher under the segregated equilibrium provides a reason to think that the segregated equilibrium would be selected over the benchmark equilibrium, at least when the exogenous probability of match break up is low.

1.7 Other Equilibria

The segregated equilibrium imposes a hierarchical relationship between groups, in which members of higher ranked groups refuse matches with members of lower ranked groups but members of lower ranked groups accept matches with members of higher ranked groups. This linear hierarchy seems to correspond to the real hierarchy that characterizes the Indian caste system. However, the equilibrium is not unique. One way in which the equilibrium fails to be unique is that the segregated equilibrium provides no guidance as to which groups are ranked higher and which groups are ranked lower in the hierarchy. If there is a segregated equilibrium in which group 1 is ranked higher than group 2, then there is also a segregated equilibrium in which group 2 is ranked higher than group 1. This indeterminacy seems to correspond to real indeterminacy in the Indian caste system, in which the relative caste rankings may differ in different places. For example, Mayer (1960) writes 'The Potter,

as her past match set grows. In this case an agent refuses to interact with members of lower ranking groups only if by doing so she stands to lose the ability to cooperate with many other partners, that is, if N is large.

for example, is of much lower standing in Uttar Pradesh than he is in Malwa'. (Mayer, 1960, p. 27) The relative ranking of the Potter caste apparently varies across locations in India.

Another way in which the segregated equilibrium fails to be unique is that the linear hierarchy of the segregated equilibrium is not the only possible pattern of segregation. From the discussion in section 1.4, there does not exist an equilibrium in which all agents reject matches with all groups other than their own after all histories. Moreover, with two groups the segregated equilibrium is essentially the only equilibrium that features group segregation, with the exception of knife-edge cases in which parameter values are such that agents are just indifferent between accepting and rejecting matches with members of other groups. However, with three or more groups there do exist other patterns of segregation that are equilibria. For example, with three groups, if a segregated equilibrium exists then there also exists an equilibrium that is a cycle. In this cyclical equilibrium, on the equilibrium path members of group 1 reject matches with members of group 2, members of group 2 reject matches with members of group 3, and members of group 3 reject matches with members of group 1, while all other matches are accepted. As in the segregated equilibrium, in this cyclical equilibrium agents interact only with members of their own groups on the equilibrium path. Other patterns are possible as well. The linear hierarchy of the segregated equilibrium seems in some sense simpler than these other equilibria, which perhaps provides a reason to think that the segregated equilibrium would be selected over the other equilibria. Admittedly, this is a somewhat weak justification, and it would be interesting to try to find examples of societies with a cyclical or some other pattern of segregation, as these other patterns also seem to be allowed by the model. However, I am not aware of any such societies.

2 Centralized and Decentralized Segregation

So far I have developed a theory of social division in which members of different groups do not interact with each other due to a reputation effect. The reputation effect makes interactions between members of different groups less profitable than interactions between members of the same group, solving the renegotiation-proofness problem in Eeckhout (2006). Akerlof (1976) and Pęski and Szentes (2013) (henceforth APS) solve the renegotiation-proofness problem in a different way. The overall setup of APS is similar to the setup of my model, but the information structures of the two models are different. In APS, agents observe with whom their partners have interacted in the past, with whom their partners' partners have interacted in the past, and so on to infinity, in contrast to my model in which agents observe with whom their partners have interacted in the past and nothing more. Thus agents in APS have much more information than agents in my model. The information structure in APS allows for an equilibrium in which there is a punishment for agents who interact with members of other groups, a punishment for failure to punish such interactions, a punishment for failure to punish failure to punish, and so on, unlike in my model in which there is a punishment for interacting with a member of a different group but no punishment for failure to punish. The infinite regress in APS implies that punishments can never be renegotiated without triggering some further punishment, solving the renegotiation-proofness problem. In this section I discuss whether my model or APS better accounts for features of real Indian castes as described in the anthropological literature.

One way in which my model improves on APS is that my model better accounts for the hierarchical structure of the Indian caste system. In APS, the motivation for members of one group to discriminate is essentially independent of the behaviour of any other group. Agents refuse to interact with members of other groups for fear of punishment by members of their own groups, they inflict those punishments for fear of further punishment by members of their own groups, and so on. Agents' beliefs about how members of other groups will behave are irrelevant for this chain of punishments. Because agents' optimal actions are independent of their beliefs about how members of other groups will behave, APS allow for equilibria in which all groups discriminate against each other symmetrically as well as equilibria in which some groups discriminate while other groups do not. In contrast, as explained in section 1.4, my model does not allow for equilibria with perfectly symmetric discrimination, so my model explains the existence of a group hierarchy in a way that APS do not.

