February 2013 No.115

# A Theory of Cooperation through Social Division, with Evidence from Nepal

James P. Choy

## WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

# A Theory of Cooperation through Social Division, with Evidence from Nepal*

James P. Choy†

November 12, 2012

**Abstract**

Informal, kin-based groups play an important role in developing country economies. I point out two facts: communities are divided into smaller groups, and many groups prohibit interactions with outsiders. These facts are rationalized in a model in which division of the community into non-interacting groups allows agents to support higher levels of cooperation. Group segregation is sustained in equilibrium through a reputation effect. I test the empirical implication that there should be less cooperation between members of groups that make up a larger percentage of their communities. I discuss implications for underinvestment in education, misallocation of resources, and institutional change.

**Keywords:** Cooperation, Caste, Social Institution
**JEL Classification Numbers:** C7, O12, O17

*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, and Joe Vavra. Seminar participants at Yale University, the University of Warwick, the University of Namur, and the NEUDC provided helpful comments. I acknowledge research funding from the Yale Economic Growth Center.

†Department of Economics, University of Warwick. E-mail: j.choy@warwick.ac.uk

# 1 Introduction

Recent research has shown that informal, kin-based groups play an important role in regulating the economies of developing countries. These groups have different names in different places, such as castes, tribes, or clans, but everywhere they serve some of the same functions. Some functions that have been discussed in the literature include enforcing trade contracts (Greif 1993), providing mutual insurance (Grimard 1997, Munshi and Rosenzweig 2009, Mazzocco and Saini 2012), and connecting job seekers to employment opportunities (Munshi and Rosenzweig 2006).

The previous list of functions raises a puzzle. For each of the functions listed, there is a natural argument that the function is most efficiently performed by a large group. If the group regulates trade, then a large group maximizes the gains from trade. If the group provides mutual insurance, then a large group better diversifies risk. If the group facilitates job matches, then a large group will have information about more job seekers and more opportunities and so will be able to generate better matches. For this reason, one might think that there would be economic pressure for one group to expand to include the entire community. However, this does not seem to happen.

Here is another puzzle. Many groups have a social norm that prohibits certain kinds of interactions between group members and outsiders. Group members who violate the norm are subject to various kinds of social penalties. For example, in the Nepalese caste system, which will be my primary example in the empirical section of this paper, members of one caste are often prohibited from marrying, socializing with, or in some cases even touching members of other castes.[1] In a Walrasian framework, however, any norm that constricts the choice set, in this case the choice set of potential relationship partners, can only be welfare reducing. So it is difficult to understand what purpose this norm serves, and how it can persist.[2]

In this paper, I develop a framework to try to resolve these puzzles. I construct a model in which there are a large number of agents who search for relationship partners. These relationships potentially benefit both partners but are also vulnerable to cheating, and so they can be modeled using a version of the repeated prisoner's dilemma. Into this framework, I introduce the two stylized facts discussed above, that society is divided into groups and that there is a norm prohibiting agents from forming relationships with members of

[1] Höfer (2004) describes the Muluki Ain, a document that had the status of constitutional law in Nepal until 1963, and that is largely concerned with delineating the various ways in which members of different castes are and are not permitted to interact. Caplan (1970) is a good general reference for information on the Nepalese caste system.

[2] While the South Asian caste system is perhaps the best known example of a system of social norms that prohibits interactions between members of different groups, it is not the only example. Weyrauch and Bell (1993) study traditional law in Gypsy communities in the United States. One punishable offense in Gypsy law is "familiarity with the *gaje*", i.e. non-Gypsies. Berman (2000) discusses social norms among ultra-orthodox Jews in Israel. He argues that the extremely strict interpretation of religious law among the ultra-orthodox effectively makes it impossible for the ultra-orthodox to interact in certain ways with non-ultra-orthodox Jews without breaking the law. Related to my thesis, Berman emphasizes the high levels of cooperation within the ultra-orthodox community. The plot of Shakespeare's *Romeo and Juliet* depends on a social norm prohibiting interactions between the Montagues and the Capulets. The protagonists of the play defy the social norm, and, as predicted by my model, they suffer bad consequences. (I thank Chris Udry for suggesting this example.)

different groups. This group segregation makes it more difficult for any agent to find a new relationship if her current relationship breaks up, compared to the situation without segregation. Thus, the outside option to any given relationship is lower, and so it is possible to sustain a higher level of cooperation within any given relationship. If the benefits of cooperation are sufficiently important, then group segregation is welfare improving compared to the situation where groups are irrelevant. I argue that this can help explain why groups and the group segregation norm are able to persist.

A natural question is why agents are willing to follow the social norm. I assume that while agents cannot observe what their current partners have done within their past relationships, they can observe with whom their current partners have interacted with in the past. In addition, I suppose that there is a hierarchy of groups. In the equilibrium I construct, agents are willing to interact only with members of their own or higher ranking groups, and agents actively search for relationship partners only within their own groups. If an agent deviates from the equilibrium strategy profile and interacts with a member of a lower ranking group, all other agents come to believe that the deviating agent will continue to search over all lower ranking groups in the future. The cost of searching over a larger subset of the community is lower, and so the deviating agent is believed to have a higher outside option to any relationship. Thus, the deviating agent can sustain a lower level of cooperation in her future relationships. This penalty for interacting with members of lower ranking groups maintains the group segregation norm in equilibrium, even though all agents are rational and care only about material outcomes and there are no intrinsic differences between groups. Moreover, the reputation effect in my model appears even though there is no incomplete information about types. In this respect my model is similar to Greif (1993).

In the literature, the theoretical paper that is most closely related to mine is Eeckhout (2006). Like my model, Eeckhout's model introduces payoff irrelevant groups into a setting in which agents search over the community to find relationship partners, and Eeckhout shows that allowing each agent to condition her behavior on her partner's group membership can be welfare improving. Unlike my model, Eeckhout's model does not contain any mechanism that enforces differential treatment of same-group versus other-group partners. As a result, Eeckhout's equilibrium does not satisfy a natural renegotiation-proofness condition called bilateral rationality, introduced by Ghosh and Ray (1996). Because my model does contain a mechanism that enforces group segregation, my equilibrium does satisfy bilateral rationality.

An empirical implication of my model is that the level of within-group cooperation is decreasing in the percentage of the group within the community. I test this implication by studying informal credit relationships between villagers in rural Nepal. A credit relationship can benefit both parties to the relationship, but it is also vulnerable to cheating if one partner refuses to repay his loans, and so it is a good example of the kind of cooperative relationship that I describe. Moreover, these relationships take place primarily within castes.[3] My model predicts that the amount of informal borrowing and lending between caste members should decrease in the percentage of the caste in an administrative unit called a ward, and this prediction is borne out in the data. I also show that my model can be distinguished empirically from a number of related models, most notably the model of Dixit (2003). Dixit argues that it is harder to support cooperation in large groups because information flow is noisier in large groups, and so it is less likely that all group members will become aware of cheating by any one group member. While my model predicts that the level of cooperation should be decreasing in the percentage of the group in the community, Dixit's model predicts that the level of cooperation should be decreasing in the absolute size of the group. I show that the data support the predictions of my model against those of Dixit.

The advantage of group segregation is that it promotes cooperation, but my model also demonstrates the disadvantages of the group segregation norm. One disadvantage is that group segregation reduces the return to education. This effect appears through two separate channels. First, there is value to interacting with a wide variety of partners, and education acts as a multiplier on this value. By restricting the range of partners with whom each agent can interact, the group segregation norm also reduces the value of education. This effect may be especially important when the benefit from a relationship is taken to include the spread of ideas. Second, education makes it harder for an agent to commit to interacting only within the group, thereby making it harder for the agent to cooperate. This effect introduces an endogenous non-concavity into the education return function, which can generate a poverty trap if agents are credit constrained. A second disadvantage of the group segregation norm is that by preventing trade between groups with different endowments, group segregation can lead to the misallocation of resources.

I argue that groups and group divisions arise as adaptations to economic circumstances by rational agents who care only about material outcomes, despite the fact that there are no intrinsic differences between groups. My approach contrasts with another strand of the literature that suggests that group divisions arise from more fundamental differences between groups, such as differences in preferences. For example, Alesina, Baqir, and Easterly (1999), argue that the difference between different groups is that members of different groups have different preferences over public goods. In the final section of the paper, I discuss the relationship between my approach and this other approach, and I discuss whether and how we might expect group divisions to change.

[3]Ligon, Thomas, and Worrall (2002) develop the theory of informal credit relationships with limited commitment. Munshi and Rosenzweig (2009) and Mazzocco and Saini (2012) provide empirical evidence that these credit relationships take place primarily within castes.

# 2 Theory

## 2.1 Setup

There is a continuum of agents with mass 1. This mass is divided into $G$ groups, with group $g$ having mass $\lambda_g$, and with $\sum_{g=1}^{G} \lambda_g = 1$. These groups are payoff irrelevant, but group membership is observable. The groups are ranked hierarchically, with group 1 ranked the highest and group $G$ ranked the lowest. I defer momentarily the explanation of the significance of the group hierarchy. Time is discrete and all agents have a fixed discount factor $\delta$. In each period the following things happen:

1. Each unmatched agent chooses a subset of groups $\mathcal{G} \subseteq \{1, ..., G\}$ over which to search for new partner. I refer to this subset as the search set. Each unmatched agent $i$ pays a search cost $c(\pi_i^c)$, where $\pi_i^c$ is the proportion of agents in agent $i$'s search set within the pool of unmatched agents. We have $c(\pi_i^c) \geq 0$ for all $\pi_i^c$ and $c'(\pi_i^c) < 0$.

2. Each agent is matched with another agent according to a uniform probability function weighted by the measure of groups over which each agent chooses to search. Specifically, suppose that there are a finite number of kinds of agents, where all agents of the same kind choose the same search set. Suppose also that the set of kinds of agents is a refinement of the partition of the community into groups, so that each kind is a subset of some group. Let the kinds be denoted by $\kappa$, let the proportion of kind $\kappa$ within the pool of unmatched agents be $\alpha_\kappa$, and let $\mathcal{G}(\kappa)$ be the search set of agents of kind $\kappa$. Then given an unmatched agent of kind $\kappa$, the probability that the agent is matched with an agent of kind $\kappa'$ is given by:

$$\frac{\frac{\alpha_{\kappa'}}{\sum_{\kappa'' \in \mathcal{G}(\kappa)} \alpha_{\kappa''}} + \alpha_{\kappa'} \frac{\alpha_\kappa}{\sum_{\kappa'' \in \mathcal{G}(\kappa')} \alpha_{\kappa''}}}{1 + \sum_{\kappa'' \text{ s.t. } \kappa \subseteq \mathcal{G}(\kappa'')} \alpha_{\kappa''} \frac{\alpha_\kappa}{\sum_{\kappa''' \in \mathcal{G}(\kappa'')} \alpha_{\kappa'''}}}$$

This expression represents the following intuition. An agent can be matched either by finding a partner through her own effort, or by being found through her partner's search effort. Suppose that conditional on an agent finding her partner through her own effort, the agent has a uniform probability of matching with any agent in her search set. Then the conditional probability that an agent of kind $\kappa$ finds a partner of kind $\kappa'$ is given by the first term in the numerator of the expression above. Similarly, conditional on the agent being found through her partner's search effort, the probability that she is found by a partner of kind $\kappa'$ is given by the second term in the numerator. When normalized by the factor in the denominator, the sum of these terms then represents the total probability that an agent of kind $\kappa$ is matched with an agent of kind $\kappa'$.

