

Girard, Victoire

Conference Paper

The impact of inter-group relationships on intra-group cooperation. A case study in rural India.

Proceedings of the German Development Economics Conference, Berlin 2011, No. 32

Provided in Cooperation with:

Research Committee on Development Economics (AEL), German Economic Association

Suggested Citation: Girard, Victoire (2011) : The impact of inter-group relationships on intra-group cooperation. A case study in rural India., Proceedings of the German Development Economics Conference, Berlin 2011, No. 32, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/48327>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The impact of inter-group relationships on intra-group cooperation. A case study in rural India.

Girard Victoire

February 2011, preliminary version

Abstract : We study the impact of inter-group relationships, with inter-group distance, on intra-group cooperation behavior for Indian rural households. This is an application to a real world case of some experimental results of the identity economics literature. This literature offers insight of channels through which inter-group relationships affect in-group actions, with identification to the in-group, and the resulting norm enforcement behavior. We proxy distance with differences of returns to attributes to one traditionally low status group (the Scheduled Castes, SC, standing for traditionally so-called “untouchables”), compared to the rest of the population (reference group). We then study the effect of this distance variable on in-group cooperation. In our data set, a cooperative behavior corresponds to the involvement in a collective action for water supply. Inter-group relationships appear to have the expected effect on intra-group cooperation for SC and households: the worst inter-group relationships, the more intra-group cooperation.

I / Introduction

How does inter-group relationships affect individual intra-group cooperative behavior? This question is of major importance since cooperation, through its importance for public goods, is necessary for economic development to take place¹. Indeed cooperation can for example mean the acceptance of a local tax, to finance a school, a well, or another shared good; or a political lobbying action, if the attribution of the public good is decided at local level (Banerjee et al. (2008)). Taking this form of a collective action, cooperation faces two issues that make her prone to suboptimal equilibrium: first, with the disconnection of private and public interest, when there are positive externalities; or, second, with free riding, if the produced good is not excludable. One way out of these suboptimal equilibrium is coercive institutions. Coercive institutions will comply everybody to participate in the collective action at a given level of commitment, ruling-out free ridding behavior. These institutions can be more or less formal, and more or less able to play their part, according to the different economic or social contexts. And the building and functioning of coercive institutions is not an obvious one, so collective action (depending upon them) remains a puzzling topic. Moreover, this puzzling topic happens to be detrimental for developing countries, as the lack of public goods has now been identified as a serious handicap in their expansion paths. And as States do not always have the power and or resources to play their part of coercive institutions it is often up to social group to handle it (see for example Miguel and Gugerty (2005) with parent

¹. “Cooperation is said to occur when two or more individuals engage in joint actions that result in mutual benefits” Bowles and Gintis (2008)

groups in Kenya), with intra- or inter group- cooperation according to what public good is at stake.

This article is an empirical work on the determinants of intra group cooperation in rural India (database presented below). The hypothesis to be tested is that inter-group distance has an impact on cooperation behaviors. Then, not only would cooperation depend on relationships between people supposed to cooperate together (a question extensively studied by a literature briefly introduced below), but also relationships between people supposed to cooperate together and people around them. We will concentrate our analysis on Scheduled Castes (SC) households, standing for the so called untouchables, a rather well identified group that is still today at the margin of Indian society (cf Thorat Newman (2010)).

Answering to that question leads us between two main stream of the literature: empirical work on cooperation and the experimental and theoretical literature about identity. Previous empirical works have first considered the impact of economic attributes, and mainly the impact of inequalities. Whether economic inequality is good, or bad, for collective action is a controversial question empirically (Bandiera et al. (2005), Khwaja (2009)) and theoretically (see Bardhan et al. (2007)). More recently social aspects have also been taken in account with a strong debate between tenants of fragmentation measure vs. tenants of polarization measure. The first consider that the more different groups in a community, the worst for cooperation (Banerjee et al. (2005), Vigor (2004)); meanwhile Montalvo and Reynal-Querol (2003), tenants of the second position, argue that the worst is to have two groups of equal size, opposing one another. Very few empirical works are made at individual level because of data limitations. Olken's (2008) study of Indonesian villagers' willingness to contribute to a program, according to the retained type of decision process, is a remarkable exception. But his problematic was the effect of direct participation in decision processes, not social relationships. Aggarwal's (2000) study on group owned wells in south India is another nice exception. His attempts at explaining intra-group variation in cooperation behavior according to the action at stake is stimulant. But he does not deal with the inter-group relationships question. Closer from our research question, and having had engaging formalization and results for us, are some theoretical or experimental work. First the formalization of identity proposed by Akerlof and Kranton (2000), insisting on how economic decisions are not just driven by economic payoffs. Also, in laboratory settings, authors have shown, with populations as different as Papua New Ginea tribes or Swiss army recruits, that norms enforcement and level of cooperation are greater in the in-group (respectively Bernhard et al. (2006), and Goette et al. (2006)). As for the fact that a negative out-group perception might enforce in-group identity it has been shown by Chen and

Chen (2009) or McLeish and Oxoby (2007).

The value added of this article is on three main grounds. First, this article uses a unique dataset, allowing us to observe cooperation behavior at the individual level and test the effect of individual predictors but also the aggregated covariates used in most previous empirical studies (that use aggregated data). Second, the hypothesis we test is in line with the current work on how much our actions are shaped by social and identity considerations², our results then offer a new example of the importance of identity questions in the economic life. And, last but not least, we put forward a new channel through which social capital affects cooperation. First regressions ran with basic social controls, as group dummies or share of group in the village, show no effect of them or small and non robust ones, which is surprising given the impact caste is still supposed to have in various ground in today India. So finding a pattern through which social setting can impact cooperation is important.

The next section gives some theoretical justifications of how distancing may impact cooperation. The dataset and variables (including the estimated distance variable) are presented in the third section, and the results are in the fourth one. The last section concludes.

II / Why should individual discrimination impact cooperation?

This article works on the effect of inter-group relationships on cooperative behavior, at the individual level. How are these two things linked? This section offers a very simple formalization answering to that question, and some examples of previous studies for how this formalization can fit facts.

2.1. Including distance effects in identity payoffs

Bringing social context as an argument of individual choice is a way opened by George A. Akerlof and Rachel E. Kranton, with their work on identity economics. Their augmented utility function, including identity, allow us to study differences in identity payoffs for SC according to their cooperative behavior in different social settings. Let us first describe the utility function. There are three arguments. The two first are traditional with a_i the vector of I's actions and a_{-i} the vector of other's actions. Then I_i , i's identity or self image is added.

So $U_i = U(a_i, a_{-i}, I_i)$

². A branch opened by Akerlof and Kranton (2000)

And I_i itself depends of five arguments. Once again a_i and a_{-i} . But also g_i , person i 's assignment (for herself and for others) to social group G . Individual i 's own given characteristics, e_i , that can match more or less its category's ideal P . P stands for ideal and prescriptions about what is the appropriate behavior to have according to social categories (different social categories can have different prescriptions).

$$\text{Hence } I_i = I(a_i, a_{-i}, g_i, e_i, P)$$

We are interested in the impact of social distance on SCs decisions so we will concentrate mainly on SCs' payoffs. Actions will depend on social categories so let us first present them. For the argument g_i consider two social groups, SCs and non SCs. Here the social group is not a choice: being SC or not (that is NCS) is an inherited identity (given in e_i). And in the hierarchy of the ancient Hindu texts SCs are at the bottom of the social ladder. Then there are two possible activities a_i for non SCs. Either to agree for close relationships with SCs with action No Distance (ND), but this will lower their identity payoffs by x because their ideal is to consider themselves and to be considered of higher social status, or to maintain a high Distance (action D) restoring their identity utility by reminding to SC their supposedly lower social status at cost z . This maintenance of social distance is a way to keep SCs apart from the rest of the society that lowers SCs' identity utility of $-s$. Consider then two possible activities for SCs, within group cooperation (C) or non cooperation (NC). Cooperation is socially valorized by group members, it would be the ideal prescription P , bringing utility H . Non cooperation would bring a lower identity utility $H-n$.

One gets the payoff matrix:

		SC	
		Cooperation	No Cooperation
NSC	Distance	$-z ; H-s$	$-z ; H-n-s$
	No Distance	$-x ; H$	$-x ; H-n$

Table 1: payoffs matrix for SC and non SC.