While my model is better than APS at explaining the caste hierarchy, in other respects the evidence distinguishing between my model and APS is more mixed. Some castes seem to have both punishments for interacting with members of other groups and punishments for failure to punish, as in APS, while in other castes the chain of punishments seems to end after the first step, as in my model. An example of a caste that seems well described by APS comes from Majumdar (1958), who describes the norms of the Chamar caste in the state of Uttar Pradesh. Majumdar emphasizes the role of the caste panchayat in enforcing caste norms. The panchayat is a council of caste elders that meets to judge offenses against the rules of the caste and that may also pronounce punishments. Majumdar describes a case where the norms were violated and the response of the panchayat as follows:

> Even if a person gives food or water to an outcaste, or invites him for a smoke, without knowing the stigma attached to the recipient of his kindness, the unwitting offender also relinquishes his membership of the caste.... An instance of this occurred in May 1954, when K-Chamar of Bijapur village visited B-Chamar of Mohana. K-Chamar had been, for some reason or other expelled from his caste by the Chamar *biradari* [that is, the local subcaste] of Bijapur. He came to Mohana without letting anyone know of the disgrace, and B-Chamar as is the custom treated his guest very hospitably, and they took their midday meals together. Soon it was known that K-Chamar was an outcaste. Consequently B-Chamar was declared an outcaste by the Chamar caste-panchayat of Mohana.
>
> (Majumdar, 1958, p. 94)

K-Chamar violates the caste norms and is punished with complete ostracism, and then B-Chamar incurs the same punishment merely for eating a meal with K-Chamar. This chain of punishments seems to correspond to the chain of punishments in APS.

A contrasting example of a caste that seems better described by my model comes from Hayden (1983). Hayden describes norms in the Nandiwalla caste in the state of Maharashtra. Among the Nandiwallas a person who has violated the caste norms is said to be *eli*. The consequences of violating caste norms among the Nandiwallas seem to be less severe than the consequences of violating caste norms among the Chamars. Hayden writes:

> The [Nandiwallas] say that they 'won't give even fire' to one who is *eli*. However, there is a certain literal quality to this pronouncement. They won't give him fire, but they will give him matches. They won't take food with him, but they will certainly drink liquor and take *pa:n* with him. One should not quarrel with someone who is *eli*, but the latter may argue in panchayat. What seems to happen is that, although certain specific commensal activities with other caste members are limited for one who is *eli*, most aspects of his life remain unchanged. He still puts his tent in the same place in both the large triennial encampment and in smaller camps on the road. People come to visit, and he can reciprocate. In most ways, life goes on normally.
>
> (Hayden, 1983, p. 299)

Among the Nandiwallas, people who violate caste norms may continue to interact with other members of the caste, albeit in somewhat limited ways, and there do not seem to be any further punishments for people who interact in these limited ways with *eli* partners. These norms seem to correspond to the strategy in my model, in which agents who interact with members of other groups can continue to interact with members of their own groups at a lower level of cooperation, and in which there is no further punishment for interacting with agents who have interacted with members of other groups.

The difference between Chamar and Nandiwalla social punishments seems to be related to differences between Chamar and Nandiwalla caste institutions. In the Chamar caste, the caste panchayat plays an important role in collecting and disseminating information about violations of caste norms. In the vignette above, the caste panchayat announces that B-Chamar interacted with K-Chamar, thereby triggering B-Chamar's punishment. Without this announcement it seems unlikely that caste members would receive the necessary information, as evidenced by the fact that B-Chamar himself does not receive information about K-Chamar's norm violation in time to punish K-Chamar. In contrast, because the Nandiwallas are nomadic the Nandiwalla panchayat only meets once every three years. As a result, information about norm violations is spread through informal gossip rather than panchayat pronouncements. Hayden writes, '*Eli* is not a status that is *imposed* on a person for his actions. Rather it is an automatic reaction to the fact that one automatically becomes polluted by an improper act.... It does not have to be pronounced by anyone'. (Hayden, 1983, p. 297) Thus it seems that in order for the equilibrium in APS to function, it is necessary to have a centralized institution to collect and broadcast the detailed information required to enforce the required punishment chains. On the other hand, the less stringent information requirements of my model can be satisfied through decentralized gossip. For this reason I refer to the institution described by APS as centralized segregation and to the institution described by my model as decentralized segregation.

3 Conclusion

In this paper I have described a model of segregation in a population of *ex ante* identical agents. By way of conclusion, I reiterate some of the main novel insights from the paper:

1. Group segregation may be enforced through a reputation effect, under which people who interact with members of different groups are believed to be less trustworthy by members of their own groups.

2. The reputation effect endogenously creates a hierarchical relationship between groups, under which members of higher ranking groups refuse to interact with members of lower ranking groups, while members of lower ranking groups are willing to interact with members of higher ranking groups.