3. Matched agents observe all observable information about their partners, and may then choose to accept or reject the match. If either agent rejects the match, then both agents return to step 1. Otherwise a match forms and both agents continue to step 4. Let $m_i(j) = A$ if agent $i$ would accept a match with agent $j$ given the information that agent $i$ can observe. Let $\pi_i^b$ be the proportion of agents $j$ within the community as a whole such that $j$ is in $i$'s search set and such that $m_i(j) = A$ and $m_j(i) = A$. Under the strategy profiles I consider, each agent has an equal probability of being unmatched in any period, and all agents accept all but at most a measure-zero subset of matches on the equilibrium path, so $\pi_i^c = \pi_i^b$; thus when discussing these strategy profiles I write $\pi_i^c = \pi_i^b = \pi_i$. However, the distinction between $\pi_i^c$ and $\pi_i^b$ plays a role in the interpretation of the model.

4. Every agent $i$ receives a payoff $y_i b(\pi_i^b)$, where $b(\pi_i^b) \geq 0$ for all $\pi_i^b$ and $b'(\pi_i^b) > 0$. In the main model I assume that $y_i = y_j$ for all $i$, $j$, and so for the remainder of the discussion of the main model I write $y_i = y_j = y$. In the extensions I allow $y$ to vary across agents.

5. All matched agents play a stage game, described below.

6. Each matched agent can choose to continue or break her match. If either agent in a match chooses to break the match, then both agents begin the next period unmatched. Otherwise, play continues to the final step.

7. All remaining matches break up exogenously with probability $p$.

The stage game is as follows.[4] Both partners in the relationship simultaneously choose a stage game action $a \in [0, \infty)$. An agent's payoff is $\Pi(a, a')$, where the agent chooses action $a$ and her partner chooses action $a'$. Define $v(a) = \Pi(a, a)$ and $d(a) = \Pi(0, a)$. I make the following assumptions on $\Pi$, $v$, and $d$:

**Assumption 1.** *1. For all $a > 0$ and all $a'$, $\Pi(0, a') > \Pi(a, a')$.*

*2. $v(0) = d(0) = 0$*

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

*4. $v'(0) = d'(0)$*

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

*6. $\lim_{a\to\infty} \Pi(a, 0) = -\infty$*

[4]This stage game was first described in Ghosh and Ray (1996).

Part 1 of the assumption states that 0 is the strictly dominant action in the stage game, which can be interpreted as a generalized prisoner's dilemma with a continuum of actions. If both players play $a$ then both receive a payoff $v(a)$, and I will sometimes refer to this as the value of cooperation at level $a$. If one player plays $a$ and the other plays 0, then the player who plays 0 gets $d(a)$, and I will sometimes refer to this as the value of cheating at level $a$. Parts 2 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. Lastly, part 6 states that it is technologically possible to inflict arbitrarily severe punishments.

The unique Nash equilibrium of the stage game is for both players to play $a = 0$. However, it may be possible to sustain higher levels of cooperation through intertemporal incentives, as is standard in the literature on repeated games. Thus $v(a)$ represents the value of a *long-term relationship* at level $a$. In contrast, players automatically receive the payoff $yb(\pi)$ merely from being matched, and so these parameters represent the value of a *short-term relationship*, which does not depend on intertemporal incentives. Many kinds of relationships have both a long-term component and a short-term component. For example, the short-term component of a trade relationship might be spot exchange of goods whose quality is observable. A long-term component of this relationship might be exchange of goods whose quality is not observable. In the long-term trade relationship there is greater opportunity to cheat by providing goods of low quality, and so intertemporal incentives are more important for maintaining the relationship. Similarly, the short-term component of a credit relationship might be a collateralized loan. The collateral provides the incentive to repay the loan, and so intertemporal incentives are not necessary. In contrast, the long-term component of a credit relationship might be an uncollateralized loan. In this case the incentive to repay the current loan is the prospect of receiving additional loans in the future. A relationship may also combine short-term and long-term components of different kinds. For example, two agents may exchange goods of observable quality, and they may also provide each other with uncollateralized trade credit. Much of the action in my model comes from the tradeoffs faced by different agents in trying to maximize their gain from short-term versus long-term relationships.

The search cost $c(\pi_i^c)$ for an agent to find a partner and the short-term benefit from being matched with that partner $yb(\pi_i^b)$ depend on the proportions $\pi_i^c$ and $\pi_i^b$. This can be thought of a reduced form model of the following search process.[5] Short-term relationships differ in quality, and each agent has the ability to evaluate a fixed number of potential partners to discover the quality of a short-term relationship with those potential partners. The agent finds potential partners by choosing a set of target groups, drawing potential

[5] I do not claim that it is *a priori* obvious that the true search process is one that generates this reduced form. In the empirical section, I consider the implications of some alternative search processes, and I conclude that there is evidence that the true search process is similar to the one I describe.

partners randomly from the set of unmatched agents with a fixed cost to each draw, and then throwing away any candidates who are not in the target groups. The cost of assembling a candidate pool of fixed size is therefore decreasing in the proportion of members of the target groups within the set of unmatched agents. This cost is $c(\pi^c)$. Having assembled the candidate pool, the agent then evaluates each member of the candidate pool and chooses to match with the partner who offers the highest quality short-term relationship. The expected quality of a relationship is the same with all agents, but the variance of relationship qualities differs across the population. However, agents do not know which segments of the population have high variance. The expected variance of short-term relationship qualities within the candidate pool is therefore maximized by sampling a large fraction of the population to form the candidate pool. The greater the variance of the candidate pool, the higher will be the quality of the highest quality candidate in expectation. The function $b(\pi^b)$ can be thought of as the expected quality of the highest quality candidate in the candidate pool, conditional on the candidate being willing to accept a match with the agent and the agent being willing to accept a match with the candidate. This expected quality is increasing in $\pi^b$. The multiplier $y$ indexes the relative value of short-term and long-term relationships.[6][7]

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 her 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 whom her partner has chosen to accept matches with in the past. Specifically, for each group $g$, an agent can observe whether her current partner has ever been matched with any agent in group $g$. In the informal description of the search process above, it can be thought that an agent observes the information about her partner's past matches after she has already discarded the other candidates from her pool of potential matches. After observing the information about her partner's past matches, an agent can choose to reject the match, but she must then pay the search cost again to assemble a new pool of potential partners from which to choose a new match.

[6]In effect, the model described here is an optimal stopping model from which I have removed the uncertainty regarding the amount of time and/or effort it takes for each agent to find a partner. In the full optimal stopping model, each agent would observe a stream of potential partners at some cost in time or effort for each potential partner observed. The agent would observe the group membership of each potential partner and then, depending on whether the potential partner is in her search set, she choose whether to pay some additional cost to evaluate the quality of a short-term relationship with that partner. Upon evaluating the partner and discovering her match value, the agent would then either choose to stop and accept the match or to continue searching. The fixed number of potential partners who can be evaluated by each agent in the description in the text then corresponds to an assumption in the full optimal stopping model that agents who choose search sets of the same size evaluate the same number of partners in expectation, regardless of the multiplier $y$. This assumption is helpful for tractability. However, in order for the qualitative conclusions of my model to hold I only need the weaker assumption that any agent who searches over her own and all lower ranked groups will spend less time/effort searching than any agent who searches only over her own group. This assumption holds as long as the elasticity of demand for match quality in terms of the time/effort cost of search is not too high.

[7]In the informal description of the search process in the text I ignore complications that arise from the fact that the search process is two-sided, and so agents may be matched not through their own search effort but through the search effort of their partner. Including these complications would not change the qualitative results of the model.

Let $a_i^\tau$ be the action played in the $\tau$th period of the current match by player $i$, denote agent $i$'s partner by $\mu(i)$, and let $a_{\mu(i)}^\tau$ be the action played by agent $i$'s partner $\mu(i)$. Then the history of a match that has lasted for $\tau$ periods is $h^\tau = \{(a_i^1, a_{\mu(i)}^1), ..., (a_i^\tau, a_{\mu(i)}^\tau)\}$. I consider equilibria in which agents condition their action only on the history of the current match and not on the history of any previous match.

Each agent conditions her action only on whether her partner has ever been matched with a member of a lower ranked group, and not on any other information about her partner's previous partners. This asymmetry in the treatment of agents who have been matched with members of higher ranked groups and agents who have been matched with members of lower ranked groups is the only difference between groups in the model, and is the only point where the group hierarchy is significant. I will say that an agent is "clean" if she has never been matched with any agent from a lower ranked group, and that she is "dirty" otherwise. Denote agent $i$'s cleanliness by $z_i \in \{C, D\}$. Then a (pure) strategy for agent $i$ is a tuple $s_i(h^\tau, g_i, g_{\mu(i)}, z_i, z_{\mu(i)}) = \{\mathcal{G}_i(g_i, z_i), m_i(g_i, g_{\mu(i)}, z_i, z_{\mu(i)}), a_i(h^{\tau-1}, g_i, g_{\mu(i)}, z_i, z_{\mu(i)}), n_i(h^\tau, g_i, g_{\mu(i)}, z_i, z_{\mu(i)})\}$. Here $\mathcal{G}_i$ the set of groups over which agent $i$ chooses to search, $m_i \in \{A, R\}$ is agent $i$'s decision to accept or reject the match after observing her partner's cleanliness as well as her group, $a_i$ is the stage game action, and $n_i \in \{B, K\}$ is the decision whether to break or continue the match. I only consider equilibria in which all players of the same group and cleanliness choose the choose the same actions conditional on their partners groups, cleanlinesses, and the history of the relationship, that is, I only consider equilibria in which $s_i = s_j$ for all $i$, $j$.

## 2.2 The Information Assumption

Before continuing, I briefly discuss the information assumption. Most models in the random matching/community enforcement literature follow Kandori (1992) and Okuno-Fujiwara and Postlewaite (1995) in assuming that players can observe something about what their partners have done *within* their past relationships. In contrast, I assume that players have no information about what their partners have done within their past relationships, but that players can observe something about *with whom* their partners have interacted in the past. I claim that this information assumption better reflects the true village information environment than do the assumptions of the previous literature.

Empirical evidence supports my characterization of the village information environment. Udry (1990) notes that villagers have good information about their neighbors' social and economic situations. In particular, he writes that villagers are able to report accurately about the ceremonies that their neighbors have given and attended, which supports my claim that villagers keep track of with whom their neighbors interact. In contrast, it is much harder to observe what happens within many kinds of relationships. For example, consider a credit relationship. In this kind of relationship, the stage game actions correspond to financial transactions between the relationship partners. However, various papers have shown that it is frequently very difficult for people to observe their neighbors' financial transactions. Goldstein (2000) and Anderson and Baland (2002) show that people in village environments are frequently unaware even of their own spouses' financial transactions. In other kinds of relationships the stage game action might correspond to whether participants shirk or not, which again is very difficult for third parties to observe. So my information assumption seems like a reasonable approximation to the true information environment in villages, at least for many kinds of relationships.

## 2.3 Equilibrium Concept

An equilibrium of my model is defined by three properties. First, an equilibrium must be a steady state. In a steady state, no agent can change cleanliness from one period to the next on the equilibrium path. Second, an equilibrium must satisfy an individual incentive compatibility condition analogous to subgame perfection. This is the familiar requirement that the strategy profile must be robust to the possibility of individual deviations. However, intuitively this requirement is not sufficient to prevent all plausible deviations. In an environment in which matched partners can communicate prior to choosing their actions, it is plausible to suppose that two matched agents could agree to deviate simultaneously from any given strategy profile. This intuition motivates my third requirement, that an equilibrium strategy profile must also be robust to joint deviations by any pair of matched agents. I adapt this requirement from Ghosh and Ray (1996), and following their lead, I call the requirement *bilateral rationality*. It is closely related to the various renegotiation-proofness concepts discussed in Bernheim and Ray (1989) and Farrell and Maskin (1989).