By construction C oftentimes yield higher payoffs to SC than NC³. So equilibrium outcome is maintained distance (D), and cooperation (C) if $z < x$. Or no distance (ND) and cooperation (C) if $z > x$. In a lot of cases distance is indeed maintained in India so $z < x$ condition is not an aberrant condition (cf Thorat Newman (2010), or Borooah (2001)). This D action does not change the equilibrium strategy for SCs since Cooperation is a dominant strategy, but it will lower SCs' identity payoffs. And if one makes the standard assumption of a concave utility function, here

³. This does not mean that SCs would oftentimes cooperate as identity payoffs are not the only ones taken in account in the utility function: the status gain is not necessarily high enough to rule out free riding.

concave over identity payoffs, identity gains will have decreasing marginal returns. Then Cooperation becomes a more interesting choice when one has its status lowered (this holds no matter if $n > s$ or $n < s$). All else being equal, if a high social distance is maintained by non SCs, the Cooperation action will have higher marginal utility returns for SCs. This result depends on three assumptions: that utility function includes identity and is concave, that maintained social distance induces a status loss for SCs, and that Cooperation induces a status gain for SCs. This extremely simple formalization shows how the free rider problem could be ruled off by identity concerns. People set apart by the high status people in the society could seek for counterbalancing that loss of status with a higher social consideration from their pairs.

2.2. Examples that low social status goes along with high identification and high identification with high cooperation.

One can also justify these assumptions and interpretations with results found by behavioral economics, showing how group settings modify group members' actions. The reasoning is in two steps. First, the effect of being distanced by others -previously formalized by a change in payoffs- would here be assimilated to an increased *identification to the group*. And second, the more one feels (or want to feel) part of a group, the more she will internalize its ideal, and adopt and enforce its norms, leading to increased in group cooperation.

About the first step of identification, the idea is that group identity becomes all the more important for an individual that she is rejected by other group members. Examples of such a behavior are found by Branscombe et al. (1999). In a study of US racial prejudices they show that *"the generally negative consequences of perceiving oneself as a victim of racial prejudice can be somewhat alleviated by identification with the minority group"*. McLeish and Oxoby (2007) found result of the same vein, using artificially built groups and make a bridge to the second step of cooperation. For them *"lower out group opinion (relative to own opinion) will reinforce in-group identity, resulting in greater cooperation within the in-group"* (and conversely for a higher out-group opinion). What they call *"lower out group opinion"* is what we want to capture with our distance variable (to be presented in next section). And being distanced by out group members going along with feeling close to in-group members, that is identifying to the in-group.

Introduced by McLeish and Oxoby example, the increased compliance to group norms, in case of higher identification, is also shown by Goette et all. (2006). Random assignment of new Swiss army recruits to different platoons allow them to show that when an agent identifies herself as a member of a social group, she will modify her behavior. She will adopt roles, norms and values of this group. And cooperation between two people of the same group will be higher than between two

recruits of different groups. Following Tajfel (1982) work numerous other studies show an in-group bias for cooperative behavior, even with groups built on minimal criteria (like a pretended preference for Klee or Kandinsky to justify random assignment).

III / Data

In order to bring these predictions to data we will use the Indian Human Development Survey, collected in 2005 by researchers from the University of Maryland and the National Council of Applied Economic Research (NCAER), New Delhi. This database covers 26,734 rural⁴ households, in 1,521 villages, comprised in 1,123 teshils, themselves included in 287 districts⁵; 32 Indian states are represented⁶. Besides having this large coverage the designers of IHDS2005 have carefully measured income, that will be a key variable for our measure of social distance. The survey design and sampling seem to be reliable one, and consumption estimates converge with the NSS consumption estimate for most states. Still there is no convergence for three states that will later be excluded from the sample as a robustness check⁷. The number of observations by village is at least 7 in 90% of the cases, the minimum is 1, the maximum 72, with a mean of 17,6 observations by village and a median of 20. Village questionnaires where hold in parallel to household questionnaire, which allows us to get valid information at village level, even if the household sampling is not representative for this scale. From this dataset we obtain the explained variable of cooperation, but also significant information in order to estimate our distance variable, and a number of controls for our model.

3.1. Explained variable: cooperation

The IHDS 2005 contains the question: “In some communities, when there is a water supply problem, people bond together to solve the problem. In other communities, people take care of their own families individually. What is your community like?” Possible answer: “Bond together” or “Each family individually”. We consider that this dyadic variable discloses the cooperative behavior of the household. Let us develop and justify this interpretation.

⁴. With the 2001 census definition of rural areas

⁵. If you keep both rural and urban data, IHDS 2005 covers 382 of the 612 Indian Districts of 2001.

⁶. There is no rural observations for Chandigar and Lakshadweep

⁷. We thank Ashwini Deshpande and Himanshu for that remark

First, we are in a collective action problem. Indeed for the first 99.42% of the sample water availability appears to be a shared good (table 1). Some households can own well(s), thus having an access to several water sources which reduces the incentive to fix problems when they occur to one of the sources, hence shall reduce the cooperative behavior (Aggarwal (2000)). We will include in regressions a control for the 9,5% of sample households owning a tube well. Unfortunately we do not know if other kinds of wells are owned, but we know if the household owns any pump (used to exploit wells). We will hence control for tube well and pump owning. We hope this well owning question does not bias our results for three reasons. First, about 90% of the sample is not concerned by these elements, and our controls shall avoid biasing the estimation with the 10% of observations concerned with one of the aforementioned elements. Second, among these small percentages, one cannot say that because a household owns well(s) it does not also rely on a public water source: a lot of private wells are built for agricultural purpose, not for drinking purpose (Aggarwal (2000)). Third, the owned wells can be common properties. Often, when a well owner dies, his heritage is shared between his sons, and if had a well this well becomes a shared good. In the two villages sampled for Aggarwal's (2005) study on group-owned wells, 61,5% of households own no well at all, 32,1% own group-well(s), and 6,3% of households own well(s) alone (he actually included in that sample households that would own a well alone and another one shared). Hence the alternate water source possessed by households owning a well shall actually correspond for a number of them to a shared good (hence the expected effect of discrimination variables shall go in the same direction as for public sources), and it does not imply that they do not any more rely on other shared water sources, according to which purpose they have to fill (irrigation, cooking, or religious acts) since different sources are used according to the purpose (Singh (2004)). This is why we have considered that potentially all households depended -at least partially- upon shared sources, and have consequently chosen to keep all variables in our sample. Besides, our results in next section give further weight to the shared well hypothesis, as tube well owning has a positive impact on cooperation (which we interpret as small group property of the source, with norms within that small group being easier to enforce than within a larger one, like a neighbourhood). So we would consider that water sources are widely shared ones. And the act of fixing the water supply problem is comparable to building a new source of water. When there is a problem nobody can use the source (no matter if it is a well, a pond, etc), and everybody has to find alternatives. If the problem is fixed, everybody (neighbour inhabitants for a public source or well owners for a shared one) is free to use the source and nobody can be excluded. The traditional free-riding problem exposed in the introduction is then illustrated by the collective action aiming at fixing the problem. The household can decide to care only of himself, waiting for others to fix the problem without having to bear the

collective action costs. Everybody has an interest for the collective action to emerge, but without including himself in.

	Frequence	Percent	Cumul
Hand pump	9,259	34.68	34.68
Piped	8,452	31.66	66.34
Open well	4,228	15.84	82.18
Tube well	3,321	12.44	94.62
Covered well	680	2.55	97.17
River	308	1.15	98.32
Pond	149	0.56	98.88
Truck	143	0.54	99.42
Rainwater or Bottled	27	0.1	99.52
Other	129	0.48	100
	26,696	100	

Table 2: household usual water source (IHDS 2005)

Second, we tend to interpret this variable as inter-group cooperation. There is no clear cut proof in any known data for this interpretation. But it is sustained by several facts and traits of Indian society. These traits lead us to think that SCs will not share the water source of any other cast group, more generally each social group (caste or religious one) is likely to have his own source(s). So then source users, supposed to fix the source, are socially homogeneous. There are three main arguments for this interpretation. First, in Hinduism, inter-caste contacts are reduced, in order to reduce the risk of “ritual pollution”. And water is a major vector of ritual pollution, so it is likely that different castes will not share the same well (Singh (2004)). Second, and as a consequence of contacts limitation, Indian villages are segregated. Each social group lives in a given neighborhood; and the extreme is for people considered as “untouchable”: historically their neighborhood were constructed next to the village, rather than as a part of it⁸ (Anderson (2007), Deliege (2004)). Third, there are often several sources in a village. In 85,7 per cent of our sample villages at least two different type of water sources are registered. This imply that there are at least two different sources in the village (one of each type), and there can be a lot more (several of each type) according to the village size. So one can assume that each neighborhood often has his own water source. Hence our interpretation of the cooperation variable as standing for intra-group cooperation: if households sharing sources are often all from the same social group it is likely that cooperation to fix the source

⁸. this is far less clear cut in cities, but we have excluded them from our sample

will come from these user groups. Numerical evidence provided by our data are scarce: in our sample 16,8% of households appear to share their sources with only people of their caste. But these numbers are very likely to be underestimated, as we only have the type of source for the household. So, if you have two hand-pumps in the village, one for each of the two neighborhoods (hence social groups), we cannot see it on our data, and these people will be considered as sharing the same source. Our interpretation of cooperation variable thus mainly depend of what we know of the Indian context. Our regression results show signs for discrimination variables that are coherent with this interpretation.