3. An alternative method for enforcing group segregation is to require group members to punish people who interact with members of other groups, to punish people who fail to inflict these punishments, to punish failure to punish failure to punish, and so on. This form of group segregation requires the existence of a centralized information-processing institution to keep track of the relevant punishments. In contrast, reputation-based segregation as described in my paper can be sustained through decentralized gossip without formal information-processing institutions.

In my model, group segregation is welfare improving because it increases the level of cooperation in the community. This positive result arises because my model is static and does not consider the effects of group segregation on growth. It is likely that group segregation has negative consequences for growth that are not included in my model. For example, group segregation may reduce growth by slowing the spread of new ideas through the population. These negative effects seem potentially very large, and may help to explain some of the negative relationships between measures of social division and growth that have been found in the literature. Further theoretical and empirical exploration of the relationship between segregation and growth may prove fruitful.

University of Warwick and CAGE

References

[1] Ali, S.N. and Miller, D.A. (2016). 'Ostracism and Forgiveness', *American Economic Review*, vol. 106(8), pp. 2329-2348.

[2] Akerlof, G. (1976). 'The Economics of Caste and of the Rat Race and Other Woeful Tales', *Quarterly Journal of Economics*, vol. 90(4), pp. 599-617.

[3] Anderson, S. (2011). 'Caste as an Impediment to Trade', *American Economic Journal: Applied Economics*, vol. 3(1), pp. 239-263.

[4] Bernheim, B.D., and Ray, D. (1989). 'Collective dynamic consistency in repeated games', *Games and Economic Behavior*, vol. 1(4), pp. 295-326.

[5] Deb, J. (2008). 'Cooperation and Community Responsibility: A Folk Theorem for Repeated Matching Games with Names', Working Paper, Yale University.

[6] Dixit, A.K. (2003). 'Trade Expansion and Contract Enforcement', *Journal of Political Economy*, vol. 111(6), pp. 1293-1317.

[7] Eeckhout, J. (2006). 'Minorities and Endogenous Segregation', *Review of Economic Studies*, vol. 73(1), pp. 31-53.

[8] Ellison, G. (1994). 'Cooperation in the prisoner's dilemma with anonymous random matching', *Review of Economic Studies*, vol. 61(3), pp. 567-588.

[9] Farrell, J. and Maskin, E. (1989). 'Renegotiation in repeated games', *Games and Economic Behavior*, vol. 1(4), pp. 327-360.

[10] Fehr, E., Gächter, S. and Kirchsteiger, G. (1997). 'Reciprocity as a Contract Enforcement Device: Experimental Evidence', *Econometrica*, vol. 65(4), pp. 833-860.

[11] Fudenberg, D. and Tirole, J. (1991). 'Perfect Bayesian Equilibrium and Sequential Equilibrium', *Journal of Economic Theory*, vol. 53(2), pp. 236-260.

[12] Ghosh, P. and Ray, D. (1996). 'Cooperation in Community Interaction Without Information Flows', *Review of Economic Studies* vol. 63(3), pp. 491-519.

[13] Greif, A. (1993). 'Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition', *American Economic Review*, vol. 83(3), pp. 525-548.

[14] Grimard, F. (1997). 'Household consumption smoothing through ethnic ties: evidence from Cote d'Ivoire', *Journal of Development Economics*, vol. 53(2), pp. 391-422

[15] Halac, M. (2012). 'Relational Contracts and the Value of Relationships', *American Economic Review*, vol. 102(2), pp. 750-779.

[16] Hayden, R.M. (1983). 'Excommunication as Everyday Event and Ultimate Sanction: The Nature of Suspension from an Indian Caste', *Journal of Asian Studies*, vol. 42(2), pp. 291-307.

[17] Jackson, M.O., Rodriguez-Barraquer, T. and Tan, X. (2012). 'Social Capital and Social Quilts: Network Patterns of Favor Exchange', *American Economic Review*, vol. 102(5), pp. 1857-1897.

[18] Kandori, M. (1992). 'Social Norms and Community Enforcement', *Review of Economic Studies*, vol. 59(1), pp. 63-80.

[19] Majumdar, D.N. (1958). *Caste and Communication in an Indian Village*, Asia Publishing House: Bombay.

[20] Mayer, A.C. (1960). *Caste and Kinship in Central India*, Berkeley: University of California Press.

[21] Mazzocco, M. and Saini, S. (2012). 'Testing Efficient Risk Sharing with Heterogeneous Risk Preferences', *American Economic Review*, vol. 102(1), pp. 428-468.

[22] Munshi, K. and Rosenzweig, M. (2006). 'Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy', *American Economic Review*, vol. 96(4), pp. 1225-1252.