Formally, an equilibrium consists of an assignment of cleanliness to each agent and a strategy profile $s$. Let $\gamma_g$ denote the proportion of clean agents within group $g$, and let $\gamma$ denote the vector of all the $\gamma_g$'s. I assume that $\gamma$ is commonly known. Define the discounted expected utility for player $i$ by $EU_i[s_i, s_{\mu(i)}, s_{-i}, h^\tau, \gamma]$, which is the expected utility that player $i$ receives when matched with player $\mu(i)$ when all other players are playing strategy profile $s_{-i}$, given $\gamma$, after match history $h^\tau$. Note that, due to the steady state condition, the expected utility does not depend on the period or on any previous history from any past match. In addition, define $\frac{s_i'}{s_i}$ to be the strategy of playing $s_i'$ for the duration of the current match and $s_i$ thereafter. Such a collection is a bilaterally rational equilibrium if:

1. No agent changes cleanliness from one period to the next on the equilibrium path.
2. $EU_i[s_i, s_{\mu(i)}, s_{-i}, h^\tau, \gamma] \geq EU_i[s_i', s_{\mu(i)}, s_{-i}, h^\tau, \gamma]$ for all strategies $s_i'$ and for all $h^\tau$, $g_i$, $z_i$, $g_{\mu(i)}$, and $z_{\mu(i)}$.
3. There do not exist any $h^\tau$, $g_i$, $z_i$, $g_{\mu(i)}$, $z_{\mu(i)}$, $s_i'$, $s_{\mu(i)}'$ such that

$$EU_i\left[\frac{s_i'}{s_i},\frac{s_{\mu(i)}'}{s_{\mu(i)}},s_{-i},h^\tau,\gamma\right]\geq EU_i\left[s_i'',\frac{s_{\mu(i)}'}{s_{\mu(i)}},s_{-i},h^\tau,\gamma\right]\text{ for all } s_i'' \text{ and}$$
$$EU_{\mu(i)}\left[\frac{s_{\mu(i)}'}{s_{\mu(i)}},\frac{s_i'}{s_i},s_{-i},h^\tau,\gamma\right]\geq EU_{\mu(i)}\left[s_{\mu(i)}'',\frac{s_i'}{s_i},s_{-i},h^\tau,\gamma\right]\text{ for all } s_{\mu(i)}''$$

and

$$EU_i\left[\frac{s_i'}{s_i},\frac{s_{\mu(i)}'}{s_{\mu(i)}},s_{-i},h^\tau,\gamma\right]\geq EU_i\left[s_i,s_{\mu(i)},s_{-i},h^\tau,\gamma\right]\text{ and}$$
$$EU_{\mu(i)}\left[\frac{s_{\mu(i)}'}{s_{\mu(i)}},\frac{s_i'}{s_i},s_{-i},h^\tau,\gamma\right]\geq EU_{\mu(i)}\left[s_{\mu(i)},s_i,s_{-i},h^\tau,\gamma\right]$$

with at least one of the last two inequalities strict.

Condition 1 is the steady state condition. Condition 2 is the individual incentive compatibility condition. Any possible agent $i$ with any possible combination of group and cleanliness must prefer to follow the strategy profile instead of playing any alternate strategy when matched with a partner with any other possible combination of group and cleanliness, after any possible match history. This includes combinations of groups and cleanliness and match histories that cannot occur on the equilibrium path. Note that I do not require the individual incentive compatibility condition to hold for any possible $\gamma$, only for the actual steady state $\gamma$. Implicitly I am requiring the strategy profile to continue to be optimal after any countable number of deviations, which can produce matches between agents with any possible combination of groups and cleanliness. However, I do not require the strategy profile to be optimal after a positive mass of agents have deviated, which could change $\gamma$. Condition 3 is the bilateral rationality condition. It states that it should not be possible for any two agents to renegotiate to a new strategy profile that is individually incentive compatible for both of them and that is weakly pareto superior. Once again, I require the bilateral rationality condition to hold for any possible combination of groups, cleanliness, and match history, including combinations that only occur off the equilibrium path, but I only require the condition to hold for the actual $\gamma$.

## 2.4 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 cleanliness. If the benchmark strategy profile is part of an equilibrium, I will refer to that equilibrium as a benchmark equilibrium.

In the benchmark strategy profile, every agent chooses to search over every group each period. Thus, $\pi = 1$ for all agents in all periods.[8] Every agent accepts every match. In the first period of each match and in any subsequent period when neither agent has deviated from the equilibrium strategy in the previous period, each agent chooses action $\bar{a}$. If either player deviates, then the deviating player must choose a punishment action $\bar{a}^p$ in the next period while her opponent chooses $a = 0$. If either player deviates during the punishment phase then that player becomes subject to the punishment in the next period, while if both players follow the equilibrium strategy during the punishment phase then in the next period play returns to the cooperative phase where both partners play $\bar{a}$.[9] Joint deviations are ignored. At the end of each period, all agents choose to continue every match.

Since groups and cleanliness are strategically irrelevant under the benchmark strategy profile, it is unnecessary to specify the probabilities of matches between agents of different groups or cleanliness.

Let $V^m$ be the value of being an agent in a relationship that is not in the punishment phase at the beginning of a period. Let $V^u$ be the value of being an agent who is unmatched at the beginning of a period. In a bilaterally rational equilibrium, any two matched agents will renegotiate to a strategy profile in which the punishment action $\bar{a}^p$ is as large as possible, since the worst possible punishment supports the highest possible level of cooperation. In this case, $V^u$ is also the value of being an agent who is in the punishment phase of a relationship. It is not possible to inflict a punishment that reduces an agent's payoff in the punishment phase below $V^u$, since an agent who expected to receive a payoff of less than $V^u$ in the next period could choose to break off the relationship, thereby getting $V^u$ next period instead. Finally, let $V^f$ be the value that an agent believes that she will receive from any future match with any other agent. It is helpful to distinguish $V^f$ from $V^m$ since agents may be able to affect $V^m$ through renegotiation, but they cannot affect $V^f$. Bilateral rationality dictates that the two agents should choose the stage game action that maximizes their joint payoffs. That is, $V^m$ must satisfy:

$$V^m = \max_a (1 - \delta)[v(a) + yb(1)] + \delta[pV^u + (1 - p)V^m] \quad (1)$$

subject to the constraint

$$V^m \geq (1 - \delta)[d(a) + yb(1)] + \delta V^u \quad (2)$$

[8]Recall that $\pi$ denotes both the proportion of agents in the search set as a fraction of the set of unmatched agents and the proportion of agents in the search set as a fraction of the set of all agents, which are equal in the strategy profiles I consider because every agent has the same probability of being unmatched each period.

[9]Thus after any deviation, play within a match follows an optimal punishment path in the spirit of Abreu (1988). There is no mutually profitable renegotiation from the punishment path, and so the punishment path satisfies the bilateral rationality requirement (van Damme 1989).

Equation (1) says that a matched agent gets $v(a) + yb(1)$ in the current period. The match then breaks up exogenously with probability $p$, in which case the agent gets the payoff to being unmatched $V^u$ next period, and otherwise the match continues providing the player with payoff $V^m$. The constraint (2) is the individual incentive compatibility constraint. It states that the value of cooperating must be greater than the payoff that the agent receives from cheating, in which case the agent receives $d(a) + yb(1)$ this period and then receives the worst possible punishment payoff $V^u$ with probability 1 next period. The payoff to being an unmatched non-myopic agent $V^u$ is defined by:

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

Equation (3) says that an unmatched agent must pay the search cost in the current period before being matched with a new partner.

It must also be the case in equilibrium that each agent optimally chooses to search over all groups, and that it is optimal for each agent to accept every match. These conditions are trivial in the benchmark case, since in the absence of concerns about cleanliness it is always optimal to search over the entire community, as this both minimizes search costs and maximizes short-term relationship quality. All match partners are identical, so it is also optimal to accept all matches.

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

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

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

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*

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

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

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

All proofs are given in the appendix.

The interpretation of the expression for the equilibrium 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 $d(\bar{a})$ and the value of cooperating $v(\bar{a})$. The cost of cheating is that in the next period she will have to pay the search cost to find a new partner with certainty, rather than with probability $p$, so the net cost, discounted for one period, is $\delta(1-p)[c(1)]$. 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 $c(1)$ 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 $c(1)$, 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 $c(1)$ 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 incentive compatibility 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 incentive compatibility 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 $c(1)$ gets larger, higher levels of cooperation become compatible with the individual incentive compatibility constraint, and for $c(1)$ sufficiently large there exist levels of cooperation that are high enough to satisfy the bilateral rationality requirement while still satisfying the individual incentive compatibility constraint.[10]

## 2.5 The Segregated Equilibrium

In this subsection I propose what I will call the segregated strategy profile. As before, if the segregated strategy profile is part of an equilibrium, I refer to the equilibrium as a segregated equilibrium. The key difference between the segregated equilibrium and the benchmark equilibrium is that on the segregated equilibrium path agents interact only with members of their own group, while on the benchmark equilibrium

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

path agents interact with all other members of the community.

The segregated strategy profile is as follows. Clean agents choose to search only over their own group, while dirty agents choose to search over their own group and all lower ranking groups. Clean agents accept matches only with members of their own or higher ranking groups, while dirty agents accept all matches. Defining $\pi_{g,z}$ for agents with group and cleanliness $(g, z)$, we therefore have $\pi_{g,C} = \lambda_g$ and $\pi_{g,D} = \sum_{g'' \geq G} \lambda_{g''}$. Each group-cleanliness combination is associated with a level of cooperation $\bar{a}_{g,z}$. If an agent with group and cleanliness $(g, z)$ is matched with an agent with group and cleanliness $(g', z')$, then in the first period of the match both agents choose action $\min\{\bar{a}_{g,z}, \bar{a}_{g',z'}\}$. In the second and subsequent periods of each match, all agents continue to choose the level of cooperation that they chose in the first period of the match, unless there was a deviation in the preceding period. If any agent deviates, then in the next period that player must choose a punishment action while her opponent chooses $a = 0$. The punishment is sufficiently large as to reduce the punished player to her outside option utility, as in the benchmark strategy profile. After the punishment phase, players return to choosing the cooperative action. Finally, as in the benchmark strategy profile, all agents choose to continue each match at the end of each period.

I look for steady states in which all but at most a measure-zero set of agents are clean. The matching probabilities are simple as long as no more than a measure-zero set of agents deviate from the strategy profile. An unmatched clean agent from group $g$ who follows the strategy profile will match with another clean agent from group $g$ with probability 1. An unmatched dirty agent or a deviating agent who searches over a set $\mathcal{G}$ is matched with a clean agent from group $g \in \mathcal{G}$ with probability

$$\frac{\lambda_g}{\sum_{g'' \in \mathcal{G}} \lambda_{g''}}.$$

I begin by deriving expressions for the levels of cooperation $\bar{a}_{g,z}$, where $\bar{a}_{g,z}$ can be thought of as the maximum symmetric cooperation level that could be sustained by an agent of group $g$ and cleanliness $z$ if the agent chose her search set and match acceptance/rejection strategy as specified in the strategy profile, but if the agent expected to be matched each period with a partner who had an infinitely negative outside option. Define $V^m_{g,z}$ to be the value of being an agent of group and cleanliness $(g, z)$ who expects to be matched each period with a partner who has an infinitely negative outside option. Define $V^u_{g,z}$ to be the value of being an unmatched agent with the same expectations. We have:

$$V^m_{g,z} = \max_a (1 - \delta)[v(a) + yb(\pi_{g,z})] + \delta[pV^u_{g,z} + (1 - p)V^m_{g,z}] \tag{6}$$

subject to the constraint

$$V^m_{g,z} \geq (1-\delta)[d(a) + yb(\pi_{g,z})] + \delta V^u_{g,z} \tag{7}$$

where we have

$$V^u_{g,z} = -(1-\delta)c(\pi_{g,z}) + V^m_{g,z} \tag{8}$$

These conditions are analogous to the conditions for the benchmark equilibrium discussed earlier. The following lemma follows from simple manipulation of (6), (7), and (8), and so I present it without proof:

**Lemma 1.** *Define $\bar{a}_{g,z}$ to be the value of a that solves (6) subject to (7) and (8). Then $\bar{a}_{g,z}$ is the value of a that solves:*

$$d(\bar{a}_{g,z}) - v(\bar{a}_{g,z}) = \delta(1-p)c(\pi_{g,z}). \tag{9}$$

The expression for $\bar{a}_{g,z}$ can be compared to the very similar expression for the equilibrium level of competition $\bar{a}$ in the benchmark equilibrium. Notice that the maximum level of cooperation that can be sustained by an agent of group and cleanliness $(g, z)$ is decreasing in $\pi_{g,z}$, because agents with higher $\pi_{g,z}$ face a lower penalty for breaking a relationship.