Third, the household answer is, in our opinion, interpretable as the household action. The variable measures the household perception of collective action in its community. This is a first nice information. But we also consider that the household answer is true for itself. Behavioral experiments tell us that people internalize their group norms, rules, and behavior. This has been shown by different authors and in different contexts (for example Bernhard et al. (2006)), (Goette et al. (2006)). We hence consider that in our setting, when families of the reference group of a household bond together, the household will internalize this norm and join them (hence cooperate). Conversely, if everyone stays alone to deal with the problem, so will do the household. Answering “each family individually” would sound absurd for a household which cooperates. And answering “bond together” when you stay on your own is strange. A free rider might do so, and you might have a bias because of households that would not assume their “selfishness”, which will over-declare cooperative behavior. We consider that these types of households are not a majority. A comforting observation is the variation observed for the mean answer for cooperation, if you first consider it only by village, and if then you add social group distinctions. If you consider only the people of the same social group, and using the same water source, within each villages, answers means are more polarized than if you consider the village as a whole⁹. But there are still cases when a part of people declare to cooperate when their pairs declare not to, recognizing then that they are free riders.

3.2. Variable of interest: social distance

To account for social distance we will build a variable measuring the difference of yield to similar characteristics for different social groups. The building will be in two steps with first the building with a Propensity Score Matching of a counterfactual group with characteristics similar to SCs characteristics. Then we will run income regressions over the SC group and this counterfactual

⁹. which might be interpreted as within a village some group have only cooperating members, and some group have none, hence the polarization of answers. But it might also be due to the resulting sample reduction. So the polarization of answers is hardly interpretable if alone

to use the difference of predicted income with each regression coefficients as our distance variable (following the procedure of the Oaxaca (1973) decomposition).

3.2.1 *The building of SCs' counterfactual*

We match the 4,290 SC households to the 4,290 non SC neither ST (scheduled tribe, another historically disadvantaged group) households that have characteristics similar to SCs of the 16,141 contained in the database. The propensity score is estimated with SC being assimilated to “being treated” and with the outcome being income. We perform matching over an extensive list of variables accounting for individual and local settings. For individual controls: the type and place of activity, animals owned, the land possessed if any and part of it irrigated if any, education level, an index for productive goods owned, the kind of earnings, the household size and composition and hiring, whether the household is migrant, whether it is part of the most numerous group in the village. And for local controls: number of households in the village, what kind of road the village possesses, its electrification, and if it possesses a market. As well as district and state identifiers.

We perform nearest neighbour matching with no replacement, as we are not interested in who is matched to whom, but just in who among non SCST is selected as having similar characteristics. The imposition of common support implies a loss of only 8 observations as shown in table below. Even if it is obtained with a very rough technique it is worth noting that matching results show a negative average treatment effect (being SC) on the income, at one percent confidence level.

	Off support	On support	Total
Untreated	0	11,577	11,577
Treated	8	4,290	4,298
Total	8	15,867	15,875

Table 3: Common support for the PSM

We check that the counterfactual built indeed has similar characteristics to our sample of SCs by pretending to perform another match on the already matched households sample. Compared to first matching only six covariates remain significant, three of them at five percent confidence interval and three of them at ten percent, with a R-squared falling from fourteen to 0.004 (tables presented in annex). So the matching is not perfect but hopefully we have erased most of the differences in attributes between SCs and their counterfactual so that decreasing returns or other elements do no bias next step estimates (remaining significant variables are categorical ones or

dummies). All variables that remain with significant differences were included as potential covariates in the next step.

3.2.2. *The estimation of an individual distance to the reference group*

The individual distance variable is estimated by us, using an extension of the Oaxaca-Blinder decomposition (Blinder (1973) and Oaxaca (1973)).

The Oaxaca-Blinder decomposition is traditionally used to compute average discrimination from one group to the other. It considers the income gap, and explains a part of it with the difference of characteristics between groups, and the other part of it with “discrimination”. What is called discrimination is the fact that comparable characteristics have different yield according to the pertaining group (Oaxaca (1973)). The average discrimination for the average individual is then computed. But it is likely that an individual behaviour is affected by the discrimination she really suffers from. And a comparable average discrimination, from one geographic unit to the other, can correspond to extremely variable individual situations within these units. Hence having information at individual level is nicer to explain an individual behaviour, and it is also more relevant to account for individuals environment as precisely as possible.

Also, discrimination and distance are highly correlated, one being the translation of the other, respectively in the economic or in the social sphere. We will usually call the result social distance because what we are interested in here is discrimination impact on identity, that is related to the social sphere side¹⁰.

In order to estimate that individual discrimination, for members of different groups, we follow Rio et al. (2010) method. We estimate income equations for each group. We then compute regression coefficients and predict everybody's revenue with it. Each group coefficients indicate the yield of its characteristics. For household i of group g we compute two predicted incomes: the first one, \hat{y}_i^g , with his group coefficients; and the second one, $\hat{y}_i^{\bar{g}}$, with “reference” group coefficients. The difference between the two is the discrimination/distance variable of household i , belonging to group g :

$$\text{Discr}_i^g = \hat{y}_i^g - \hat{y}_i^{\bar{g}}$$

From there we will compute relative discrimination. It can be computed with respect to household predicted income with group coefficient ($\text{RDiscr}_i^g = D_i^g / \hat{y}_i^g$) or with reference group coefficients ($\text{RDiscr}_i^g = D_i^g / \hat{y}_i^{\bar{g}}$). These different specifications do not change the sign of our results even if they change their magnitude, and we will concentrate on discrimination relative to

¹⁰. We would have focused more on the discrimination with difference of returns aspect if people discriminating and discriminated had to cooperate together, because differences of return would then have had an impact. But here it is not the case.

household own income, as it is supposed more relevant for the household perception of her situation.

3.2.3. *Income equations and predictions*

Inspired by Borooah (2005) work on inequality and poverty in India, we estimate a log-linear model: the log of the revenue¹¹ is explained by a set of different variables. Explaining variables are inspired by Borooah (2005) with augmentations. The first set of covariates, W , includes land owned, the number of adult workers, the number of teenage workers, an index for the amount of (non-land) productive assets owned¹² and the maximum level of education for adults in the household. All these are interacted with regional dummies. The second set, X , is made of dummies for the different type of professional activities being the main income source, and for if the household belongs to the most numerous group in the village. And the third set, L , is made of three local controls: the number of households in the village, the accessibility of the village by road, and distance to the closest market. And we add District fixed effects. Previous studies at the aggregate level show significant, historically rooted, inter-district variations of inter-caste relationships (see Banerjee and Somanathan (2001) or Banerjee et al. (2005)). The econometric specification is thus¹³:

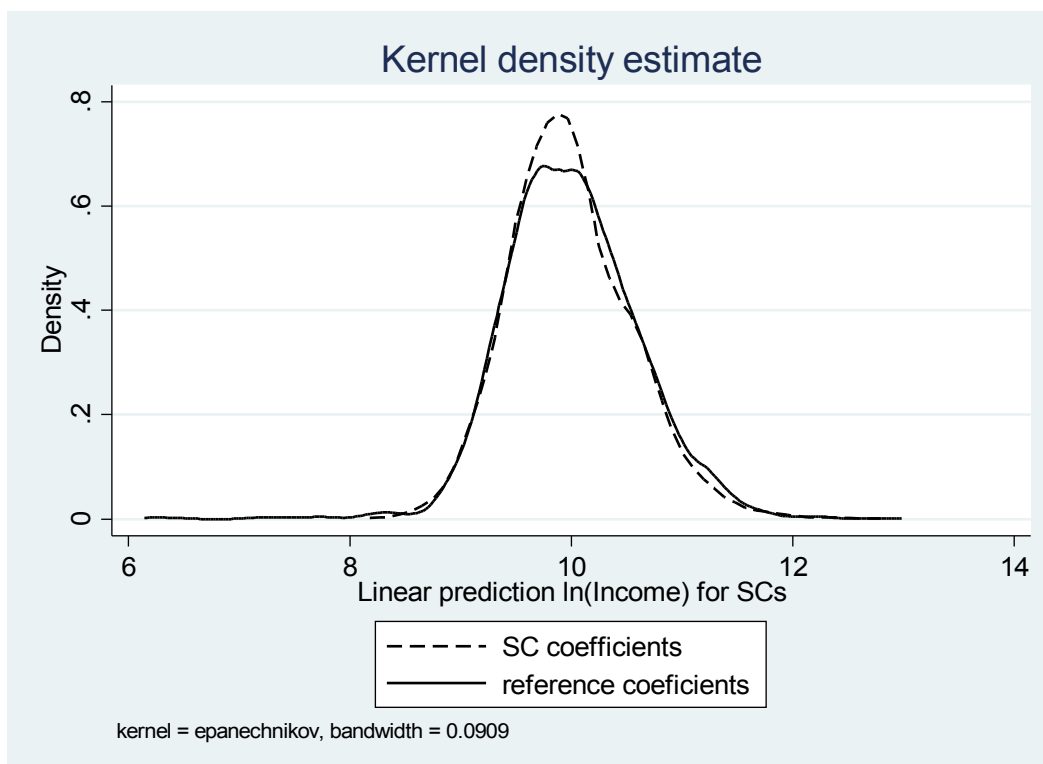
$$y_i^g = \ln(\text{income}_i^g) = \alpha_0^g + \alpha_1^g W_i + \sum_{\text{Reg}} \alpha_{1\text{Reg}}^g \text{Region}_i * W_i + \alpha_2^g X_i + \alpha_3^g L_i + \varepsilon_i$$

Regressions are run for the two different social groups. Results of regressions over income and fit with actual income are presented in annex. Both regressions have R-squared above 0.40 and graph are engaging for the quality of predictions: except for some households at low income levels predictions are close from declared income. From there we also predict SCs income with reference coefficients; and compute the difference between that prediction and the one made with their actual coefficients. The resulting distributions of income are presented below. One can observe confounded patterns for low attributes rewards. Then appears some “glass ceiling” with SCs' coefficients predicting SCs to be more numerous at medium level income, and less at higher level income, than reference coefficients.

