Capturing minds of the oppressed: Evidence from a lab-in-field experiment

Bruhan Konda[1a], Eleonora Nillessen[a] and Stephan Dietrich[a]

[a] *United Nations University (UNU MERIT), Maastricht, Netherlands*

[1] Corresponding author. Email: konda@merit.unu.edu

Abstract

Social identity and differences in social power between groups play a crucial role in determining the success of a collective action. More recent research points to the possibility that inter-group relationships rather than diversity per se, determine the nature of cooperation. We hypothesize that perceived high-status identity associated with elite groups perpetuate normative behaviour that favours them and groups with perceived low-status identity anticipate and conform to these norms.

Using individuals from the top and bottom of the caste hierarchy in a lab-in-field experiment, we empirically test the role of caste differences in collective action. In particular, we test whether contribution and enforcement behaviour in a public goods game varies with group (caste) heterogeneity and differences in marginal returns and assess its implications on social welfare. We find that the individuals from both high and low castes are very sensitive to the differences in marginal returns from the public good.

Both these caste groups exhibit a caste conflict behaviour in the norm enforcement wherein we observe a higher punishment levels in the heterogeneous caste groups as compared to homogeneous caste groups. It reflects on the existence of hostile relationship between these groups due to persistent dominance and discrimination by high castes against the low castes over centuries. Among homogenous caste groups, high caste members demonstrate higher cooperation levels than the low caste members due to a strong perception of punishment and stronger affinity towards their own social identity (homophily). Higher social welfare is obtained when high caste gets higher marginal returns from the public good.

We also use a priming exercise prior to the second set of games where individuals from both high and low castes are exposed to the stories of prominent individuals (role models) from their own caste groups who have excelled and contributed in various walks of life. It is attempted to test if it can nullify the caste differences in cooperation due to increase in confidence and positive self-image of their identity. We witness a reduction of differences in cooperation between high and low caste individuals in the homogeneous caste groups with the caste conflict behaviour remain intact. Our results suggest that although the low caste members have developed courage to strongly retaliate the discrimination by high caste members, the lack of affinity towards their own identity still remain a great concern. The policies should therefore not limit their focus only on strengthening the economic conditions and political mobilization of the low caste individuals, but also on addressing the image of the negative stereotype they carry about their own identity.

Keywords: Social identity, Caste, Public goods, Punishment, Social welfare

1. Introduction

How social identity, group membership and intergroup differences affect an individual's behavior in a collective action has been one of the important features of economic and behavioral studies in the recent times. Commonly individuals will be part of certain identities such as religion, race, nationality, caste, tribe etc. Akerlof and Kranton (2000) considers how one's sense of self (identity) affect their behavior and thereby economic outcomes. In their utility function they include an individual belonging to certain identity share some common values, follows the norms/prescriptions set by the groups and thus various social categories in societies live together with multiple norms being abide by the members of its groups. These norms affect an individual's choice in terms of food, marriage, people they frequently interact with, labor market, politics etc. Any country would contain many social groups that are formed on certain basis. For example, in United States there are blacks and whites where the groups are formed on the basis of race. In India there are many religious groups such as Hindus, Muslims, Christians, Buddhists, Sikhs etc. Among Hindus, there are several castes and each of them act as one social group. Understanding the relationship and differences among these groups are important for the development of a nation. Social identities also create homophily due to certain commonness among the members of the same identity and differences across social groups. Individuals also derive self-esteem from the positive views about their social group (Köszegi, 2006) and are known to conform to the stereotypes when the group identity is salient (Benjamin et al., 2010).

There are some studies which have examined the role of inter-group relationships on cooperation by focusing on specific social identities in diverse countries like India (Hoff et al., 2011; Waring and Bell, 2013), and Kenya (Habyarimana et al., 2007). These studies showed the role of differences in the punishment behavior, social dominance and differences in tastes and preferences in affecting their cooperative behavior. Therefore, it suggests that, it is not merely the presence of several social groups, but it is the inter-relationships between them that determine their cooperation.