I can now define a segregated equilibrium. Let $V^{g',z'}_{g,z}$ be the value of being an agent of group and cleanliness $(g, z)$ matched with an agent of group and cleanliness $(g', z')$, and let $V_{g,z}$ be the value of being an agent of group and cleanliness $(g, z)$ who is unmatched at the beginning of a period. Let $V^{g',z',f}_{g,z}$ be the value that an agent of group and cleanliness $(g, z)$ expects to receive if matched with an agent of group and cleanliness $(g', z')$ in the future. Lastly, let $q^{g',z'}_{g,z}$ be the probability that an unmatched agent with group and cleanliness $(g, z)$ is matched with an agent of group and cleanliness $(g', z')$ during the search phase if all agents follow the strategy profile. A segregated equilibrium is a tuple $\{V_{g,z}, V^{g',z'}_{g,z}, V^{g',z',f}_{g,z}, \bar{a}_{g,z}\}$ which satisfies the following conditions:

1. The levels of cooperation $\bar{a}_{g,z}$ are defined by the expression in Lemma 1.

2. For all $g$ and $z$, the values $V_{g,z}, V^{g,',z'}_{g,z}$, and $V^{g,',z',f}_{g,z}$ satisfy

$$V^{g',z'}_{g,z} = V^{g',z',f}_{g,z} = (1-\delta)[v(\min\{\bar{a}_{g,z}, \bar{a}_{g',z'}\}) + yb(\pi_{g,z})] + \delta[pV_{g,z} + (1-p)V^{g',z'}_{g,z}] \tag{10}$$

$$V^{g',z'}_{g,z} \geq (1-\delta)[d(\min\{\bar{a}_{g,z}, \bar{a}_{g',z'}\}) + yb(\pi_{g,z})] + \delta V_{g,z} \tag{11}$$

and

$$V_{g,z} = -(1-\delta)c(\pi_{g,z}) + \sum_{g',z'} q_{g,z}^{g',z'} V_{g,z}^{g',z'}. \tag{12}$$

3. Clean agents prefer to accept matches with dirty agents from the same or higher ranking groups rather than rejecting those matches and continuing to search. That is, for all $g$ and for all $g' \leq g$,

$$V_{g,C}^{g',D} \geq V_{g,C} \text{ for all } g'. \tag{13}$$

4. Clean agents prefer to reject matches with agents from lower ranking groups and to continue to search rather than accepting those matches and becoming dirty. That is, for all $g$ and for all $g' > g$ and all $z'$,

$$V_{g,C} \geq V_g^X \tag{14}$$

where $V_g^X$ is defined by

$$V_g^X = (1-\delta)[v(\bar{a}_{g,D}) + yb(\lambda_g)] + \delta[pV_{g,D} + (1-p)V_g^X] \tag{15}$$

5. Clean agents prefer to search over only their own groups rather than searching over their own and any set of lower ranking groups and incurring the possibility of becoming dirty. That is, for all $g$ and all $\mathcal{G} \subseteq \{g, g+1, ..., G\}$, we have

$$V_{g,C} \geq -(1-\delta)c\left(\sum_{g'' \in \mathcal{G}} \lambda_{g''}\right) + \frac{\sum_{g'' \in \mathcal{G} \setminus g} \lambda_{g''}}{\sum_{g'' \in \mathcal{G}} \lambda_{g''}} V_{g,D}^{g',C} + \frac{\lambda_g}{\sum_{g'' \in \mathcal{G}} \lambda_{g''}} V_g^Y. \tag{16}$$

where $V_g^Y$ is defined by

$$V_g^Y = (1-\delta)\left[v(\bar{a}_{g,C}) + yb\left(\sum_{g'' \in \mathcal{G}} \lambda_g\right)\right] + \delta[pV_{g,C} + (1-p)V_g^Y] \tag{17}$$

We have:

**Proposition 2.** *A segregated equilibrium exists if and only if*

*1.*

$$c(1) \geq \frac{1}{\delta(1-p)} \left[d(\hat{a}(p)) - v(\hat{a}(p))\right] \tag{18}$$

*2. For all $g$,*

$$(1-\delta)c(\lambda_g) \geq \frac{1-\delta}{1-\delta+\delta p}[v(\bar{a}_{g,C}) - v(\bar{a}_{1,D})] \tag{19}$$

*3. For all $g < G$,*

$$(1-\delta)c(\lambda_g) + \delta p \left[c(\lambda_g) - c\left(\sum_{g'' \geq g} \lambda_{g''}\right)\right] + \frac{\delta p}{1-\delta+\delta p} y \left[b\left(\sum_{g'' \geq g} \lambda_{g''}\right) - b(\lambda_g)\right] \tag{20}$$
$$\leq [v(\bar{a}_{g,C}) - v(\bar{a}_{g,D})]$$

*4. For all $g < G$ and for all $\mathcal{G} \subseteq \{g, g+1, ..., G\}$,*

$$y\left[b\left(\sum_{g'' \in \mathcal{G}} \lambda_{g''}\right) - b(\lambda_g)\right]\left(1 - \frac{\lambda_g}{\sum_{g'' \in \mathcal{G}} \lambda_{g''}} \frac{\delta p}{1-\delta+\delta p}\right) \tag{21}$$
$$\leq \frac{\sum_{g'' \in \mathcal{G} \setminus g} \lambda_{g''}}{\sum_{g'' \in \mathcal{G}} \lambda_{g''}}[v(\bar{a}_{g,C}) - v(\bar{a}_{g,D})] - \left(1 - \delta + \frac{\sum_{g'' \in \mathcal{G} \setminus g} \lambda_{g''}}{\sum_{g'' \in \mathcal{G}} \lambda_{g''}}\right)\left[c(\lambda_g) - c\left(\sum_{g'' \in \mathcal{G}} \lambda_g\right)\right]$$

*If a segregated equilibrium exists, then the equilibrium level of cooperation $\bar{a}_{g,C}$ achieved by clean agents from group $g$ is defined by*

$$d(\bar{a}_{g,C}) - v(\bar{a}_{g,C}) = \delta(1-p)c(\lambda_g). \tag{22}$$

The first necessary condition for the existence of segregated equilibrium is the same as the condition for the existence of a benchmark equilibrium, and it is necessary for the same reason. Under the segregated equilibrium the individual incentive compatibility constraint and the bilateral rationality constraint must be satisfied for each group-cleanliness combination, and the constraints cannot be jointly satisfied unless the search cost is sufficiently large. It is most difficult to satisfy both constraints for dirty agents from group 1, who search over the entire community (i.e. they choose $\pi = 1$) and who thus have the lowest search costs. If both constraints can be satisfied for dirty agents of group 1, then they can be satisfied for all other group-cleanliness combinations.

The second necessary condition is the condition that clean agents must prefer to accept matches with dirty agents from the same or higher ranking groups rather than rejecting those matches and continuing to search. The condition is most difficult to satisfy when the agent is matched with a dirty partner from group

1, since these partners can sustain the lowest levels of cooperation. The gain from rejecting the dirty partner is that the agent will then be matched with a clean partner from her own group with probability 1, and will be able to achieve a higher level of cooperation for a period of time that depends on the expected length of the relationship, parameterized by $p$. The loss is that she will have to pay the search cost in the current period. The condition states that the loss from rejecting a dirty partner from a group 1 is greater than the gain. If the condition holds for dirty partners from group 1, then an agent is also willing to accept matches with dirty partners from her own or all other higher ranking groups.

The third necessary condition is the condition that clean agents must prefer to reject matches with agents from lower ranking groups and to continue to search rather than accepting those matches and becoming dirty. If an agent accepts a match with a lower ranking agent, she benefits by not having to pay the search cost again, and by becoming dirty and searching over a larger group in the future she will also pay lower search costs and gain higher quality short-term relationships. However, she will be able to support lower levels of cooperation in the future. The condition states that the gain from rejecting a partner from a lower ranking group is greater than the loss. Notice that conditions 2 and 3 can only be satisfied simultaneously if $p$ is sufficiently large. This reflects the fact that an agent's incentive to remain clean is that by doing so she will be able to support higher levels of cooperation in her future relationships. This incentive is sufficiently strong only if the agent expects her current relationship to break up and a new relationship to form reasonably quickly.

The final necessary condition is the condition that clean agents must prefer to search over their own group rather than searching over their own and any set of lower ranking groups and incurring the possibility of becoming dirty. The condition states that the multiplier $y$ must be sufficiently small, that is, the value of short-term relationships must not be too large in relation to the value of long-term relationships. Otherwise, agents would be willing to forgo the benefits of higher levels of cooperation from being clean in order to achieve higher quality short-term relationships by searching over a larger set of potential partners. I defer further discussion of this condition to the extensions.

The segregated equilibrium features a reputation effect in that it is believed that agents who have interacted with members of lower ranking groups in the past will continue to interact with members of lower ranking groups in the future. In the literature, the most common way to generate reputation effects is to assume that each agent has incomplete information about the other agents' payoffs. In contrast, in my model the reputation effect appears even though all agents have the same payoffs. The reputation effect appears because the model has multiple equilibria. In particular, there is an equilibrium in which all agents are clean, and there is an equilibrium in which one agent is dirty and all other agents are clean. When an agent becomes dirty, the community shifts to the new equilibrium, in which the dirty agent is strictly worse off, because she can support lower levels of cooperation. The penalty from moving to the new equilibrium is sufficient to allow agents to commit to interacting only within their own groups, even though in the absence of concerns about cleanliness it would be optimal for each agent to search over every group.[11]

## 2.6 Welfare and Optimal Group Size

An immediate corollary of proposition 2 is:

**Corollary 1.** *For any group $g$ with $\lambda_g < 1$,*

1. *The segregated equilibrium level of cooperation $\bar{a}_{g,C}$ is higher than the benchmark equilibrium level of cooperation $\bar{a}$.*
2. *The segregated equilibrium search cost $c(\lambda_g)$ is higher than the benchmark equilibrium search cost $c(1)$ and the segregated equilibrium short-term relationship quality $b(\lambda_g)$ is lower than the benchmark equilibrium short-term relationship quality $b(1)$.*

Corollary 1 demonstrates both the advantage and some of the disadvantages of the group segregation norm. On the one hand, group segregation allows group members to support higher levels of cooperation. On the other hand, group segregation both increases the search costs paid by each agent and reduces the quality of each agent's short-term relationships. Mathematically, define the average welfare under segregation $W_g^S$ in group $g$ by taking the mean of per-period utility in a given period across all members of group $g$. Similarly, let $W^B$ be the average welfare in the benchmark equilibrium. Then we have:

$$W_g^S = v(\bar{a}_{g,C}) + yb(\lambda_g) - pc(\lambda_g) \tag{23}$$

and

$$W^B = v(\bar{a}) + yb(1) - pc(1) \tag{24}$$

The net gain from segregation is:

[11] The observant reader may have noticed a tension in this discussion of the model mechanism. I impose a renegotiation-proofness requirement on each match, and the point of renegotiation-proofness refinements is to rule out equilibria in which there are histories that contain pareto-ranked states. The intuition is that if any two matched partners were to arrive at a pareto inferior state, then they would be able to renegotiate to the pareto dominating state. However, the equilibrium I propose contains (weakly) pareto ranked states. In particular, the state in which all agents are clean weakly pareto dominates the state in which all agents except one are clean, holding constant the set of agents who are matched. What prevents the dirty agent from renegotiating back to cleanliness? The answer is that getting back to the state in which all agents are clean would require all members of the community to renegotiate simultaneously, not just any two matched partners. In effect, then, I assume that while renegotiations between matched partners are costless, renegotiations that include all members of the community are prohibitively costly. A similar implicit assumption is necessary for the equilibrium in Greif (1993).

$$W_g^S - W^B = [v(\bar{a}_{g,C}) - v(\bar{a})] - y[b(1) - b(\lambda_g)] - p[c(\lambda_g) - c(1)] \tag{25}$$

Welfare is higher under group segregation if $p$ and $y$ are sufficiently small.