¹¹. We consider the declared income of the household minus the remittances declared, as these do not depend on the surveyed household characteristics. Moreover remittances often come from persons exiled in a city or a foreign country, where caste or religious distinction shall be far less relevant than in the village of origin.

¹². Index equals to a weighted sum of the following goods (weights in parentheses): draft animals (4); pumps (5); tubewell (10); bullock carts (4); tractor (10); threshers or bio-gas plants (3).

¹³. The final specification was chosen among others using Akaike Information Criteria. Except for irrigation that is a very poorly filled variable: including it leads to the loss of 1000 SCs observations for prediction. So we do not include it as a covariate as matching has smoothed it among groups, and its suppression does not bias our results (hausman test). Still its inclusion is presented in robustness checks.



Graph 1: distribution of predicted incomes for SC households

Discrimination variable further used to proxy distance is presented in table 2. SCs appear to be on average disadvantaged by their coefficients. This is consistent with average treatment effect obtained from the matching, and also with replication for our data of Borooah (2005) results: with the full sample we find that 22,2% of income differences between SCs and non SCSTs is attributable to discrimination¹⁴. These results are also consistent with Throat, Mahamanlik and Sadana survey results (in Throat Newman (2010)). They show that SCs are either excluded from market or suffer from systematic deviation from market price at their disadvantage, being sold/rented out products at high price, and bought/rented in products at low price. The estimated variable is negative for about 52% of the sample. We call negative discrimination the absolute value of the variable for SCs when it is negative, that is to say when their predicted income is higher with non SCSTs coefficients than with SCs coefficients. For the remaining 48% of the SC sample, discrimination would be “positive”. Descriptive statistics show a credible range for discrimination values. The low rate of negative discrimination may sound surprising but at least two things can contribute for it. First, the existence of various positive discrimination programs that have targeted

¹⁴. This is far lower than Borooah finding of 36% of income difference attributable to discrimination, but India is evolving and in 10 years discrimination against SCs seems to have been significantly reduced

SCs and STs households since independence. Second, and more importantly, our variable is very likely to be an underestimation of the real difference of returns between SCs and the reference group. Indeed in the building of the counterfactual group we privileged similarities of characteristics. This leads us to keep in the sample Muslims, Christians, or other religions households, as well as households categorized as Other Backward Castes although all of them will occupy somewhat despised ranks in the social ladder and often suffer from economic discrimination (Thorat and Newman (2010)). So the counterfactual is made of very different kind of households. Data limitations did unfortunately not allowed us to concentrate only on Hindu neither OBS nor SC households which would have led a far more clear cut picture.

But, although certainly undermined and not precise enough, we think that at least the negative outcomes of that discrimination variable does give a relevant ranking information, so we will use them to proxy social distance¹⁵.

		Obs.	Mean	Std. Dev.	Min.	Max.
Absolute ¹⁶	Positive	2601	5543.41	5421.6	2.279	32131.4
	Negative	2812	6634.6	6515.4	3.168	35544.5
Relative	Positive	2606	0.208	0.149	0.0001	0.709
	Negative	2794	0.321	0.278	0.0001	1.463

Table 4: Descriptive statistics for predicted individual discrimination for SC households.

3.3. Control variables

Economic situation (household level). Variables at stake: income, land owned, difference from land owned to average land owned in the village (obtained from the village survey) and its square, difference from land owned to average land owned by the households group members and its square. These variables aim at accounting for the Olson effect at individual level. The idea is that people with a high income, or with a lot of land, are the ones who shall benefit more from the collective action, hence they might cooperate more. On the contrary Bardhan et al. (2007) show

¹⁵. We would still keep the positive outcomes as a control variable among others, that one could choose to interpret as a social distance variable, but the interpretation would be less straightforward than for negative outcomes. Shall one consider it as a lowering of SCs status, because it still marks social distance, or as an increasing of their status, because it is as if they were “rewarded” for being SCs? Knowing that it is likely to be overestimated. Plus, the correlation it shows with the outcome variable was less robust to alternate specifications than the one for negative outcomes. And, the sign of the effect makes it prone to interpretation as an Olson effect of the more well offs cooperating more, but also raises concerns of reverse causality. All these reasons make that we will not focus on these outcomes.

¹⁶ To be compared to annual incomes going from zero to Rs. 564000, with a mean of Rs. 31044.17 and a standard deviation of Rs. 34420.38.

theoretically that efficiency within a group rises when within group equality rises. The inequality effect can be relevant at village level or at social group level and its effect is not necessarily linear.

Farming activity (household level). Variables at stake: a dummy for firstly farmer households, land owned, tubewell owned. A household mainly depending on cultivation for living might be more interested in water source functioning than the one who use it only in order to drink or cook (it is easier in this case to use another water source, and have less long term consequences). On the other hand, farmers are also more likely to own their own tube well (this is the case for 8,7% of our sample), and then depend less on the public source.

Social group (household level). Variables at stake: group dummies for household caste or religion, and for if their group is the most numerous in the village, interaction term between group dummies and share of the corresponding social group in the village (provided by village survey), and square of this interaction term. A household's social group is its reference of people to cooperate with. Within a group people share the same norms, and members of the group make sure these norms are respected (Bernhard et al. (2006), Goette et al. (2006)). Group dummies allow accounting for different functioning from one group to the other. Some can be more prone to cooperate than others. As for the share of household's social group in the village, it may also change cooperative behavior: the more numerous its pair's proportion, the higher the probability that it has to cooperate only with them, with no out-side group intruders (and conversely for a small share). The shares might be interpreted as the household utility surplus coming from cooperation with its pairs (Vigdor (2002)), with a positive impact. But, on the other hand, larger groups can have more coordination problems (Bardhan (2000)). Also, belonging to the most numerous group raises your chance to have a second water source accessible and used by your group members, which shall decrease the likelihood to cooperate. So both signs are possible for these variables.

Other capital (household level). Variables at stake: number of adult members, maximum level of education of an adult male member, fact to belong to village official or having a close relative in (these people might tend to be overoptimistic about cooperative behavior within the circumscription).

Economic differences (village level). Variables at stake: gini and gini squared on income, or on land, inter-group component of the Theil index on land (built with Village survey) and its square. We can

thus control for the effect of village level economic heterogeneity. For Bardhan et al. (2007), there exist an optimal level of between group inequality and we account for this one.

Social fragmentation (village level). Variables at stake: fragmentation and polarization indexes, computed for all social groups distinguished in the village survey, and for major caste or religious groups. On the one hand we use village survey questions concerning the different “jatis” (traditional social groups) present in the village. There can be up to eight different ones within a village. We obtain general fictionalization and polarization indexes. These indexes are the most synthetic insofar as they account for caste and religious distinctions at the same time. But a higher aggregation level might be more relevant at a time when caste distinctions are said to be fading. So we also extract from the survey data the shares of people of each religion and of each broad caste categories (Brahmans, OBC, SC, ST) to compute other indexes.

Political context (village level). Variables at stake: a dummy for the presence of the local assembly building within the village (Pani Panchayat), and another dummy if the village benefices of a public program promoting safe drinking water in the village. These political facts might enhance the coercive institutions promoting cooperation.

Water sources (individual and village level). Dummies for household main water source, dummies for if he owns a tubewell or a pump, dummies for village main water source (provided by village survey), and the interaction between the two when corresponding.

All these control variables are used in our regressions, in order to account for households background, with a household individual characteristics, and its living place characteristics. Descriptive statistics of main variables are presented in annex. We choose the final specification using Akaike information criteria.