[23] Munshi, K. and Rosenzweig, M. (2016). 'Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap', *American Economic Review*, 106(1), pp. 46-98.

[24] Nava, F. and Piccione, M. (2014). 'Effciency in Repeated Games with Local Interaction and Uncertain Local Monitoring', *Theoretical Economics*, vol. 9(1), pp. 279-312.

[25] Pęski, M. and Szentes, B. (2013). 'Spontaneous Discrimination', *American Economic Review*, vol. 103(6), pp. 2412-2436.

A Proofs

A.1 Proof of Proposition 1

Plugging equation (3) into the constraint (2) and equation (1) and rearranging yields

$$V_B^m = \max_a \frac{1-\delta}{1-\delta(1-p)} v(a) - \frac{(1-\delta)p}{1-\delta(1-p)} c + \frac{\delta p}{1-\delta(1-p)} V_B^f \tag{12}$$

subject to

$$V_B^f \leq \frac{1}{\delta(1-p)}[v(a) - (1-\delta(1-p))d(a)] + (1-\delta)(1-p)c \tag{13}$$

Recall that $\hat{a}$ was defined as the value of a that solves

$$\max_a v(a) - (1-\delta(1-p))d(a).$$

Since v is strictly concave and d is strictly convex, there exists a finite value of $\hat{a}$ that maximizes a. Since a has a maximum value, there exists $\hat{V}_B^f$ such that the constraint (13) can be satisfied for $a \geq 0$ if and only if $V_B^f \leq \hat{V}_B^f$, with $\hat{V}_B^f$ defined by

$$\hat{V}_B^f = \frac{1}{\delta(1-p)}[v(\hat{a}) - (1-\delta(1-p))d(\hat{a})] + (1-\delta)(1-p)c. \tag{14}$$

Now, define a function $\phi(x)$ by

$$\phi(x) = \max_a \frac{1-\delta}{1-\delta(1-p)} v(a) - \frac{(1-\delta)p}{1-\delta(1-p)} c + \frac{\delta p}{1-\delta(1-p)} x \tag{15}$$

subject to

$$x \leq \frac{1}{\delta(1-p)}[v(a) - (1-\delta(1-p))d(a)] + (1-\delta)(1-p)c \tag{16}$$

Any fixed point of ϕ is a benchmark equilibrium. However, notice that ϕ is not well-defined for all x, since for $x > \hat{V}_B^f$ there is no $a \geq 0$ that satisfies (16). Since v and d are continuous and differentiable, ϕ is continuous and differentiable. By the envelope theorem,

$$\frac{\partial \phi}{\partial x} = \frac{\delta p}{1-\delta(1-p)} - \psi < 1 \tag{17}$$

where $\psi > 0$ is the Lagrange multiplier on the constraint (16). Since $\phi(x) > x$ for x sufficiently small, ϕ has exactly one fixed point if and only if

$$\phi(\hat{V}_B^f) \leq \hat{V}_B^f \tag{18}$$

Plugging in the expression for $\hat{V}_B^f$ from (14) into (18) and rearranging yields the condition that a benchmark equilibrium exists if and only if

$$c \geq \frac{1}{\delta(1-p)}[d(\hat{a}) - v(\hat{a})].$$

This completes the proof.

A.2 Proof of Lemma 3

From the definition of $\bar{a}(\gamma)$ in (6) and the fact that $v(\cdot)$ is convex and $d(\cdot)$ is concave, $\bar{a}(\gamma)$ is decreasing in γ. Since $\bar{a}(\gamma)$ is decreasing in γ and since $V^m(\gamma) = v(\bar{a}(\gamma)) - p\frac{G}{\gamma}c$, if p is sufficiently small and $\gamma < \gamma'$ then $V^m(\gamma) > V^m(\gamma')$. Since $V^u(G) = -(1-\delta)c + \delta V^m(G)$, $V^u(G) < \delta V^m(G)$. Thus it only remains to be shown that if $\gamma < \gamma'$ then $V^u(\gamma) < V^u(\gamma')$.

Define $\phi(x, \gamma)$ by

$$\phi(x,\gamma) = \max_a \frac{1-\delta}{1-\delta(1-p)}v(a) - \frac{(1-\delta)p}{1-\delta(1-p)}c + \frac{\delta p}{1-\delta(1-p)}x$$

subject to the constraint

$$x \leq \frac{1}{\delta(1-p)}[v(a) - (1-\delta(1-p))d(a)] + (1-\delta)(1-p)\frac{G}{\gamma}c$$