Notice that the disutility from high search costs under segregation decreases as the exogenous breakup probability $p$ falls, while the disutility from low quality short-term relationships does not depend on $p$. Thus in situations where relationships are relatively stable, it seems likely that the disadvantage of segregation from low quality short-term relationships is more important than the disadvantage from high search costs.

It is also possible to derive a result on the optimal group size under segregation. In general, $W_g^S$ is not concave in $\lambda_g$, and so there is not necessarily a unique maximizer of $W_g^S(\lambda_g)$ on $[0, 1]$. However, $W_g^S$ is supermodular in $(\lambda_g, y)$. Therefore, by Topkis' theorem (Topkis 1998, theorem 2.8.3), we have

**Proposition 3.** *The set* $\arg\max_{\lambda_g \in [0,1]} W_g^S(\lambda_g)$ *has a greatest element and a least element, and both the greatest and the least element are increasing in* $y$*.*

Proposition 3 gives a more precise statement of the intuition that the optimal group size is increasing as the relative importance of short-term relationships increases.

I should note that I do not necessarily believe that the optimal group size can be used to create a positive theory of group size. Because group membership must be common knowledge, no agent can unilaterally change groups, and so group sizes are sticky, at least in the short-to-medium run. For this reason observed group sizes may diverge from the optimum. Developing a positive theory to explain observed group sizes would be an interesting project for future research.

## 2.7 From Theory to Empirics

The following is another immediate corollary to proposition 2:

**Corollary 2.** *The segregated equilibrium level of cooperation* $\bar{a}_{g,C}$ *of group* $g$ *is decreasing in* $\lambda_g$*, the percentage of members of group* $g$ *in the community.*

I test this proposition by studying informal credit relationships between villagers in rural Nepal. Informal credit relationships benefit households by allowing them to borrow in response to unforeseen shocks or opportunities. However, because formal contract enforcement is unavailable in rural Nepal, households have the option of refusing to repay their loans. Households are willing to repay their loans because by doing so they increase the likelihood that they will be able to receive additional loans in the future. The informal credit relationship therefore meets my definition of a cooperative relationship. My measure of the level of cooperation engaged in by each household is the logarithm of the sum of the outstanding amount owed by the household to friends and family members and the outstanding amount owed to the household by friends and family members. My definition of a community is an administrative unit called a ward. My main empirical hypothesis is thus the following:

**Empirical Hypothesis 1.** *The amount of informal credit that a household engages in is decreasing in the percentage of the household's caste in the ward.*

The next section is devoted to testing this hypothesis.

# 3 Empirics

## 3.1 Data

My dataset is the first round of the Nepal Living Standards Survey (NLSS), collected in a joint effort by the Nepalese government and the World Bank between 1995 and 1996. The survey consisted of a nationally representative sample of administrative units called wards, stratified by the major regions of the country. Within each sampled ward, 12 households were chosen to be interviewed. A rural ward typically contains between 50 and 500 households, and consists of a single village in those parts of the country in which houses are clustered into villages. In some parts of the country houses are distributed evenly across the countryside rather than being clustered into villages, and in those parts of the country the ward is a somewhat arbitrary administrative division. Table 1 presents summary statistics. Monetary amounts are in Nepalese rupees; there were approximately 50 rupees to the dollar in 1996. Two additional rounds of data were also collected, from 2003-2004 and from 2010-2011. These rounds were affected by the fighting and aftermath of the Nepalese civil war, which began with a few minor incidents in 1996, escalated in 2001, and ended in 2006. The civil war appears to have had a dramatic effect on the structure of the Nepalese caste system, and the results from the second and third rounds of the survey are quite different from the results from the first round. I discuss the civil war and its effects in more detail in Choy (2012).

One advantage of the Nepalese context for my study is the Nepalese government has a fairly short, standardized list of castes that it uses for all surveys. This contrasts markedly with the situation in India, where there are tens of thousands of castes, and where no official data on caste membership has been collected since 1931. The NLSS 1 includes 14 named caste categories, which are the 14 largest castes from the 1991 census making up about 80% of the Nepalese population, and a 15th category, "other". I check that my regressions are robust to dropping the "other" category.

### 3.2 Methodology and Results

My basic identification strategy will be to use a difference-in-difference regression across castes and wards to examine the effects of the percentage of a household's caste in the ward on different outcomes for the household. Thus my identification assumption is that there are no unobserved caste-ward specific factors that affect informal borrowing and lending and that are correlated with the percentage of the caste within the ward. I partially test this assumption by including various covariates in some of my regressions.

My main empirical hypothesis is that the amount of borrowing and lending between caste members is decreasing in the percentage of that caste in a ward. To test this hypothesis, I estimate the following regression:

$$C^{\star}_{icw} = \exp(\beta_1 P_{cw} + \beta_2 X_{icw} + \nu_w + \nu_{ct} + \eta_{icw}) \tag{26}$$
$$C_{icw} = \ln C^{\star}_{icw} \text{ if } \ln C^{\star}_{icw} > \xi, \text{ 0 otherwise}$$

Here $C_{icw}$ is the logarithm of the sum of the outstanding loan balance owed by the household to friends and family and the outstanding balance owed to the household by friends and family, where $i$ indexes the household, $c$ indexes the caste, and $w$ indexes the ward. $P_{cw}$ is the percentage of caste $c$ in ward $w$, $X_{icw}$ are covariates, $\nu_w$ and $\nu_c$ are ward and caste fixed effects, and $\eta_{icw}$ is an error term. The censoring point $\xi$ is unknown, as it depends on factors such as the unobserved fixed cost of borrowing. Following Carson and Sun (2007), I estimate $\xi$ by $\hat{\xi} = \min_{C_{icw}>0} C_{icw} - 10^{-6}$.

I estimate (26) using the Honoré panel tobit estimator (Honoré 1992), which unlike the regular tobit estimator can consistently estimate censored regression models with fixed effects. A disadvantage of the Honoré estimator is that it assumes that the $\eta_{icw}$ are i.i.d. conditional on the fixed effects and covariates. This assumption is questionable, given that shocks may be correlated within ward-caste groups. For this reason, I also estimate a linear probability model, taking my dependent variable to be 1 if a household owes or is owed anything to or by friends or family:

$$Pr(C_{icw} > 0) = \beta_1 P_{cw} + \beta_2 X_{icw} + \nu_w + \nu_c + \eta_{icw} \tag{27}$$

For this regression, I cluster the standard errors by ward in order to account for possible correlation in the error structure.

Table 2 presents results on the effect of caste percentage on the logarithm of the sum of borrowing and lending to and from friends and family, which I refer to as $\ln(CasteCredit)$. Column 1 shows the regression without covariates, using the Honoré estimator. The effect is negative and strongly significant, as predicted by my model.

A potential concern with my identification strategy is that caste percentage and informal credit are both correlated with some third factor that varies by ward-caste. For example, maybe it is the case that large castes are able to exert disproportionate political influence within the ward, so that they are able to acquire more wealth through political connections, and because they have more wealth they need less insurance and so they engage in less informal credit. In this case it is wealth that has the true effect on informal credit, and the correlation between caste percentage and informal credit is spurious. In order to control partially for such omitted third factors, I include a number of additional covariates in my regression in column 2. The value of each household's land holdings proxies for the household's wealth, and the variable "migrant" takes the value 1 if the head of the household was born in a district other than the district he or she is currently living in. The coefficient on caste percentage is unchanged, which indicates that there are no observable third factors that are biasing my estimate.

Columns 3 and 4 of table 2 repeat the previous exercise using the linear probability model instead of the Honoré estimator. The estimate of the effect of caste percentage on caste credit is negative in all specifications.

Table 3 presents some robustness checks. Approximately 80% of the population of Nepal is Hindu and most of the remainder are Buddhist. While the Buddhist part of the population is also divided into groups that are similar in some ways to the Hindu castes, it is possible that the Buddhist groups are not sufficiently similar to the Hindu castes to make the comparison meaningful. Thus in column 1 of Table 3 I restrict my sample to the Hindu population. In column 2 I drop the "Other" castes and restrict my sample only to the 14 named castes that also make up about 80% of the population. In column 3 I drop the top 5% of households by $CasteCredit$, to check that outliers are not driving my results. Some of the ward population data appears to be affected by missing digits and other data entry errors. For example, in one ward the sum of the number of households in each caste is 84, but in the dataset the number of households in the ward as a whole is 4. In the main dataset, I do my best to correct these errors by hand. I correct the population data for 6 wards. In column 4 check that these corrections do not unduly affect my results by dropping the corrected wards. In all of these regressions I use the Honoré estimator. In each case my results are essentially unchanged.

For comparison, table 4 presents the same regressions as in table 2, but instead of using $\ln(CasteCredit)$, I use the log of lending and borrowing to and from all other sources, namely banks, NGOs, shopkeepers, employers/employees, and landlords/tenants as my dependent variable. I refer to this variable as $\ln(OtherCredit)$. If there were some bias that was affecting my regressions in table 2, this bias would likely affect the regressions in table 4 as well, leading to a negative coefficient on caste percentage in these regressions. In fact, if anything caste percentage has a positive effect on credit to and from other sources, although this effect is not statistically significant. This suggests that members of castes that make up a larger percentage of the community substitute away from caste credit and towards other sources of credit, which is consistent with my theory.

In Table 5, I run regressions to test whether the true effect on cooperation is from caste percentage, as predicted by my model, or from absolute caste size. Column 1 repeats my basic regression estimating the effect of caste percentage on $\ln(CasteCredit)$, again using the Honoré estimator. Column 2 runs the same regression using the absolute number of households in the caste-ward instead of the caste percentage as the independent variable. Absolute caste size also has a negative effect on caste credit, which is not surprising since absolute caste size and caste percentage are correlated. Column 3 includes both absolute caste size and caste percentage in a single regression. The standard errors increase, which again is not surprising due to the correlation between caste size and caste percentage. It is more interesting to look at the change in the size of the estimated coefficients. The coefficient on caste percentage falls by a little over 10%, while the coefficient on absolute caste size falls by a factor of 6. This indicates that it is caste percentage that has the true effect on caste credit, and the positive effect of caste size in column 2 is spurious and due to the correlation between caste size and caste percentage.

# 4 Alternative Theories

My results can distinguish my model from a number of other models. My first result is that the level of cooperation is decreasing in caste percentage. This result distinguishes my model from models with perfect contracting, either with perfect contracting between any two members of the community or with perfect contracting only between members of the same caste. If there were perfect contracting between any two members of the community, then we would expect that the level of cooperation would not be affected by caste percentage. If there were perfect contracting only between caste members, then we would expect that the level of cooperation would be increasing in caste percentage, as it would be easier for members of larger castes to find relationship partners and the level of cooperation within any given relationship would be unaffected by caste size. So my results reject models with perfect competition.

A different class of models that predict that the level of cooperation should be increasing in caste percentage are models in which the social network functions as collateral. These models include Greif (1993) as well as the more recent models of Karlan, Mobius, Rosenblatt, and Szeidl (2009) and Jackson, Rodriguez-Barraquer, and Tan (2012). In all of these models, the penalty for cheating in the current relationship is that the cheater will lose her ability to cooperate not just with her current partner but also with all of her other social connections. Thus an agent's social connections act as collateral to guarantee her honesty in her current relationship. In these models agents with large social networks have more to lose from cheating and so they can support higher levels of cooperation. Thus under these models we would expect that members of castes that make up a large percentage of the population would be able to support higher levels of cooperation. My results are evidence against these models.