The final model can be written:

$$\text{Cooperation}_i = \beta_0 + \beta_1 \text{Distance}_i + \beta_2 \text{IndivControls}_i + \beta_3 \text{VillageControls}_i + \varepsilon_i$$

IV / Results

We first run two parsimonious probit regressions, to check for the non linearity of the effect of discrimination variable. The effect indeed appears as non linear, with both sign of the

discrimination variable significant at very high confidence levels¹⁷. The negative discrimination, further called distance variable, has the expected effect on Cooperation: the more discriminated against (that is the higher the social distance) the more SCs cooperate. Our interpretation derives from formalization and previous finding exposed in section 2. Various interpretations are possible for the sign of the positive part of discrimination variable but as exposed above this result is not reliable (and will further appear not robust) so we will not comment upon it although the variable will be kept as control in further regressions.

VARIABLES	(1) Coop	(2) Coop
Discrimination toward SCs	-0.0473 (0.0419)	
Negative Discr toward SCs		0.192*** (0.0529)
Positive Discr toward SCs		0.166*** (0.0537)
Observations	24,213	24,213
log likelihood	-8.440e+07	-8.420e+07
Pseudo R2	0.000206	0.00256
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1		

Table 5: Cooperative behavior prediction, parcimonious specifications (marginal effects)

A first endogeneity concern arising would be reverse causality. But this does not seem to be a relevant concern for us. The reverse causality relation relies on the intuition that free riders would be the more discriminated against. But, in our setting, discriminators aren't the people you are supposed to cooperate with. And, as a consequence, empirical results show the exact opposite: the more you are discriminated/distanced, the more you will cooperate. This rules out the reverse causality concern.

We then introduce the controls, presented in the previous section, with clustered standard errors at village level (to account for Moulton's bias). Each time we have to choose between two alternative controls, we run regressions with both, holding all other elements fix, and keep the specification with the highest log likelihood. The exception to this method is for some variables that are poorly reported and the use of alternate controls implies big variations for the number of observations. This is mainly the case for village variables on land owned, that are used to compute ¹⁷. All presented regressions are without outliers though their inclusion would give exactly the same conclusions for sign and significance level of variables.

economic inequality controls, and village variables on social group shares, used for group shares and fragmentation indexes. We thus present two alternative specifications. The first includes the maximum number of observations (column 3) and only properly filled control variables. The second((column 4) includes extensive controls, no matter how well they are filled. One can see that in both cases the higher social distancing toward SCs, the more cooperation. Results are presented in table 6.

VARIABLES	(3) Coop	(4) Coop	(5) Coop	(6) Coop	(7) Coop	(8) Coop
Distanciation as SCs	0.253*** (0.0656)	0.235*** (0.0711)	0.225** (0.110)	0.199* (0.121)	0.192 (0.121)	0.292 (0.189)
Minimum Indiv Controls	yes	yes	yes	yes	yes	yes
Minimum Controls Village	yes	yes	-	-	-	-
Extensive Controls	-	yes	-	yes	-	-
Village Random Effects	-	-	yes (1,513 villages)	yes (1,290 villages)	yes (1,290 villages)	-
Village Fixed Effects	-	-	-	-	-	Yes (1,077 villages)
Observations	23,647	18,623	24,092	19,684	19,684	18,591
log likelihood	-8.000e+07	-6.110e+07	-12208	-9930	-9943	-8263
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1						

Table 6: Cooperative behavior prediction

We then introduce village random effects, both for the well filled (5) and poorly filled model (6). The significance loss in the model with maximum variables filled is actually due to observations loss, as can be seen in column 7 when we run the same model as column 5 but on observations of column 6. A similar argument can be advanced for the loss of significance when we introduce fixed effects in a logit (a p value of 12%, column 8): the z stat of column 3 model re-estimated on the restricted sample used for fixed effects decreases from 3.86 to 2.48. But a Hausman test further confirms that fixed effects are the efficient specification. This is a serious concern for the robustness of our result, though the fact that sample restriction leads to the loss of about one fifth of the sample, with among it observations leading the result, leaves us with no clear cut conclusion.

We then try to run again all the estimation procedure (from the beginning, with re-estimation of the social distance variable) without the 2,382 observations belonging to Bihar, Jharkand, and Tamil Nadu, three states for which consumption estimates do not converge with NSS data, raising concern on the sampling quality for these states. Predicted incomes do not differ from previous ones for SCs, neither with their coefficients, nor with reference coefficients (although state exclusion leads to a higher predicted income for the reference group). As a consequence the re-estimated distance variable does not significantly differ from the previous one. It remains significant at the highest level for the parsimonious specification. But it is only partially robust to the introduction of controls (with a p-value under 9% for the extended control specification). This significance drop is linked to the sample reduction, leading to key observations loss. We can unfortunately not address further that problem with existing data. So, given that our interest variable does not seem to be biased by the inclusion of these three states, we will go on working with it. Although aware that the strength of results is undermined by that potential sample bias.

VARIABLES	(9) Coop	(10) Coop	(11) Coop
	Alternate measure	Alternate measure	Initial measure
Distanciation as SCs	0.571*** (0.215)	0.444** (0.279)	0.386* (0.208)
Minimum Indiv Controls	yes	yes	yes
Minimum Controls Village	-	-	-
Extensive Controls	-	-	-
Village Random Effects	Yes (1,511 villages)	-	-
Village Fixed Effects	-	Yes (1069 villages)	Yes (1069 villages)
Observations	23,075	17,726	17,726
log likelihood	-11766	-7868	-7873

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 7: Robustness check, cooperative behavior prediction with alternate distance computation

One last robustness check is with modification of the income equations used to compute distance variable. New income equations are run, including a control for land irrigation¹⁸. Regressions results of village random and fixed effect specifications with minimum controls are shown in table 7 (columns 9 and 10), other results remain unchanged. The impact of distance measured with that alternate way is higher. This impact augmentation is partly due to sample reduction (as shown in column 11 if compared to column 8 of table 6). But not only, otherwise significance levels would be exactly the same in columns 10 and 11. Hence the alternate measure of

¹⁸ Being poorly filled using it for income prediction leads to the loss of one fifth of SC households sample.

distance -that includes irrigation and is a more accurate fit- has a better explicative power of cooperation than the previous one. This is a comforting result although it leaves us, once again, with no clear cut conclusion.

In the end what have we learned? Regressions ran do bring interesting insights and empirical support to the thesis exposed in the second section. The social distance impact is not robust to exclusion of three states with potential sample bias, this is a serious concern that we can unfortunately not address with current data. With two alternative measures of the distance variable the impact of social distance is perfectly robust to random effects inclusion, and partially robust to village fixed effects inclusion¹⁹. Whenever social distance has an impact it is the one we expected: being distanced by other households leads the set apart households to try to compensate for it. This compensation will be obtained through increased cooperation, because this goes along with a higher social status within the in-group.

As for a broad review of controls effects (full tables for some estimations are in annex), except through the amount of land owned (which effect can also be interpreted with corporatism or other reasons), neither richness nor inequality seem to influence cooperative behavior. This reinforces our interest on social variables. As a macro variable, fragmentation of the village is not relevant to explain cooperation behavior. Polarization has instead a small positive impact, that can be linked to our second section explanation, with the conflict raising to two big group opposition to each other reinforcing each in-group identity. We can also deconstruct the micro foundation of the fragmentation index, following Vigdor(2002), and consider group shares. There appear to be some convex link between the share of Muslims in a village and their behavior of cooperation. This can be interpreted as the presence of some threshold of fellow group members below which the household does not cooperate because he has nobody from his in-group to cooperate with. Then, once the threshold is passed, the more in-group fellow members are in the village the more cooperative will be the household. As for SCs their intricate characteristics tend to be of lower cooperation than the reference social group (non SCST here), in all context or mainly when they are the dominant group. That last result is seemingly the most reliable as it appears when we include village random effects and it fits in the formalization exposed in the second section: when SCs are the most numerous they are less likely to suffer from out-group despise and cooperate less. One can also notice that the three more “political” variables included have an impact. Public program for water has a negative impact, hopefully because then the State substitutes itself to user-groups. Being

¹⁹ With mitigate conclusions due to the fact that one fifth of the sample lives in villages without intra village variation of our dyadic cooperation variable.

a Panchayat member lowers Cooperation, which is a surprising result, maybe linked to officials higher expectations before they report cooperation of their group. Eventually knowing somebody who belongs to the local assembly and is close from the household has a strong positive impact, which can have various explanation, from a real greater involvement in collective actions to over-reporting. What is clear from that is that political and social context do matter. Except for land owned not a single economic variable is significant in all our regressions meanwhile social and political variables are.

V / Elements of Conclusion

Intra-group cooperation appears to depend on inter-group relationships. We have proxied intrer-group relationships with a distance variable. We have estimated how SCs individuals were set aside from Indian society today. This is blunt estimation, and it contains imperfections among which a very likely underestimation of actual social distance. The distance variable obtained is then introduced as an explicative variable of households' cooperative behavior. It appears to have a significant explicative power for SC households' behavior. This explicative power is robust to outliers removal, variation in specifications, regression model used, and village random effects. It is partially robust to village fixed effects and not robust to sample modification (with removal of three states).