Another important type of social groups which are formed mainly through the political processes and social power structures are elites. They are known to be the power centers who determine the allocation of public resources in society. These groups are known to negatively affect the welfare of the society through their formal/informal influence on the allocation of benefits and/or selection of beneficiaries for the public programs in their favor (Platteau and Abraham, 2002; Fritzen, 2007; Alatas et al., 2013; Martinez-Bravo et al., 2017). On the other hand, some other literatures have shown a positive influence of local elites in the participatory environments (Dasgupta and Beard, 2007; Lund and Jensen, 2013). The local elites (not elected) usually arise through the social power differences in communities. Therefore, understanding the complex power relationships between social groups is more important than only looking at the political legitimacy of formal elites to challenge the practices of elite capture (Musgrave and Wong, 2016).

This paper attempts to understand how the elites who are formed due to the social power differences between the groups can manage to capture higher proportion of benefits from the public goods in the absence of any power to distribute the resources but by manipulating the normative behavior of the disadvantaged groups for their benefits. In specific, we try to understand the complex relationship between the members from top and bottom of the caste hierarchy in India in determining the success of a collective action. The high caste members are referred as elites as they decorate the top of the social structure with the highest social power. On the other hand, the low caste (commonly called as

Dalits) sits at the bottom of the social structure who was highly discriminated, ostracized, oppressed, mistreated over centuries by the high caste members using severe social sanctions. Caste remain as one of the most important factors in determining social status in India in addition to wealth, education etc. This division is only based on the social status but not economic or any other parameters. This has not only lead to the loss of economic and social wellbeing but also to the loss of self-confidence (Hoff et al, 2011) among low castes due to the stigma attached to their identity (Bros, 2014). To correct the damage, the government has been following the affirmative action policies to bring them back to the main stream. But they are argued to have not solved the issue of self-image problems (Bros, 2014) and also it has not reached every low caste individual (Mondal, 2014). For instance, *Times of India (June 1, 2018)*[2] reported that a Dalit man was forced to shave his mustache for using a high caste's last name in an invitation for a religious ceremony. This incidence indicates two important elements of our discussion. Firstly, the Dalit man was attempting to change his last name into a high caste last name shows the stigma attached with their identity. Secondly, it indicates the persistent dominance of high caste in the Indian society, although it has been reduced over years. The study is conducted in Uttar Pradesh which has recorded one of the highest crimes (in the recent years) against low castes, as reported by *The Indian Express (April 14, 2018)*[3] using data from National Crime Records Bureau.

There are important differences between the political/economic elites and the social elites (eg: based on social identity: caste). There are certain possibilities for anybody to become political/economic elite whereas the caste elites are fixed as it is determined only by birth and therefore there is a seize of movement across the hierarchy. There is an information asymmetry between elites and non-elites in case of political elites whereas the information is complete in case of caste elites, i.e., both high and low caste members know each other's behavior very well as they interact in their everyday activities. In order to obtain higher benefits from the public resources, the political elites have to use the illicit means or the power of position. The caste elites on the other hand perpetuate normative behavior which benefits them and make low caste members to conform to these norms using social sanctions.

Successful economic and social development is critically dependent on well-functioning institutions (Rodrik et al, 2004). Yet many developing countries face weak formal institutional arrangements and rely on social norms for contract enforcement and solving collective action dilemmas. Individuals or groups enforce such norms through their willingness to sanction anyone who deviates from the set norm. There exists now a wealth of literature that demonstrates the general effectiveness of having second or third-party punishment on solving collective action problems (Fehr and Gächter, 2000; Fehr and Fishbacher, 2004a; Fehr and Fishbacher, 2004b; Cinyabuguma et al., 2005; Charness et al., 2007). In addition to the burgeoning literature using laboratory experiments there is an increasing interest in understanding the role of punishment in public goods provision among real-world social groups. Bernard et al., (2006) examine the extent to which punishment in a DG varies between and within native tribes in Papua New Guinea and finds that deviations from an egalitarian norm are more heavily punished if both the recipient ("victim") and third party punisher belong to the same tribe compared to a treatment where the recipient and punisher belong to different tribes. Goette et al., (2006) use random assignment of Swiss Army officers to platoons during a four-week officer training to measure cooperation between and within these groups in a series of simultaneous PDs. They find cooperation