If there exists $\bar{x}(\gamma)$ such that $\phi(\bar{x}(\gamma), \gamma) = \bar{x}(\gamma)$, then $\bar{x}(\gamma) = V^m(\gamma)$ and $\bar{a}(\gamma)$ is the solution to the maximization problem for $x = \bar{x}(\gamma)$. This fact can be seen by comparison to the analogous construction of $\phi(x)$ in the proof of proposition 1. From the proof of proposition 1, there exists $\bar{x}(\gamma)$ such that $\phi(\bar{x}(\gamma), \gamma) = \bar{x}(\gamma)$ for all γ if and only if

$$c \geq \frac{1}{\delta(1-p)}[d(\hat{a}) - v(\hat{a})]$$

Suppose that c satisfies this condition. Then $\bar{a}(\gamma)$ solves

$$\max_a \frac{1-\delta}{1-\delta(1-p)}v(a) - \frac{(1-\delta)p}{1-\delta(1-p)}c + \frac{\delta p}{1-\delta(1-p)}\bar{x}(\gamma)$$

subject to the constraint

$$\bar{x}(\gamma) \leq \frac{1}{\delta(1-p)}[v(a) - (1-\delta(1-p))d(a)] + (1-\delta)(1-p)\frac{G}{\gamma}c$$

The solution to this problem is decreasing in $\bar{x}(\gamma)$, and $\bar{a}(\gamma)$ is decreasing in γ, which implies that $\bar{x}(\gamma)$ must be increasing in γ. From the definitions of $V^m(\gamma)$ and $V^u(\gamma)$, $\bar{x}(\gamma) = V^m(\gamma) = \frac{1-\delta}{\delta}\frac{G}{\gamma}c + \frac{1}{\delta}V^u(\gamma)$, so $V^u(\gamma)$ is increasing in γ. This completes the proof.

A.3 Proof of Lemma 4

Suppose that all agents follow the segregated strategy profile and suppose that the segregated strategy profile is an equilibrium. Let $V_S^m(g, \mathcal{H}, g', \mathcal{H}')$ be the value of a relationship slot for an agent i with group and past match set $(g, \mathcal{H})$ matched to an agent j with group and past match set $(g', \mathcal{H}')$, when agent i has match profile $\mu_i \in \mathcal{M}_i$ and agent j has match profile $\mu_j \in \mathcal{M}_j$. Let $V_S^u(g, \mathcal{H})$ be the value of an empty relationship slot for an agent i with group and past match set $(g, \mathcal{H})$, when agent i has match profile $\mu_i \in \mathcal{M}_i$. Consistency of beliefs implies that an agent i believes that she will be provisionally matched with a partner with group and past match set $(g', \{g'\})$ for some group g' whenever she searches for a new partner. Thus, we have:

$$V_S^m(g, \mathcal{H}, g', \mathcal{H}') = (1-\delta)v(\bar{a}_S(g, \mathcal{H}, g', \mathcal{H}')) + pV_S^u(g, \mathcal{H}) + (1-p)\delta V_S^m(g, \mathcal{H}, g', \mathcal{H}') \tag{19}$$

$$V_S^u(g, \mathcal{H}) = -(1-\delta)c + \sum_{g' | g' \geq g \text{ and } g' \in \mathcal{H}} \frac{1}{G}\delta V_S^m(g, \mathcal{H}, g', \{g\} \cup \{g'\}) + \frac{G - \gamma(g, \mathcal{H})}{G} V_S^u(g, \mathcal{H}) \tag{20}$$

Consider first the case of an agent with group and past match set $(g, \{g\})$ for some group g. We can write $V_S^u(g, \{g\})$ as

$$V_S^u(g, \{g\}) = -(1-\delta)c + \frac{1}{G}\delta V_S^m(g, \{g\}, g, \{g\}) + \frac{G-1}{G} V_S^u(g, \{g\})$$

Since the segregated strategy profile is individually rational and bilaterally rational, the level of cooperation $\bar{a}_S(g, \{g\}, g, \{g\})$ solves

$$V_S^m(g, \{g\}, g, \{g\}) = \max_a (1-\delta)v(a) + pV_S^u(g, \{g\}) + (1-p)\delta V_S^m(g, \{g\}, g, \{g\})$$

$$\text{such that } (1-\delta)v(a) + pV_S^u(g, \{g\}) + (1-p)\delta V_S^m(g, \{g\}, g, \{g\}) \geq (1-\delta)d(a) + \delta V_S^u(g, \{g\})$$

But this implies that

$$V_S^m(g, \{g\}, g, \{g\}) = (1-\delta)d(a) + \delta V_S^u(g, \{g\})$$

Plugging in the definition of $V_S^u(g, \{g\})$ and rearranging yields:

$$d(\bar{a}_S(g, \{g\}, g, \{g\})) - v(\bar{a}_S(g, \{g\}, g, \{g\})) = (1-p)\frac{1}{G}c$$