To put it in a nutshell, the more, as a SC household, you are set apart from the rest of the society, the more you will cooperate with your in-group members. Our interpretation of this result is through identity economics literature: a distanced household shall tend to identify more to his group of origin, and a reinforced identification will implie reinforced norms, raising in-group cooperation. The household will seek for his pairs recognition, through an action they promote, to compensate for the identity loss of being despised by out-group members. Thus this article is a bridge between identity literature results, mostly theoretical or derived from experimental settings, and a growing empirical literature looking at which community characteristics are the best for cooperation and public goods existence and maintenance. The identity economics theoretical and experimental insights are confirmed by our results on survey data: identification to the in-group matters for cooperation. And this identification depends, among other things, of how out-group members treat you.

The aim of this article is not to enhance between group distance positive aspects, but to point at a mechanism reinforcing cooperation. This mechanism is in line with the work done on identity economics, showing that identity payoffs can be important determinants of one economic decision

-here cooperation (Akerlof and Kranton (2000)). Results presented in this article show that enhancing group identities reinforce cooperation. But groups can be defined along a lot of social lines and the aim is not to support caste communitarianism or anything alike. Singh (2011) article shows that in today's Kerala a state level identity has been built, that means more for its inhabitant than their caste or religious distinctions, or their Indian identity. If the growing literature in identity goes on showing its importance for major economic decisions, the way identities can be reinforced or defined along new lines, stimulating one or another aspect of identity (as done by McLeish and Oxoby (2011)), would become a major question for policy makers.

This article is a first step in showing that in-group cooperation behavior is, among other things, explained by out-group members comportment: bad inter-group relationships are detrimental for inter-group cooperation, but enhance intra-group cooperation. For future research this element shall be taken in account in aggregated work, studying at local level how many public goods are presents: when public goods can be partially privatized (each group being able to use or have its own one reserved for itself), the worst inter-group relationships, the more public goods. Wells, but also education, or health facilities, are examples of goods that can be privatized. In such cases fragmentation indexes, or polarization indexes, will not give enough information to understand communities' collective actions. Further research is needed to find the best proxy of inter-group relationships, both at individual and at aggregated level.

Bibliography

- Aggarwal, R. M. (2000) « Possibilities and Limitations to Cooperation in Small Groups : The Case of Group-Owned Wells in Southern India », *World Development* Vol. 28, No. 8, 2000
- Akerlof, G. A. and Kranton, E. R. (2000) « Economics and Identity » *Quarterly Journal of Economics* 115
- Akerlof, G. A. and Kranton, E.R. (2005) « Identity and the Economics of Organizations », *The Journal of Economic Perspectives*, 19(1)
- Akerlof, G. A. and Kranton, E. R. (2010) «Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being » Princeton University Press.
- Alesina, A. and La Ferrara, E. (2005). « Ethnic diversity and economic performance », *Journal of Economic Literature*, vol. 43(3), pages 762–800.
- Bandiera, O. Barankay, I. and Rasul, I. (2005). « Cooperation in collective action », Natural field experiments, The Field Experiments Website.
- Banerjee, A. Iyer, L. and Somanathan, R. (2005). « History, social divisions, and public goods in rural India », *Journal of the European Economic Association*, vol. 3(2-3), pages 639–647.
- Banerjee, A. Iyer, L. and Somanathan, R. (2008). « Public Action for Public Goods », *HandBook of Development economics*, Elsevier, ch. 49, pages 3117-3154.

- Banerjee, A. and Somanathan, R. (2001). « Caste, community and collective action: The political economy of public good provision in India ». mimeo, Massachusetts Institute of Technology.
- Banerjee, A. and Somanathan, R. (2007). « The political economy of public goods: Some evidence from India », [Journal of Development Economics](#), Elsevier, vol. 82(2), pages 287-314.
- Bardhan, P. (2000). « Irrigation and cooperation: An empirical analysis of 48 irrigation communities in south india », *Economic Development and Cultural Change* vol. 48(4), pages 847–65.
- Bardhan, P. Ghatak, M. and Karaivanov, A. (2007). « Wealth inequality and collective action », *Journal of Public Economics* vol. 91(9), pages 1843–1874.
- Bernhard H., Fehr E. and Fischbacher, U. (2006). « Group Affiliation and Altruistic Norm Enforcement », *American Economic Review*, American Economic Association, vol. 96(2), pages 217-221.
- Borooh, V. K. (2005). « Caste, Inequality, and Poverty in India », *Review of Development Economics*, vol. 9(3), pages 399-414.
- Bowles, S. and Gintis, H. "Cooperation." *The New Palgrave Dictionary of Economics*. Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan, 2008.
- Branscombe, N. R., Schmitt, M.T., and Harvey R. D. (1999) « Perceiving Pervasive Discrimination Among African Americans: Implications for Group Identification and Well-Being », [Journal of Personality and Social Psychology](#); 77, 135-149
- [Chen, R., Chen, Y. \(2009\) « The potential of Social Identity for Equilibrium Selection », working paper.](#)
- Del Río, C., Gradín, C. and Cantó O. (2010). « The measurement of gender wage discrimination: the distributional approach revisited », *The Journal of Economic Inequality*)
- Deliège, R. (2004). « Les castes en Inde aujourd'hui », ed. PUF.
- Desai, S., Dubey, A., Joshi, B.L., Sen, M., Shariff, A. and Vanneman, R. (2009) « India Human Development Survey: Design and Data Quality », IHDS Technical Paper 1.
- Dumont L. (1966) « Homo Hierarchicus : essai sur le système de castes », ed. Gallimard.
- Goette, L., Huffman, D. and Meier, S. (2006). « The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence using Random Assignment to Real Social Groups », *American Economic Review*, American Economic Association, vol. 92(2), pages 212-216.
- Hadi, A. S. (1992) « Identifying Multiple Outliers in Multivariate Data », *Journal of the Royal Statistical Society. Series B* Vol. 54, No. 3 (1992), pp. 761-771
- Hoff, K., and Pandey, P. (2006) « Discrimination, Social Identity, and Durable Inequalities », *The American Economic Review*, American Economic Association, vol. 96(2), pages 206-211.
- Khwaja, A. I. (2009). « Can good projects succeed in bad communities? », *Journal of Public Economics*, vol. 93(7), pages 899–916.
- McLeish, K. et Oxoby, R. (2007). « Identity, Cooperation, and Punishment », IZA Discussion Papers, Institute for the Study of Labor (IZA).
- McLeish, K. et Oxoby, R. (2011). « Social interactions and the salience of social identity », *Journal of Economic Psychology*, 32 pages 172-178.
- Miguel, E., Gugerty, M.K., (2005) « Ethnic diversity, social sanction, and public goods in Kenya », *Journal of public economics* 89(11-12)
- Montalvo, J.G. and Reynal-Querol, M., (2003) « Religious polarization and economic

development » *Economics Letters*, Elsevier, vol. 80(2), pages 201-210, August.

Oaxaca, R. (1973). « Male-Female Wage Differentials in Urban Labor Markets. » *International Economic Review* 14: 693–709.

Olson, M. (1965). « The logic of collective action: public goods and the theory of groups », Harvard Economic Studies Vol. 124. Harvard University Press, Cambridge.

Olken, B. A. (2008). « Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia », NBER Working Papers 14123, National Bureau of Economic Research, Inc.

Sethi, R. and Somanathan, E, 2006. « A Simple Model of Collective Action », [Economic Development and Cultural Change](#), University of Chicago Press, vol. 54(3), pages 725-47.

Singh (2004)

Singh, P. (2011) « We-ness and Welfare: A Longitudinal Analysis of Social Development in Kerala, India » *World Development*, Elsevier, vol. 39(2), pages 282-293, February.

Thorat, S. and Newman, K. S. (2010)« Blocked by caste », Oxford University press

Vigdor, J.L. (2002) « Interpreting Ethnic Fragmentation Effects », *Economic letters*, 75(2), 271-276

Vigdor, J.L. (2004) « Community composition and CA: Analyzing initial mail response to the 2000 Census », *Review of Economics and Statistics*, 86(1), 303-312

Annex 1. Descriptive statistics

Household variables

	Observations	Mean	Std.	Min.	Max.
Income	26733	41194.4	69805.72	-108327.8	3945151
Land	26598	2.192858	5.501518	0	200
Land Gap	23510	-2.936466	16.45299	-481.9873	199.368
Education	25137	6.157935	4.846223	0	15
Nb Adults	26734	2.820042	1.410575	0	14
Mainly Farmer*	26734	35.63	-	-	-
Main Group*	26,734	47.79	-	-	-

*:dummy variables: the mean actually indicates the proportion of mainly farmer households, or households member of the most numerous group in their village.