My second result is that the true effect on the level of cooperation is from caste percentage and not caste size. This result helps to distinguish my model from a number of other models. First, an important issue in matching models such as mine is whether there are increasing returns to scale in the matching function. A matching function has increasing returns to scale if the number of matches per person per unit time or per unit search effort expended is increasing in the size of the population.[12] In my reduced form search model, an increasing returns to scale matching function would translate into the search cost $c$ decreasing both in the proportion of the search set within the community and in the absolute size of the search set. Such a model would predict that the level of cooperation would decrease both in the caste percentage and in the caste size. The fact that the level of cooperation decreases only in caste percentage suggests that I am correct in choosing a search model that represents a constant returns to scale matching function.

The result that caste percentage and not caste size affects cooperation also helps to distinguish my model from the model of Dixit (2003). In Dixit's model, agents can sometimes observe whether their partners have cheated in past relationships, but the flow of information is imperfect and is noisier in larger groups. Thus in large groups an agent's next partner is less likely to find out if she has cheated in her current relationship, and so agents in large groups are more likely to cheat. In this model, the level of cooperation is decreasing in caste size, not caste percentage, since the noisiness of information flow within one group is not affected by the size of any other group. Thus my finding that the true relationship is between caste percentage and cooperation supports my model against Dixit's model.

In addition to the empirical results that support my model against the alternative models discussed above, there is a purely theoretical reasons to prefer my model to those other models, which is that the other models cannot rationalize the two stylized facts mentioned in the introduction. Recall that the first stylized fact is that communities remain divided into many small groups, instead of coalescing into a single group, and the second stylized fact is that there is a social norm prohibiting interactions between members of different groups. In the perfect contracting models and the models in which the social network functions as collateral, large groups are better than or at least equal to small groups from a welfare perspective. Thus we would expect that one group would tend to expand, contrary to the first stylized fact.

Unlike the perfect contracting and social network as collateral models, Dixit's model is consistent with

[12] See Petrongolo and Pissarides (2001) for a review of these issues.

the first stylized fact. However, it is not consistent with the second stylized fact. The key point to notice here is that the real world social norm prohibiting interactions between members of different groups is enforced through third party punishments. If a member of a high caste in Nepal interacts with a member of a lower caste in violation of the social norm, he will be punished by the other members of the high caste. In Dixit's model, a member of a high caste may be wary of interacting with a member of a lower caste for fear that he would be cheated, as knowledge about the cheating would be unlikely to reach the close acquaintances of the lower caste member. However, if a member of a high caste did for some reason choose to interact with a member of a lower caste, the other members of the high caste would have no reason to change their behavior towards the deviator. Thus Dixit's model cannot rationalize the third party enforcement of the social norm prohibiting interactions between members of different groups, while my model can rationalize this aspect of the social norm.

# 5 Extensions

The advantage of the group segregation norm is that it allows group members to support higher levels of cooperation, while some disadvantages include lower quality short-term relationships and higher search costs. In this section, I extend my model to demonstrate some additional disadvantages of the group segregation norm.

## 5.1 Undernvestment in education

So far I have been silent on the factors that affect the multiplier $y$, which indexes the relative value of short-term and long-term relationships. One possible factor that could increase $y$ is investment in education. This effect may be particularly relevant if the benefits of short-term relationships are taken to include the exchange of ideas, as in Lucas and Moll (2012). Education helps people to make use of ideas, and so education acts as a multiplier on the value of ideas spread through short-term relationships.[13]

If education increases $y$, then we can consider the consequences of allowing agents to invest in education in the segregated equilibrium. Suppose that all agents are clean and that all agents except one have a fixed value of $y$, where $y$ is sufficiently small that condition 4 for the existence of a segregated equilibrium in proposition 2 is satisfied. Suppose that the final agent $i$ in group $g$ is matched with another clean member of group $g$ in the current period, and suppose that at the beginning of the period, agent $i$ has the option to make an observable investment in education which would increase $y_i$. Let $V_g(y_i)$ be agent $i$'s welfare

[13] Conley and Udry (2011) show that knowledge about new technologies spreads through social networks. Foster and Rosenzweig (1996) show that education helps people to make use of new technologies.

while matched with a clean agent from her own group conditional on $y_i$. Let $y_g^*$ be the value of $y$ such that constraint 4 in proposition 2 is satisfied with equality for group $g$ and the most attractive potential deviant search set $\mathcal{G}$. If $y_i > y_g^*$, then agent $i$ can no longer commit to searching only over her own group for a new partner when her current relationship breaks up. Thus, the segregated strategy profile in which all agents including agent $i$ remain clean is not an equilibrium. However, the strategy profile in which all agents except agent $i$ follow the segregated strategy profile and agent $i$ interacts with members of her own and all lower ranking groups is still an equilibrium. Assuming that this new equilibrium is the one that obtains after agent $i$'s value of $y$ increases above $y_g^*$, then as $y_i$ increases from $y_g^*$ to $y_g^* + \epsilon$, the level of cooperation that agent $i$ can sustain falls from $\bar{a}_{g,C}$ to $\bar{a}_{g,D}$. As a result, agent $i$ faces a discrete drop in her welfare at $y_i = y_g^*$. Figure 1 shows $V_g(y_i)$ as a function of $y_i$ both in the benchmark equilibrium and in the segregated equilibrium.

The effect that investment in education can reduce welfare under segregation is a version of the "acting white" effect, discussed by Austen-Smith and Fryer (2005). The name of the effect comes from the claim that black students do not work hard in school for fear that their peers will reject them for "acting white". Austen-Smith and Fryer argue that by failing to invest in education, black students signal that they value leisure highly and thus that they make good friends. A weakness of Austen-Smith and Fryer's analysis is that it does not explain why the acting white effect should appear primarily within minority groups; their model would seem to apply equally well to white students. In my model, minorities are more likely to experience the acting white effect because members of small groups can support higher levels of cooperation, and so the value of group membership is higher for members of small groups. Thus, members of small groups are less willing to endanger their group membership by investing in education. In fact, Fryer (2010) shows that high achieving black students have fewer friends than high achieving white students, and that this effect is much more pronounced in schools that are less than 20 percent black than in schools that are majority black. This evidence supports my model.

Because of the downward jump in welfare at $y_i = y_g^*$, the function relating welfare to education is non-concave. If agents are credit constrained this non-concavity can lead to a poverty trap, as in Galor and Zeira (1993). A difference between my model and Galor and Zeira's model is that in my model the non-concavity of the educational return function is generated endogenously through group segregation, rather than being assumed to be a feature of the educational technology.

The return to education under segregation is lower than the return to education in the benchmark equilibrium even before agents reach the downward jump at $y_i = y_g^*$. The reason is that education acts as a multiplier on the quality of short-term relationships, and agents have lower quality short-term relationships under segregation than in the benchmark equilibrium. In Figure 1, this effect appears in the shallower slope of the educational return function under segregation than in the benchmark equilibrium.

Both because of the low general return to education under segregation and because of the downward jump in welfare at $y_i = y_g^*$, it is likely that the group segregation norm causes people to invest less in education than they would in the absence of the norm.[14]

## 5.2 Misallocation of resources

In the model, I suppose that all agents are identical. However, in many situations it is more realistic to suppose that different groups have different endowments on average. In this case there is an additional advantage to interactions between members of different groups, due to the gains from trade, and there is a corresponding disadvantage to the group segregation norm, which prevents trade between different groups. By cutting off potentially profitable trades between members of different groups, the group segregation norm may lead to the misallocation of resources.

Two recent papers provide evidence that this phenomenon actually occurs. Banerjee and Munshi (2004) study textile firms in Tirupur, a city in India. Ownership of textile firms in Tirupur is divided between two groups, the Gounders and the Outsiders. Roughly speaking, the Gounders are wealthier than the Outsiders, but they are less skilled at managing textile firms and their firms generate a lower return to capital. Efficiency therefore dictates that capital should be transferred from the Gounder firms to the Outsider firms. Banerjee and Munshi show that the efficient allocation is not achieved and capital remains in the Gounder firms, despite their lower returns. I argue that the mechanism in my model can partially explain this observation, since my model provides a reason why potentially profitable trades between Gounders and Outsiders might not take place.

Anderson (2011) reports similar evidence from markets for groundwater between high caste and low caste households in northeast India. Anderson shows that low caste households generate lower agricultural yields in villages in which land is owned primarily by high caste households than in villages in which land is owned primarily by low caste households. Anderson traces this discrepancy to the market for irrigation water: low caste households are unable to buy water from high caste households, and so they achieve poor outcomes in villages in which high caste households control access to water. Again, my model explains why potentially profitable trades between high and low caste households might not take place.

[14]There are many examples of minority groups that seem to underinvest in education. Some examples are discussed in Austen-Smith and Fryer, and I mention some additional cases here. Berman (2000) discusses the preference of ultra-orthodox Jews in Israel for religious education instead of secular education, despite the much higher wage premium for secular education. The Amish in the American Midwest have successfully sued in the Supreme Court to gain the right to withdraw their children from school after the eighth grade, see *Wisconsin v. Yoder*. Anthropologists studying Gypsy communities emphasize the disapproval of those communities for formal schooling (e.g. Okely 1983, Sutherland 1975).

# 6 Can group divisions change?

In this paper, I have argued that small groups are better able to promote within group cooperation, and that this is one reason why societies are divided into groups in the first place. In the final section, I discuss how my theory relates to other theories of social division, and I discuss whether and how we can expect the group segregation norm to change.

Chandra (2012) discusses the political science literature on social division, and classifies theories of social division into two main categories. First, there are *primordialist* theories, which posit that there are fundamental differences between the members of different groups, that for the most part these differences arise from impulses that are deeper than and prior to rational thought, and that these fundamental differences are the reason for observed social divisions. An implication of most primordialist theories is that since group differences reflect fundamental differences between the members of different groups, group boundaries should not change in response to changing socio-economic circumstances. In contrast, *constructivist* theories posit that groups and group divisions serve a rational function given the prevailing circumstances. Constructivist theories thus generally imply that changing socio-economic circumstances could change or eliminate group boundaries.

The leading economic theories of group divisions are primordialist. For example, Alesina, Baqir, and Easterly(1999) hypothesize that the difference between different ethnic groups is that different ethnic groups have different preferences over public goods. Preferences are fundamental features of a person that are unlikely to change significantly in response to changing circumstances. Given that different groups necessarily compete over a single pot of funding for public goods, Alesina et. al.'s theory implies that group divisions are likely to remain salient in spite of any changes in policy.

Evolutionary biologists have also proposed primordialist theories of social division. For example, in a controversial book, Wilson (2012) argues that an instinct for group solidarity is hardwired into human nature through evolution. Again, this theory implies that absent a major change in the biological conditions of life, groups and group divisions will persist.

In contrast to these theories, my theory is constructivist. I argue that group divisions arise as a way to support cooperation between group members, even though there are no fundamental differences between the members of different groups, and even though agents have no intrinsic preference for interacting with members of their own groups. My theory predicts that the right policy changes could reduce or eliminate the salience of group membership. In particular, consider again the multiplier $y$. In the previous section, I argued that $y$ could be affected by investments in education. Another factor that may increase $y$ is improvement in formal contract enforcement and the rule of law. As the rule of law becomes more entrenched, it becomes possible to support more and deeper relationships without relying on intertemporal incentives, and so the value of short-term relationships increases relative to the value of long-term relationships. For $y$ sufficiently large, condition 4 for the existence of a segregated equilibrium in proposition 2 fails to hold. Thus, as the role of formal contract enforcement expands, segregation may break down.