This is just the equation that implicitly defines $\bar{a}(1)$, so $\bar{a}_S(g, \{g\}, g, \{g\}) = \bar{a}(1)$ for all g. Plugging this expression into the expression for $V_S^u(g, \{g\})$ yields that $V_S^u(g, \{g\}) = -(1-\delta)Gc + v(\bar{a}(1)) - pGc$. Moreover, the incentives for an unmatched agent with group and past match set $(g, \{g\} \cup \{g'\})$ such that $g' \leq g$ are exactly the same as the incentives for an unmatched agent with group and past match set $(g, \{g\})$, so $V_S^u(g, \{g\} \cup \{g'\}) = V_S^u(g, \{g\})$ if $g' \leq g$.

Now consider the case of an agent i with group and past match set $(g, \mathcal{H})$ matched with a partner j with group and past match set $(g', \{g\} \cup \{g'\})$, where $g' \geq g$. $\bar{a}_S(g, \mathcal{H}, g', \{g'\})$ is individually rational for agent j if

$$v(\bar{a}_S(g, \mathcal{H}, \{g\} \cup \{g'\})) \geq (1-\delta)d(\bar{a}_S(g, \mathcal{H}, g', \{g\} \cup \{g'\})) + \delta V_S^u(g', \{g\} \cup \{g'\})$$

Plugging in the value of $V_S^u(g', \{g\} \cup \{g'\})$ derived above yields that $\bar{a}_S(g, \mathcal{H}, g', \{g\} \cup \{g'\})$ is individually rational for agent j if $\bar{a}_S(g, \mathcal{H}, g', \{g\} \cup \{g'\}) \leq \bar{a}(1)$. Therefore, $\bar{a}_S(g, \mathcal{H}, g', \{g\} \cup \{g'\})$ is individually and bilaterally rational for both agents if $\bar{a}_S(g, \mathcal{H}, g', \{g\} \cup \{g'\})$ solves

$$V_S^m(g, \mathcal{H}, g', \{g\} \cup \{g'\}) = \max_a (1-\delta)v(a) + pV_S^u(g, \mathcal{H}) + (1-p)\delta V_S^m(g, \mathcal{H}, g', \{g\} \cup \{g'\})) \tag{21}$$

$$\text{such that } V_S^m(g, \mathcal{H}, g', \{g\} \cup \{g'\}) \geq -(1-\delta)d(a) + \delta V_S^u(g, \mathcal{H}) \tag{22}$$

$$\text{and } a \leq \bar{a}(1) \tag{23}$$

The first constraint is the individual rationality constraint for agent i and the second constraint is the individual rationality constraint for agent j. Since this problem is the same for all g' such that $g' \geq g$, $\bar{a}_S(g, \mathcal{H}, g', \{g\} \cup \{g'\})$ is the same for all g' such that $g' \geq g$, which implies that I can write

$$V_S^u(g,\mathcal{H}) = -(1-\delta)c + \frac{\gamma(g,\mathcal{H})}{G}\delta V_S^m(g,\mathcal{H},g',\{g\}\cup\{g'\}) + \frac{G-\gamma(g,\mathcal{H})}{G}V_S^u(g,\mathcal{H}) \tag{24}$$

Plugging (24) into (21) and rearranging yields that $\bar{a}_S(g,\mathcal{H},g',\{g\}\cup\{g'\})$ solves

$$v(\bar{a}_S(g,\mathcal{H},g',\{g\}\cup\{g'\})) - d(\bar{a}_S(g,\mathcal{H},g',\{g\}\cup\{g'\})) = (1-p)\frac{\gamma(g,\mathcal{H})}{G}c$$

This is just the equation that implicitly defines $\bar{a}(\gamma(g,\mathcal{H}))$, so $\bar{a}_S(g,\mathcal{H},g',\{g\}\cup\{g'\}) = \bar{a}(\gamma(g,\mathcal{H}))$ for all $g,\mathcal{H}$, and g' such that $g' \geq g$. Plugging this into the expression for $V_S^u(g,\mathcal{H})$ yields that $V_S^u(g,\mathcal{H}) = -(1-\delta)\frac{G}{\gamma(g,\mathcal{H})}c + v(\bar{a}(\gamma(g,\mathcal{H}))) - p\frac{G}{\gamma(g,\mathcal{H})}c$

Finally, consider and agent i with group and past match set $(g,\mathcal{H})$ matched with a partner j with group and past match set $(g',\mathcal{H}')$. $\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')$ satisfies individual rationality for agent i if

$$v(\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')) \geq (1-\delta)d(\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')) + \delta V_S^u(g,\mathcal{H})$$