Caste groups	Freq.	Percent	Cum.
Brahmin	1,092	4.08	4.08
OBC	10,703	40.04	44.12
SC	5,952	22.26	66.38
ST	2,936	10.98	77.37
Other	6,051	22.63	100.00
Total	26,734	100.00	

Dominant Caste	Freq.	Percent	Cum
Brahmin	24	1.58	1.58
SC	74	4.87	6.45
ST	122	8.02	14.47
Other	1301	85.53	100.00
Total	1,521	100.00	

Village variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Share Muslim	1408	.088821	.215274	0	1
Share Christian	1408	.0191832	.102874	0	1
Share Tribal	1408	.0096307	.0873181	0	1
Share SC	1408	.187294	.194871	0	1
Share ST	1408	.1218253	.26002	0	1
Share Brahmin	1408	.045973	.1270782	0	1
Fractionalization Index	1408	.5918539	.2095463	0	.865
Caste Frac.	1408	.440544	.2143398	0	.9305
Religious Frac.	1408	.1493354	.1993432	0	1
Polarization Index *	1408	.668065	.1910349	0	1
Caste RQ	1408	.675351	.2732703	0	1
Religious RQ	1408	.2755035	.3404272	0	1
Gini Land	1372	.3728095	.1506648	0	.838779
Theil Inter	1223	1301.726	2263.493	.2164202	29201.26
Local Public Assembly**	1481	22.96	-	-	-
Public Water Programm**	1481	60.23	-	-	-

*: Montalvo and Reynal-Querol (2003) polarisation index.

** : dummy variables, the mean actually indicates the percentage of villages with event happening.

Village main water source	Freq.	Percent	Cum.
Piped	598	40.41	40.41
Tube well	198	13.38	53.78
Hand pump	485	32.77	86.55
Dug, open well	132	8.92	95.47
Covered well	29	1.96	97.43
River, canal	16	1.08	98.51
Pond	7	0.47	98.99
Tanker truck	9	0.61	99.59
Rainwater	2	0.14	99.73
Bottled	1	0.07	99.80
Other	3	0.20	100.00
Total	1,480	100.00	

Annex 2: Propensity score matching estimate.

VARIABLES	Full sample SC	Matched sample SC
Occupation	-0.0281*** (0.00532)	-0.0126* (0.00654)
Migrant	-0.00197 (0.0333)	-0.0734* (0.0399)
Nb ADULTS	0.0390*** (0.0102)	0.0151 (0.0123)
Nb TEEN	0.0274** (0.0121)	-0.00624 (0.0144)
Owens Land (dummy)	-0.0745 (0.0504)	-0.0249 (0.0641)
Amount Owned	-0.124*** (0.0104)	-0.000197 (0.0156)
Irrigated land (dummy)	-0.129** (0.0640)	-0.0515 (0.0834)
Share Irrigated (categorical)	0.0119 (0.0253)	0.00246 (0.0336)
Hired farm Labour	0.00714 (0.0116)	0.00704 (0.0138)
Productive Assets	-0.00666** (0.00279)	0.00239 (0.00398)
Big Animals	-0.105*** (0.0195)	-0.0574** (0.0262)
Small Animal	-0.00153 (0.00654)	0.000592 (0.00980)
Works at home	-0.324*** (0.0515)	-0.0710 (0.0669)
Works in fixed place	-0.439*** (0.0534)	-0.0466 (0.0715)
Moves for work	-0.144** (0.0602)	-0.167** (0.0713)
Hired Buissness Labour	-0.358*** (0.0862)	0.140 (0.130)
Income from Rent	-0.216** (0.0919)	-0.0387 (0.124)
Income from Pension	-0.181*** (0.0660)	-0.0174 (0.0858)
Income from Sale	-0.167 (0.104)	-0.210 (0.129)
Max Adult Education	-0.0266*** (0.00277)	-0.00720** (0.00333)
Administrative Work	-0.0909** (0.0361)	0.0141 (0.0458)
Nb of HH in Village	-9.44e-05*** (1.35e-05)	8.01e-06 (1.75e-05)
Village access by road	-0.000269 (0.0221)	0.0165 (0.0267)
Pct Village eletrified	0.00123*** (0.000394)	0.000698 (0.000470)

Closest Market	-0.000304 (0.00202)	0.00414* (0.00238)
Dominant Grp (dummy)	-0.697*** (0.0240)	0.0283 (0.0315)
District	-0.00279*** (0.000871)	-0.00125 (0.00105)
State	0.267*** (0.0871)	0.121 (0.105)
Constant	0.379*** (0.0774)	0.0590 (0.0937)
_treated		
Observations	15,876	8,580
R-squared	0.1409	0.0044
log likelihood	-7965	-5921

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

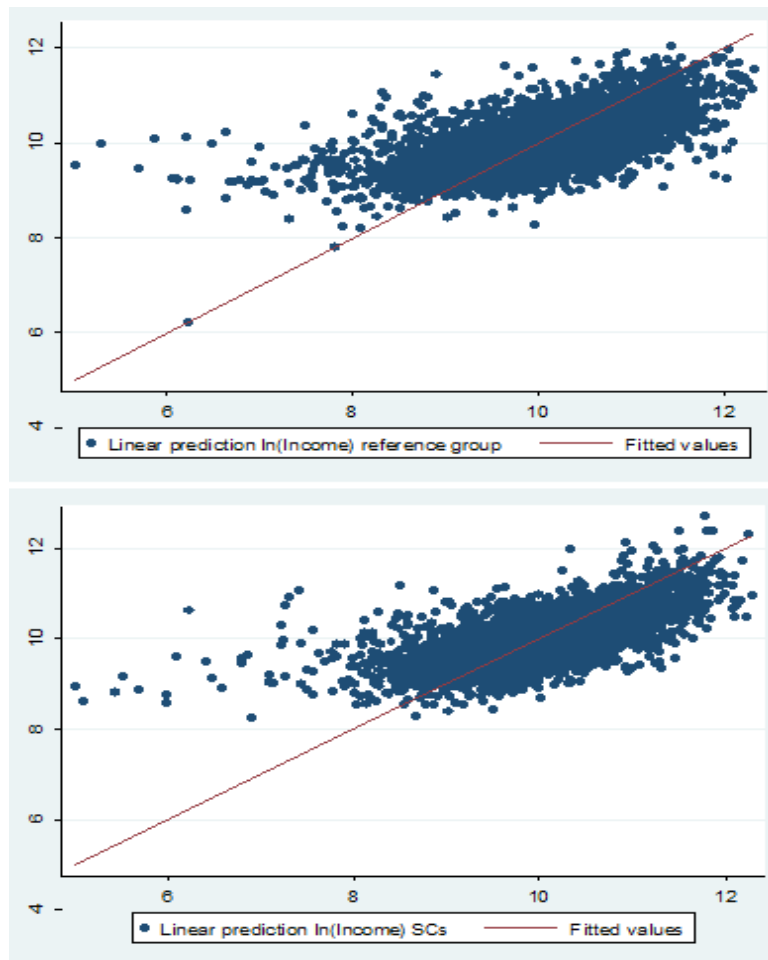
Annex 3: Income equations for SCs and their counterfactual.

VARIABLES	Reference Grp ln(Income)	SCs ln(Income)
Occup: Allied Agr	0.418*** (0.131)	0.355** (0.158)
Occup: Ag Labour	0.312*** (0.0624)	0.146*** (0.0558)
Occup: Non-Ag Labour	0.358*** (0.0624)	0.301*** (0.0579)
Occup: Artisan	0.433*** (0.0841)	0.305*** (0.0880)
Occup: Petty trade	0.298*** (0.106)	0.303*** (0.0917)
Occup: Business	0.653*** (0.107)	0.299*** (0.0836)
Occup: Salaried	0.777*** (0.0864)	0.741*** (0.0761)
Occup: Profession	0.443*** (0.155)	0.309 (0.318)
Occup: Pension/rent	0.370** (0.169)	0.290* (0.149)
Occup: Others	0.0712 (0.146)	-0.375 (0.316)
Works at home	0.196** (0.0808)	0.222*** (0.0588)
Works in fixed place	0.198** (0.0772)	0.307*** (0.0629)
Moves for work	0.148* (0.0767)	0.167*** (0.0645)
Migrant	0.0196 (0.0471)	0.0822** (0.0373)
Dominant Grp Member	0.109*** (0.0389)	0.0350 (0.0355)
Land Owner	-0.0844* (0.0488)	-0.107*** (0.0373)
Big Animals	0.0877*** (0.0286)	0.114*** (0.0286)
Nb adults in HH (NAD)	0.111*** (0.0318)	0.128*** (0.0248)
NAD*west	0.104** (0.0474)	-0.0129 (0.0482)
NAD*south	0.0420 (0.0431)	0.0660* (0.0377)
NAD*east	0.0151 (0.0437)	-0.000400 (0.0342)
NAD*center	0.0380 (0.0370)	0.0322 (0.0297)
Nb teen in HH (TEEN)	0.178*** (0.0622)	0.0964*** (0.0287)
NTEEN*west	-0.108 (0.0774)	0.0284 (0.0447)