There is some empirical evidence that this kind of change does in fact reduce the salience of group membership. Munshi and Rosenzweig (2006) study the effects of increasing economic integration with the outside world on castes in Mumbai. Over the period they study, increasing trade opportunities increased the relative value of formal sector employment as compared to informal sector employment, which in the context of my model could be thought of as an increase in $y$. Munshi and Rosenzweig show that the percentage of people marrying outside of their castes increased dramatically as formal sector employment opportunities improved. In the context of my model, this could be interpreted as a breakdown of the segregated equilibrium.

Because group segregation helps people to support cooperation, the breakdown of segregation can be welfare reducing in the short-run. Thus, measures to improve formal contract enforcement and the rule of law may face political opposition. In the long-run, however, the end of group segregation is likely to improve growth rates and welfare by increasing investment in education and improving the allocation of resources. Moreover, my model suggests that the breakdown of segregation is likely to happen suddenly, even societies that have been segregated for hundreds or thousands of years in the past. Thus at least in this case we can have hope that the institutional limitations on growth that have been bequeathed to us by history need not constrain us in the future.

# A Appendix

## A.1 Proof of Proposition 1

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

$$V^m = \max_a \frac{1-\delta}{1-\delta+\delta p}[v(a)+yb(1)] - \frac{\delta p(1-\delta)}{1-\delta+\delta p}c(1) + \frac{\delta p}{1-\delta+\delta p}V^f \tag{28}$$

subject to

$$V^f \leq \frac{1}{\delta(1-p)}[v(a)-(1-\delta+\delta p)d(a)] + yb(1) + (1-\delta)c(1) \tag{29}$$

Since $v$ is strictly concave and $d$ is strictly convex, there exists a finite value of $a$ that maximizes $v(a) - (1-\delta+\delta p)d(a)$. Recall that $\hat{a}(p)$ was defined as the value of $a$ that solves

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

Since $v(a) - (1 - \delta + \delta p)d(a)$ has a maximum value, there exists $\hat{V}^f$ such that the constraint (29) can be satisfied for $a \geq 0$ if and only if $V^f \leq \hat{V}^f$, with $\hat{V}^f$ defined by

$$\hat{V}^f = \frac{1}{\delta(1-p)}[v(\hat{a}(p)) - (1 - \delta + \delta p)d(\hat{a}(p))] + yb(1) + (1 - \delta)c(1). \tag{30}$$

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

$$\phi(x) = \max_a \frac{1-\delta}{1-\delta+\delta p}[v(a) + yb(1)] - \frac{\delta p(1-\delta)}{1-\delta+\delta p}c(1) + \frac{\delta p}{1-\delta+\delta p}x \tag{31}$$

subject to

$$x \leq \frac{1}{\delta(1-p)}[v(a) - (1 - \delta + \delta p)d(a)] + yb(1) + (1 - \delta)c(1) \tag{32}$$

Any fixed point of $\phi$ is a benchmark equilibrium. However, notice that $\phi$ is not well-defined for all $x$, since for $x > \hat{V}^f$ there is no $a \geq 0$ that satisfies (32). I will show that $\phi$ has a fixed point if and only if $\phi$ is defined on a sufficiently large domain. I prove the following lemma:

**Lemma 2.** *1. $\phi$ is continuous and differentiable.*

*2. Let $x = -(1-\delta)c(1)$. Then $\phi(x)$ is well defined and $\phi(x) > x$.*

*3. $\frac{\partial \phi}{\partial x} < 1$ for all $x$ such that $\phi(x)$ is well-defined.*

*Proof.* Part 1: This follows from the fact that $v(a)$ and $d(a)$ are continuous and differentiable.

Part 2: Let $x = -(1-\delta)c(1)$. Assumption 1 part 4 ensures that the constraint (32) can be satisfied with $a > 0$. Examination of (31) shows that $\phi(x) > 0$, so $\phi(x) > x$.

Part 3: By the envelope theorem, we have

$$\frac{\partial \phi}{\partial x} = \frac{\delta p}{1-\delta+\delta p} - \psi < 1 \tag{33}$$

where $\psi > 0$ is the Lagrange multiplier on the constraint (32). □

Lemma 1 implies that $\phi(V^f)$ has exactly one fixed point with $V^f \geq -(1-\delta)c(1)$ if and only if $\phi(\hat{V}^f) \leq \hat{V}^f$. Since $V^f = -(1-\delta)c(1)$ is the minimum individually rational value to being in any relationship, this implies that there exists a unique benchmark equilibrium if and only if $\phi(\hat{V}^f) \leq \hat{V}^f$. Plugging in the expression for

$\hat{V}^f$ from (30) into this condition and rearranging yields the condition that a benchmark equilibrium exists if and only if

$$c \geq \frac{1}{\delta(1-p)}[d(\hat{a}(p) - v(\hat{a}(p))]. \tag{34}$$

This completes the proof.

## A.2 Proof of Proposition 2

The necessity of the first condition is proved by following exactly the same steps as in the proof of proposition 1, using the individual incentive compatibility and bilateral rationality constraints for dirty agents of group 1. Dirty agents of group 1 face the lowest search costs of any group-cleanliness combination, so if the search cost is sufficiently high that both the individual incentive compatibility and bilateral rationality constraints can be satisfied for these agents, then both constraints can be satisfied for all agents. The remaining necessary conditions for equilibrium are derived through algebraic manipulation of inequalities (13), (14), and (16).

## References


[1] Abreu, Dilip (1988), "On the Theory of Infinitely Repeated Games with Discounting", *Econometrica* 56:2, 383-396.

[2] Alesina, Alberto, Reza Baqir, and William Easterly (1999), "Public Goods and Ethnic Divisions", *Quarterly Journal of Economics* 114:4, 1243-1284.

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

[4] Anderson, Siwan and Jean-Marie Baland (2002), "The Economics of ROSCAs and Intrahousehold Resource Allocation", *Quarterly Journal of Economics* 117:3, 963-995.

[5] Austen-Smith, David and Roland G. Fryer, Jr. (2005), "An Economic Analysis of 'Acting White'", *Quarterly Journal of Economics* 120:2, 551-583.

[6] Berman, Eli (2000), "Sect, Subsidy, and Sacrifice: An Economist's View of Ultra-Orthodox Jews", *Quarterly Journal of Economics* 115:3, 905-953.

[7] Bernheim, B. Douglas, and Debraj Ray (1989), "Collective dynamic consistency in repeated games", *Games and Economic Behavior* 1:4, 295-326.

[8] Caplan, Lionel (1970), *Land and social change in East Nepal: a study of Hindu-tribal relations*, London: Routledge & K. Paul.

[9] Central Bureau of Statistics, Nepal (1996), *Nepal Living Standards Survey 1995-1996*, provided by the National Data Archive, www.cbs.gov.np.

[10] Chandra, Kanchan (2012), *Constructivist Theories of Ethnic Politics*, Oxford: Oxford University Press, forthcoming.

[11] Choy, James P. (2012), "How Do Social Institutions Change? Evidence from the Nepalese Civil War," work in progress.

[12] Conley, Timothy G. and Christopher R. Udry (2010), "Learning about a New Technology: Pineapple in Ghana", *American Economic Review* 100:1, 35-69.

[13] Dixit, Avinash K. (2003), "Trade Expansion and Contract Enforcement", *Journal of Political Economy* 111:6, 1293-1317.

[14] Eeckhout, Jan (2006), “Minorities and Endogenous Segregation”, *Review of Economic Studies* 73:1, 31-53.

[15] Farrell, Joseph and Eric Maskin (1989), “Renegotiation in repeated games”, *Games and Economic Behavior* 1:4, 327-360.

[16] Foster, Andrew D. and Mark R. Rosenzweig (1996), “Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution”, *American Economic Review* 86:4, 931-953.

[17] Fryer Jr., Roland G. (2010), “An Empirical Analysis of ‘Acting White’”, *Journal of Public Economics* 94:5-6, 380-396.

[18] Galor, Oded and Joseph Zeira (1993), “Income Distribution and Macroeconomics”, *Review of Economic Studies* 60:1, 35-52.

[19] Ghosh, Parikshit and Debraj Ray (1996), “Cooperation in Community Interaction Without Information Flows”, *Review of Economic Studies* 63:3, 491-519.

[20] Goldstein, Markus Paul (2000), “Intrahousehold allocation and farming in southern Ghana”, Unpublished Ph.D. Dissertation, University of California, Berkeley.

[21] Greif, Avner (1993), “Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders’ Coalition”, *American Economic Review* 83:3, 525-548.

[22] Grimard, Franque (1997), “Household consumption smoothing through ethnic ties: evidence from Cote d’Ivoire”, *Journal of Development Economics* 53:2, 391-422

[23] Höfer, András (2004), *The caste hierarchy and the state in Nepal: a study of the Muluki Ain of 1854*, Lalitpur: Himal Books.

[24] Honoré, Bo E. (1992), “Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects”, *Econometrica* 60:3, 533-565.

[25] Jackson, Matthew O., Tomas Rodriguez-Barraquer and Xu Tan (2012), “Social Capital and Social Quilts: Network Patterns of Favor Exchange”, *American Economic Review* 102:5, 1857-1897.

[26] Kandori, Michihiro (1992), “Social Norms and Community Enforcement”, *Review of Economic Studies* 59:1, 63-80.

[27] Karlan, Dean, Markus Mobius, Tanya Rosenblat, and Adam Szeidl, “Trust and Social Collateral”, *Quarterly Journal of Economics* 124:3, 1307-1361.

[28] Ligon, Ethan, Jonathan P. Thomas, and Tim Worrall (2002), "Informal Insurance Arrangements with Limited Commitment: Theory and Evidence from Village Economies", *Review of Economic Studies* 69:1, 209-244.

[29] Lucas, Robert E. and Benjamin Moll (2012), "Knowledge Growth and the Allocation of Time", *mimeo*, Princeton University.

[30] Mazzocco, Maurizio and Shiv Saini (2012), "Testing Efficient Risk Sharing with Heterogeneous Risk Preferences", *American Economic Review* 102:1, 428-468.

[31] Munshi, Kaivan and Mark Rosenzweig (2006), "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy", *American Economic Review* 96:4, 1225-1252.

[32] Munshi, Kaivan and Mark Rosenzweig (2009), "Why is Mobility in India so Low? Social Insurance, Inequality, and Growth", *mimeo*, Brown University.

[33] Okely, Judith (1983), *The Traveller-Gypsies*, Cambridge: Cambridge University Press.

[34] Okuno-Fujiwara, Masahiro and Andrew Postlewaite (1995), "Social Norms and Random Matching Games", *Games and Economic Behavior* 9:1, 79-109.

[35] Petrongolo, Barbara and Christopher A. Pissarides (2001), "Looking into the Black Box: A Survey of the Matching Function", *Journal of Economic Literature* 39:2, 390-431.

[36] Sutherland, Anne (1975), *Gypsies: The Hidden Americans*, New York: The Free Press.

[37] Topkis, Donald M. (1998), *Supermodularity and Complementarity*, Princeton: Princeton University Press.

[38] Udry, Christopher (1990), "Credit Markets in Northern Nigeria: Credit as Insurance in a Rural Economy", *World Bank Economic Review* 4:3, 251-269.

[39] van Damme, Eric (1989), "Renegotiation-Proof Equilibria in Repeated Prisoners' Dilemma", *Journal of Economic Theory* 47:1, 206-217.

[40] Weyrauch, Walter Otto and Maureen Anne Bell (1993), "Autonomous Lawmaking: The Case of the 'Gypsies'", *Yale Law Journal* 103:2, 323-399.

[41] *Wisconsin v. Yoder*, 406 U.S. 205 (1972)

[42] Wilson, Edward O. (2012), *The Social Conquest of Earth*, New York: Liveright Publishing Corporation.