[2] https://timesofindia.indiatimes.com/city/rajkot/thakor-youth-thrashed-for-using-sinh-with name/articleshow/64389408.cms

[3] https://indianexpress.com/article/india/crimes-against-sc-sts-140-higher-than-that-of-general-public-in-gujarat-data-5136267/

between members of the same group is higher than that between members of different groups, yet punishment is (only) stronger in cases where defection impacts members of the in-group, suggesting group assignment is conducive to enforcing norms of cooperative behavior. Hoff et al., (2011) focus on social groups that are defined by their position in the extremely hierarchical Indian caste system and measures individuals' willingness to punish norms violators. They specifically test three hypotheses that also feature in our study: the "caste conflict hypothesis", predicting higher levels of punishment in mixed caste groups vis-à-vis groups whose members all belong to the same caste; the "caste submission hypothesis" where members of the low caste are expected to tolerate norm violations by members of the high caste (but not others) due to fear and experience of retaliation by high caste members; and "the caste culture hypothesis" where high caste members punish norm violators more severely when the affected person belongs to in-group than low caste members. Results show strong support for the caste culture hypothesis while rejecting the caste conflict and caste submission hypothesis, suggesting in-group affiliation is higher among members of the high caste.

Collective action has been the strategy for conservation and development of many resources such as forestry, fishery, irrigation etc. which involve community participation. Since all these resources are embedded in a complex socio-ecological environment, it is necessary to consider all the multiple subsystems in order to successfully manage these resources (Ostrom, 2009). Ostrom has designed a general framework for better understanding these systems for its sustainable use. It includes understanding interlinkage between several subsystems such as resource units, resource system, governance system, user's characteristics and their interactions with social, economic and political settings and other related ecosystems. Inspired by this framework, we have also included caste identity as the user characteristics, voluntary cooperation with punishment of norm violators as the governance system, with a resource (public good) generating heterogeneous marginal benefits. The heterogeneous benefits are commonly seen in many public goods in real life and this will also help us to understand the elite capture effect better. That is, we attempt to answer how the high and low castes behave differently under high and low marginal returns in a heterogenous society.

Presence of returns heterogeneity introduces tension between the individuals not only to cooperate or not but also to decide on the level of cooperation based on their incentives. A few studies which analyzed the role of heterogeneous public goods with respect to marginal returns on contributions have obtained mixed results (Fisher et al., 1995; Reuben and Riedl, 2009; Reuben and Riedl, 2013). This raises the speculation on the differences in the way the benefits from the public goods are interpreted by different individuals. It also reflects on the different contribution norms entailed by the individuals or groups in a society. All these studies have been conducted with the standard subject pool but on the other hand, the social norms are specific to location and context. Therefore, it is important to conduct these studies using the individuals from the field which helps us to understand their contribution behavior under a given context. Our study therefore merges both caste and returns heterogeneity by conducting a public good experiment involving individuals from high and low castes and by introducing heterogeneous returns to the public good. The design to include different caste individuals to the public good game is adapted from the study by Hoff et al (2011).

Our research also extends the study by Hoff et al (2011) in a number of ways. First, we randomly vary the marginal returns to investing in the public good, extending the evidence from laboratory experiments on this topic (see e.g. Fehr and Schmidt, 1999; Reuben and Riedl, 2013). Specifically, we

expect differences in contributions as a result of different marginal returns to co-vary with caste type. Second, we use a priming exercise to test whether any possible differences in contribution and punishment between high and low caste members (possibly due to differences in social identity) are sensitive to exposure to a role model of their own caste. Our results are as follows.