NTEEN*south	0.0290 (0.0700)	0.145*** (0.0529)
NTEEN*east	-0.0822 (0.0697)	0.0679* (0.0391)
NTEEN*center	-0.0753 (0.0658)	0.0611* (0.0352)
Index Pive Assets	-0.0169* (0.00906)	-0.00147 (0.00972)
PiveA*west	0.0138 (0.0133)	0.0149 (0.0137)
PiveA*south	0.0164 (0.0172)	-0.0254 (0.0160)
PiveA*east	0.0292** (0.0136)	-0.0167 (0.0118)
PiveA*center	0.0218** (0.00973)	-0.00366 (0.0107)
Max education (HHEDU)	0.0320*** (0.0106)	0.0363*** (0.0114)
HEDU*west	-0.00225 (0.0139)	-0.0275* (0.0151)
HEDU*south	-0.0160 (0.0131)	-0.0211 (0.0135)
HEDU*east	0.0220* (0.0133)	-0.00775 (0.0130)
HEDU*center	-0.0101 (0.0119)	-0.0259** (0.0122)
Land owned	0.136*** (0.0402)	0.155*** (0.0510)
Land*west	-0.0942* (0.0568)	-0.130** (0.0575)
Land*east	-0.0253 (0.0739)	0.0516 (0.0586)
Land*south	-0.0714 (0.0660)	-0.0735 (0.0595)
Land*center	-0.0559 (0.0436)	-0.0913* (0.0531)
Village road	0.103*** (0.0296)	0.0210 (0.0314)
Closest Market	-0.00875** (0.00400)	-0.00361 (0.00293)
Nb HH in Village	1.08e-05 (2.67e-05)	3.84e-05* (2.11e-05)
Constant	8.545*** (0.326)	8.868*** (0.183)
District Fixed Effects	yes	yes
Observations	3,975	4,019
R-squared	0.422	0.425

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Annex 4: Predicted incomes compared to actual incomes for the reference group and SCs



Annex 5: Cooperation determinants, with village random effects.

VARIABLES	(S3) Coop	(S4) Coop	(S5) Coop	(S8) Coop
Distance Variables				
Distanciation as SC	0.253*** (0.0656)	0.235*** (0.0711)	0.225** (0.110)	0.199* (0.121)
Individual Controls				
Positive Discrimination	0.276*** (0.0658)	0.231** (0.0963)	0.150 (0.174)	0.136 (0.189)
SC	-0.0587** (0.0290)	-0.0696* (0.0415)	0.00506 (0.0489)	-0.0902 (0.0761)
ST	-0.0547 (0.0359)	0.0504 (0.0518)	0.0321 (0.0674)	0.0276 (0.103)
OBC	-	-	-	-
Brahmin	-	-	-	-
Muslim	-0.0355 (0.0374)	0.0565 (0.0460)	0.0254 (0.0698)	-0.00594 (0.107)
Christian	0.144 (0.0987)	-	0.140 (0.257)	-
Tribal	0.338*** (0.0526)	-	0.757 (0.546)	-
Member dominant group	-0.0456 (0.0418)	0.0352 (0.0483)	-0.00861 (0.0769)	0.0271 (0.0866)
SC*DominantGroup	0.0553 (0.0428)	-0.0594 (0.0695)	-0.0726 (0.0859)	-0.268** (0.122)
ST*DominantGroup	0.0846 (0.0559)	0.0503 (0.102)	-0.122 (0.108)	-0.346 (0.216)
Brahmin*DominantGroup	0.0263 (0.0402)	-0.0211 (0.0481)	0.0510 (0.0756)	0.0817 (0.0865)
Muslim'DominantGroup	0.0701 (0.0510)	0.0940 (0.0750)	-0.0223 (0.113)	-0.0325 (0.198)
Tribal*DominantGroup	-	-	5.135 (419.4)	-
Christian*DominantGroup		-0.326* (0.174)		0.142 (0.469)
Muslim*ShareMuslim		-0.820*** (0.315)		0.117 (0.749)
Christian*ShareChristian		-		-
Tribal*ShareTribal		-		-
SC*ShareSC		0.0982 (0.209)		0.148 (0.419)
ST*ShareST		-1.129*** (0.366)		-0.647 (0.717)
OBC*ShareOBC		-0.0321 (0.0393)		-0.162** (0.0696)
Brahmin*ShareBrahmin		-0.284 (0.350)		-0.245 (0.736)
(Muslim*ShareMuslim) ²		0.698** (0.293)		-0.227 (0.689)

(Christian*ShareChristian) ²		-		-
(Tribal*ShareTribal) ²		-		-
(SC*ShareSC) ²		-0.0618 (0.211)		0.544 (0.536)
(ST*ShareST) ²		1.276*** (0.331)		1.088* (0.648)
(OBC*ShareOBC) ²		-		-
(Brahmin*ShareBrahmin) ²		0.0763 (0.318)		-0.0651 (1.050)
Owned Land	0.00250 (0.00289)	0.00361 (0.00376)	0.0163*** (0.00527)	0.00669 (0.00832)
Income	4.60e-08 (1.98e-07)	3.27e-08 (2.25e-07)	-5.25e-07 (3.77e-07)	-6.05e-07 (5.66e-07)
Gap between Average and Household owned Land		0.00191 (0.00119)		0.00860 (0.00556)
(Gap between Average and Household owned Land) ²		-2.11e-06 (6.94e-06)		-1.31e-05 (6.53e-06)
Mainly farmer household	-0.0147 (0.0156)	-0.0196 (0.0175)	-0.0629** (0.0279)	-0.0607* (0.0314)
Owns Tubewell	-0.00353 (0.0226)	-0.0200 (0.0247)	0.0576 (0.0381)	0.0561 (0.0434)
Owns Electric Pump		-0.0126 (0.0261)		-0.00837 (0.0477)
Owns Gaz Pump		-0.0214 (0.0242)		-0.0972* (0.0515)
Maximum Adult Education	-0.000691 (0.00129)		0.00320 (0.00263)	
Maximum Male Education		0.000392 (0.00145)		0.00570** (0.00289)
Number of adults	0.00384 (0.00463)	0.00598 (0.00522)	-0.00237 (0.00907)	0.00214 (0.0100)
Member local assembly	-0.0301 (0.0427)	-0.0571 (0.0416)	-0.167** (0.0829)	-0.231** (0.0900)
Relative member local assembly	0.0618*** (0.0216)	0.0828*** (0.0239)	0.181*** (0.0370)	0.208*** (0.0412)
Main Water Sources: for household, interaction with village one, village one				
Piped	-0.0135 (0.0672)	-0.0702 (0.0659)	0.0110 (0.0973)	0.179 (0.110)
Handpump	-0.0135 (0.0553)	-0.0293 (0.0543)	0.139 (0.0902)	0.215** (0.100)
Openwell	-0.0159 (0.0552)	-0.0427 (0.0536)	-0.0130 (0.0892)	0.113 (0.0989)
River	0.164*** (0.0572)	0.0920 (0.0916)	0.360** (0.142)	0.149 (0.170)
Tubwell	-0.0736 (0.0677)	-0.0923 (0.0638)	0.117 (0.0998)	0.169 (0.110)
Piped*VillagePiped	0.0859* (0.0510)	0.111** (0.0542)	0.130* (0.0730)	0.0513 (0.0831)
HandPump*VillageHandPump	-0.0481	-0.0397	-0.113*	-0.0944

OpenWell*VillageOpenWell	(0.0458) -0.0603 (0.0606)	(0.0496) -0.0403 (0.0651)	(0.0662) -0.139 (0.0978)	(0.0719) -0.198* (0.107)
River*VillageRiver	-	-	5.235 (2,101)	5.466 (1,188)
TubeWell*VillageTubewell	-0.146** (0.0696)	-0.148** (0.0687)	-0.398*** (0.101)	-0.432*** (0.111)
VillagePiped	0.0421 (0.0648)	-0.00571 (0.0896)		
VillageTubewell	0.0700 (0.0608)	0.0129 (0.0888)		
VillageHandPump	0.0971 (0.0633)	0.0305 (0.0873)		
VillageOpenwell	0.0389 (0.0705)	0.0195 (0.0938)		
VillageCoveredwell	0.189*** (0.0594)	0.185** (0.0885)		
VillageRiverCanalPond	0.138 (0.0932)	0.0483 (0.117)		
Village Controls				
Polarization index		0.128* (0.0668)		
Inter Group Land Disparity		5.53e-06 (4.49e-06)		
ShareMuslim		(0.237) 0.230		
ShareChristian		(0.250) 0.576		
ShareTribal		0.623** (0.277)		
ShareSC		0.0790 (0.267)		
ShareST		-0.138 (0.268)		
Share OBC		-0.0591 (0.263)		
ShareBrahmin		-		
Public programm Safe Water	-0.0554** (0.0229)	-0.0633** (0.0252)		
Official local assembly building	-0.0275 (0.0260)	-0.0200 (0.0290)		
Village Random Effects	no	no	Yes (1,513 villages)	Yes (1,290 villages)
Observations	23,647	18,623	24,092	19,684
log likelihood	-8.000e+07	-6.110e+07	-12208	-9930

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1