Table 1: Summary Statistics
NLSS 1995-1996 round

| VARIABLES | mean (sd) |
|---|---|
| Caste credit | 3,880 |
| | (11,270) |
| Other credit | 4,985 |
| | (22,785) |
| Caste percentage | 0.511 |
| | (0.315) |
| Land value | 185,995 |
| | (638,633) |
| HH head age | 44.75 |
| | (14.46) |
| HH head education | 2.054 |
| | (3.669) |
| Female HH head | 0.134 |
| | (0.341) |
| HH size | 6.114 |
| | (2.987) |
| Migrant HH head | 0.142 |
| | (0.349) |
| Number of observations with caste credit$>0$ | 804 |
| Caste credit conditional on caste credit$>0$ | 11,038 |
| Number of observations with other credit$>0$ | 897 |
| Other credit conditional on other credit$>0$ | 12,694 |
| Observations | 2,279 |

Monetary amounts are in Nepalese rupees. Standard errors in parentheses.

Table 2: Effect of Caste percentage on credit to and from friends and family
NLSS 1995-1996 round

| VARIABLES | (1)<br>Honoré panel tobit<br>ln(CasteCredit) | (2)<br>Honoré panel tobit<br>ln(CasteCredit) | (3)<br>OLS<br>=1 if CasteCredit> 0 | (4)<br>OLS<br>=1 if CasteCredit> 0 |
|---|---|---|---|---|
| Caste percentage | -1.953*** | -2.012*** | -0.144** | -0.144** |
| | (0.618) | (0.592) | (0.0643) | (0.0642) |
| Land value | | 8.85e-08 | | -5.85e-09 |
| | | (2.52e-07) | | (1.57e-08) |
| HH head age | | -0.00785 | | -0.00143* |
| | | (0.00737) | | (0.000737) |
| HH head education | | 0.0517 | | 0.00124 |
| | | (0.0338) | | (0.00298) |
| Female HH head | | -0.824*** | | -0.0641** |
| | | (0.302) | | (0.0281) |
| HH size | | 0.101** | | 0.00327 |
| | | (0.0404) | | (0.00350) |
| Migrant HH head | | 0.702* | | 0.0707* |
| | | (0.407) | | (0.0361) |
| Observations | 2,279 | 2,279 | 2,279 | 2,279 |
| Number of Wards | 205 | 205 | 205 | 205 |
| Caste and Ward FE? | YES | YES | YES | YES |

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Dependent variable in columns 1 and 2 is $\ln(CasteCredit)$, where $CasteCredit$ is the sum of borrowing and lending to and from friends and family. Dependent variable in columns 3 and 4 is equal to 1 if $CasteCredit > 0$, and 0 otherwise. Columns 1 and 2 estimated using the Honoré panel tobit estimator. Columns 3 and 4 estimated using OLS. Standard errors in parentheses in columns 1 and 2. Robust standard errors clustered by ward in parentheses in columns 3 and 4.

Table 3: Robustness checks
NLSS 1995-1996 round

| VARIABLES | (1) Drop non-hindus | (2) Drop "other" castes | (3) Drop top 5 percent | Drop corrected wards |
|---|---|---|---|---|
| Caste percentage | -2.280*** | -2.077*** | -1.828*** | -2.235*** |
| | (0.587) | (0.661) | (0.626) | (0.570) |
| Land value | 1.47e-07 | 2.90e-08 | -4.58e-07 | 8.95e-08 |
| | (2.01e-07) | (2.77e-07) | (2.87e-07) | (2.57e-07) |
| HH head age | -0.00987 | -0.00938 | -0.0182** | -0.00693 |
| | (0.00808) | (0.00737) | (0.00831) | (0.00743) |
| HH head education | 0.0332 | 0.0456 | -0.0107 | 0.0528 |
| | (0.0359) | (0.0378) | (0.0363) | (0.344) |
| Female HH head | -0.847** | -0.778** | -0.688** | -0.806*** |
| | (0.332) | (0.320) | (0.333) | (0.302) |
| HH size | 0.101** | 0.0544 | 0.0819** | 0.103** |
| | (0.0417) | (0.0447) | (0.0389) | (0.408) |
| Migrant HH head | 0.639 | 0.387 | 0.744* | 0.795* |
| | (0.460) | (0.447) | (0.408) | (0.413) |
| Observations | 2,002 | 1,890 | 2,156 | 2218 |
| Number of Wards | 201 | 195 | 205 | 199 |
| Caste and Ward FE? | YES | YES | YES | YES |

*** p<0.01, ** p<0.05, * p<0.1. All regressions using Honoré panel tobit estimator. Standard errors in parentheses.

Table 4: Effect of caste percentage on credit to and from all other sources
NLSS 1995-1996 round

| VARIABLES | (1)<br>Honoré panel tobit<br>ln(OtherCredit) | (2)<br>Honoré panel tobit<br>ln(OtherCredit) | (3)<br>OLS<br>=1 if OtherCredit> 0 | (4)<br>OLS<br>=1 if OtherCredit> 0 |
|---|---|---|---|---|
| Caste percentage | 0.822 | 0.799 | 0.0750 | 0.0713 |
| | (0.597) | (0.545) | (0.0548) | (0.0551) |
| Land value | | -2.79e-07 | | -2.02e-08* |
| | | (4.42e-07) | | (1.09e-08) |
| HH head age | | -0.00437 | | -0.00111 |
| | | (0.00780) | | (0.000704) |
| HH head education | | 0.0728** | | 0.000457 |
| | | (0.0367) | | (0.00337) |
| Female HH head | | -0.945*** | | -0.0555* |
| | | (0.346) | | (0.0319) |
| HH size | | 0.217*** | | 0.0135*** |
| | | (0.0347) | | (0.00370) |
| Migrant HH head | | 0.241 | | 0.0100 |
| | | (0.418) | | (0.0435) |
| Observations | 2,279 | 2,279 | 2,279 | 2,279 |
| Number of Wards | 205 | 205 | 205 | 205 |
| Caste and Ward FE? | YES | YES | YES | YES |

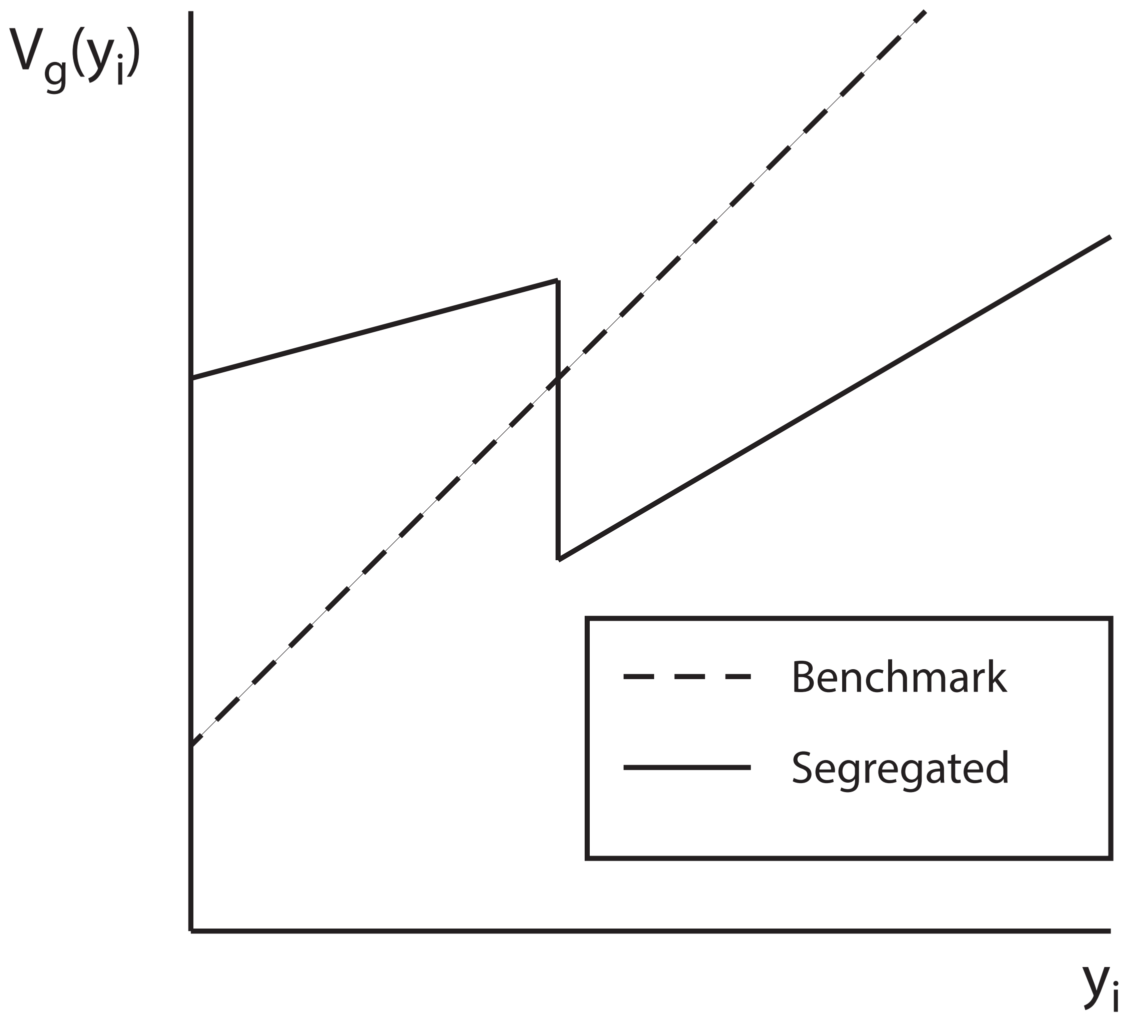
*** p<0.01, ** p<0.05, * p<0.1. Dependent variable in columns 1 and 2 is ln($OtherCredit$), where $OtherCredit$ is the sum of borrowing and lending to and from all sources other than friends and family. Dependent variable in columns 3 and 4 is equal to 1 if $OtherCredit > 0$, and 0 otherwise. Columns 1 and 2 estimated using the Honoré panel tobit estimator. Columns 3 and 4 estimated using OLS. Standard errors in parentheses in columns 1 and 2. Robust standard errors clustered by ward in parentheses in columns 3 and 4.

Table 5: Does caste percentage or caste size matter for caste credit?
NLSS 1995-1996 round

| VARIABLES | (1) ln(CasteCredit) | (2) ln(CasteCredit) | (3) ln(CasteCredit) |
|---|---|---|---|
| Caste percentage | -2.012*** | | -1.788** |
| | (0.592) | | (0.864) |
| Caste size | | -0.0105* | -0.00184 |
| | | (0.00575) | (0.00498) |
| Land value | 8.85e-08 | 1.05e-07 | 9.94e-08 |
| | (2.52e-07) | (3.07e-07) | (2.54e-07) |
| HH head age | -0.00785 | -0.00782 | -0.00789 |
| | (0.00737) | (0.00747) | (0.00737) |
| HH head education | 0.0517 | 0.0528 | 0.0517 |
| | (0.0338) | (0.0346) | (0.0339) |
| Female HH head | -0.824*** | -0.818*** | -0.827*** |
| | (0.302) | (0.302) | (0.302) |
| HH size | 0.101** | 0.0995** | 0.101** |
| | (0.0404) | (0.0405) | (0.0402) |
| Migrant HH head | 0.702* | 0.711* | 0.701* |
| | (0.407) | (0.416) | (0.408) |
| Observations | 2,279 | 2,279 | 2,279 |
| Number of Wards | 205 | 205 | 205 |
| Caste and Ward FE? | YES | YES | YES |

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. All estimations using the Honoré panel tobit estimator. Dependent variable is $\ln CasteCredit$. Standard errors in parentheses.

Figure 1




The horizontal axis is the value of the multiplier $y_i$ for agent $i$. The vertical axis is the value of being a matched agent with multiplier $y_i$ in either the benchmark or the segregated equilibrium, when all other agents have some fixed $y$. In the segregated equilibrium it is assumed that agent $i$ is clean and is matched with another clean agent from the same group.