Plugging in the value for $V_S^u(g,\mathcal{H})$ derived above and rearranging yields that $\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')$ is individually rational for agent i if $\bar{a}_S(g,\mathcal{H},g',\mathcal{H}') \leq \bar{a}(\gamma(g,\mathcal{H}))$. A similar argument shows that $\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')$ is individually rational for agent j if $\bar{a}_S(g,\mathcal{H},g',\mathcal{H}') \leq \bar{a}(\gamma(g',\mathcal{H}'))$. Thus $\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')$ is individually and bilaterally rational for both agents if $\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')$ solves

$$\begin{aligned}
&V_S^m(g,\mathcal{H},g',\mathcal{H}') = \max_a (1-\delta)v(a) + pV_S^u(g,\mathcal{H}) + (1-p)\delta V_S^m(g,\mathcal{H},g',\mathcal{H}') \\
&\text{such that } a \leq \bar{a}(\gamma(g,\mathcal{H})) \\
&\text{and } a \leq \bar{a}(\gamma(g',\mathcal{H}'))
\end{aligned}$$

Plugging in the value of $V_S^u(g,\mathcal{H})$ derived above and rearranging yields that $\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')$ solves

$$d(\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')) - (\bar{a}_S(g,\mathcal{H},g',\mathcal{H}')) = (1-p)\frac{\min\{\gamma(g,\mathcal{H}),\gamma(g',\mathcal{H}')\}}{G}c$$

This is just the expression that implicitly defines $\bar{a}_S(g,\mathcal{H}),g',\mathcal{H}')$. So $\bar{a}_S(g,\mathcal{H},g',\mathcal{H}') = \min\{\bar{a}_S(\gamma(g,\mathcal{H}),\gamma(g',\mathcal{H}')\}$. This completes the proof.

A.4 Proof of Lemma 5

Consider an agent i with group and past match set $(g_i,\mathcal{H}_i)$. Consistency of beliefs implies that agent i expects that all of her future provisional matches will have group and past match set $(g_j,\{g_j\})$ for some group j. Consistency of beliefs also implies that agent i believes that all of those future provisional matches will have match profiles $\mu_j \in \mathcal{M}_j$. If both agents follow the segregated strategy profile, then the match between agent i and j will form if and only if $g_j \geq g_i$ and $g_j \in \mathcal{H}_i$, and by lemma 4, the agents will cooperate at level $\bar{a}(\gamma(g_i,\mathcal{H}_i))$. The expected utility that agent i receives from an unfilled relationship slot in which she expects to cooperate at level $\bar{a}(\gamma(g_i,\mathcal{H}_i))$ in all future periods is $V^u(g_i,\mathcal{H}_i)$, which establishes the first part of the lemma.

Now suppose that agent i is presently matched with a partner with group and past match set $(g_j,\mathcal{H}_j)$ such that $\mathcal{H}_j$ does not include any groups ranked below g_j and such that $g_j \in \mathcal{H}_i$. Consistency of beliefs implies that agent i believes that agent j has match profile $\mu_j \in \mathcal{M}_j$. Lemma 4 then implies that agent i expects to cooperate at level $\bar{a}(\gamma(g_i,\mathcal{H}_i))$ for the duration of the present relationship. By the previous paragraph, agent i also expects to cooperate at level $\bar{a}(\gamma(g_i,\mathcal{H}_i))$ in all future matches in the relationship slot, which implies that the expected utility that agent i agent receives from the relationship slot at the beginning of the period is $V^m(\gamma(g_i, \mathcal{H}_i))$. This establishes the second part of the lemma.

A.5 Proof of Proposition 2

I begin by formally restating the conditions for the existence of a segregated equilibrium presented informally in section 1.5. Given a segregated strategy profile, let $V_S^m(g, \mathcal{H}, g', \mathcal{H}')$ be the value that an agent with group and past match set $(g, \mathcal{H})$ receives from a relationship slot in which she is matched with a partner with group and past match set $(g', \mathcal{H}')$. Let $V_S^u(g, \mathcal{H})$ be the value of an empty relationship slot for an agent with group and past match set $(g, \mathcal{H})$. Let $V_S^f(g, \mathcal{H})$ be the value that an agent with group and past match set $(g, \mathcal{H})$ expects to receive from a match with any future relationship partner. Note that lemma 4 implies that under the segregated strategy profile, an agent expects to receive the same value from any partner with whom the agent expects to be matched in any future period with positive probability, which is why $V_S^f(g, \mathcal{H})$ does not depend on the group or past match set of the partner with whom the agent expects to be matched. A segregated strategy profile exists that is an equilibrium if $V_S^f(g, \mathcal{H}) = V_S^m(g, \mathcal{H}, g', \mathcal{H}')$ for all g' such that $g' \leq g$ or $g' \in \mathcal{H}$, and if the following conditions hold:

1. For all $g, \mathcal{H}, g', \mathcal{H}'$, $\bar{a}_S(g, \mathcal{H}, g', \mathcal{H}')$ solves

$$\begin{aligned}
&V_S^m(g, \mathcal{H}, g', \mathcal{H}') = \max_a (1-\delta)v(a) + pV_S^u(g, \mathcal{H}) + (1-p)\delta V_S^m(g, \mathcal{H}, g', \mathcal{H}') \\
&\text{such that } (1-\delta)v(a) + pV_S^u(g, \mathcal{H}) + (1-p)\delta V_S^m(g, \mathcal{H}, g', \mathcal{H}') \geq (1-\delta)d(a) + \delta V_S^u(g, \mathcal{H}) \\
&\text{and } (1-\delta)v(a) + pV_S^u(g', \mathcal{H}') + (1-p)\delta V_S^m(g', \mathcal{H}', g, \mathcal{H}) \geq (1-\delta)d(a) + \delta V_S^u(g', \mathcal{H}')
\end{aligned}$$

   where

$$V_S^u(g, \mathcal{H}) = -(1-\delta)c + V_S^f(g, \mathcal{H})$$

2. For all $g, \mathcal{H}$ and for all $g', \mathcal{H}'$ such that $g' \leq g$ or $g' \in \mathcal{H}$,

$$\delta V^m(g, \mathcal{H}, g', \mathcal{H}') \geq V^u(g, \mathcal{H})$$

3. For all $g, \mathcal{H}$, for all $g', \mathcal{H}'$ such that $g' > g$ and $g' \notin \mathcal{H}$

$$V_S^u(g, \mathcal{H}) + \delta^{1/N} V_S^m(g, \mathcal{H}, 1, \{1, ..., N\}) + \sum_{t=2}^{N-1} \delta^{t/N} V_S^m(g, \mathcal{H}, g, \{g\}) \geq \sum_{t=1}^{N} \delta^{t/N} V_S^m(g, \mathcal{H} \cup \{g'\}, g, \{g\})$$

Condition 1 states that no agent has a profitable individual deviation and that no pair of matched agents have a profitable joint deviation in any relationship. Condition 2 states that an agent with group and past match set $(g, \mathcal{H})$ prefers to accept a match with a partner with group and past match set $(g', \mathcal{H}')$ if $g' \leq g$ or $g' \in \mathcal{H}$. Condition 3 states that an agent with group and past match set $(g, \mathcal{H})$ and match profile $\mu_i \in \mathcal{M}_i$ prefers to reject a match with a partner with group and past match set $(g', \mathcal{H}')$ if $g' > g$ and $g' \notin \mathcal{H}$. The left hand side of condition 3 is the minimum utility that an agent with group and past match set $(g, \mathcal{H})$ and match profile $\mu_i \in \mathcal{M}_i$ could receive from rejecting a potential match. The right hand side of condition 3 is the maximum utility that agent with group and past match set $(g, \mathcal{H})$ and match profile $\mu_i \in \mathcal{M}_i$ could receive from accepting a match with a partner with group and past match set $(g', \mathcal{H}')$ such that $g' > g$ and $g' \notin \mathcal{H}$. Further details on the derivation of this inequality are in the text of section 1.5.

The argument in the text of section 1.5 shows that condition 2 is satisfied for all N and that there exists $\bar{N}$ such that for $N > \bar{N}$, condition 3 is satisfied. Thus it only remains to show that condition 1 is satisfied.

Without loss of generality, suppose that the first constraint in condition 1 binds. Following the same steps as in the proof of proposition 1, I can rewrite the equations in condition 1 as:

$$V_S^m(g, \mathcal{H}, g', \mathcal{H}') = \max_a (1-\delta)v(a) + pV_S^u(g, \mathcal{H}) + (1-p)\delta V_S^m(g, \mathcal{H}, g', \mathcal{H}')$$

subject to

$$V_S^f(g, \mathcal{H}) \leq \frac{1}{\delta(1-p)}[v(a) - (1-\delta(1-p))d(a)] + (1-\delta(1-p))\frac{G}{\gamma(g, \mathcal{H})}c$$

Then, continuing to follow the steps in the proof of proposition 1, I eventually get that there exists a level of cooperation a that satisfies condition 1 if and only if

$$\frac{G}{\gamma(g, \mathcal{H})}c \geq \frac{1}{\delta(1-p)}[d(\hat{a}) - v(\hat{a})]$$

Since $\gamma(g, \mathcal{H}) \leq G$, condition 1 can be satisfied if and only if

$$c \geq \frac{1}{\delta(1-p)}[d(\hat{a}) - v(\hat{a})]$$

This condition holds by assumption, completing the proof.