We find that the individuals are highly sensitive to changes in marginal returns from the public good. The cooperation is stronger among high castes compared to low castes due to higher response by the high caste members under punishment. Comparing the actual punishment behavior, we find high and low caste individuals strongly punish the other than their own caste members showing a caste conflict behavior. Under both caste and returns heterogeneity, high caste members captured higher payoff from the public good when they are given higher marginal returns, whereas low caste's payoff is indifferent between high and low returns scenario indicating the strong elite capture effect. The total welfare is also higher when high caste gets higher marginal returns, but it is coupled with inequality in the gain from pubic good.

Priming with role models improved low caste's cooperation among homogeneous groups and also reduced the significance of elite capture effect. Therefore, the low castes managed to obtain higher payoffs when they are given higher marginal returns. This also reduced difference in welfare between the conditions of high and low castes getting higher returns from the public good. Overall priming also reaffirmed the role of caste identities in determining the success of collective action.

1.1. Context

A. Caste system

Caste is a social system of identification and segregation followed mostly in India among Hindus. Any individual subscribes to a caste by the virtue of his/her birth and it is fixed for life time indicating no free mobility across castes. Caste is derived from *Varnas* - an ancient system of categorizing societies based on the occupation and skills (class system) pursued by the individuals (Deshpande and Kerbo, 2010). According to Varna system, any individual will fall into one of these ordered social categories, viz., Brahmins (Knowledge), Kshatriyas (military), Vysyas (trade) and Shudras (labor) and the Dalits/untouchables being the last category was kept outside the system of varnas. Each of these four caste categories contain number of sub-categories called as *Jati*. The system of endogamy was followed in all Jatis and therefore, mobility across Jatis is largely absent.

Although this was believed to be a fluid system where any individual can be part of any of these categories based on their skills, they became fixed over time where the subscription into a particular caste is determined only by birth. Thus, it can be described as a system of graded inequality (Deshpande, 2008). Therefore, caste system has been largely seen as a result of differentiation in the societies based on ritual purity where the caste at the top is considered as ritually pure and the bottom most as impure (Dumont, 1970). On the other hand, there is another theory which argues that this system of segregation and discrimination is the result of British colonization who strengthened caste affiliation and introduced preferential treatments based on castes for their economic and political interests (Dirks, 2011).

Post-independence, Indian constitution ensured affirmative action policies for the castes that are at the base of the caste pyramid with reservations in education, jobs and politics. This was done with an intent of correcting the past discrimination and oppression of those castes. For the administrative

purposes these castes are named as Scheduled Castes (SC) and Scheduled Tribes (Indigenous tribes - ST). Both SC's and ST's are commonly referred as low castes. The castes which were at the top of the pyramid are named as General castes which were commonly known as high castes. The middle of the pyramid is occupied by other castes which were neither at top nor at bottom and are named as Other Backward Castes (OBC). All these categories are made based only on one's caste identity but not through any other parameters.

B. Irrigation technology

Our study is also motivated by the design of an irrigation technology UTFI[4], which is installed at the community (village) level in Uttar Pradesh for the dual purpose of reducing the impact of floods and to recharge groundwater. The groundwater stored is pumped and used by the farms around the structure. The construction and maintenance of these structures require collective action from communities either in terms of contributing labor or money or both. Therefore, it can be viewed as a public good where there is a possibility to free ride but still to get benefit from the contribution of others. An important feature of this technology is that the farms close to the structure gets more recharge of groundwater than the farms located away from the structure. It introduces the variation in returns (in terms of amount of groundwater) across the farms from the public good (UTFI). On the other hand, the farms in villages belong to the members from different castes and therefore it creates a complex environment for cooperation with both project (technology) specific and social heterogeneities in which the public goods need to establish.

C. Combining both social and technological aspects of public good

A public good (Irrigation system) which yield different benefits and is jointly used by the farmers from different castes (social identity) is an essential element of our experimental design. Combining both the social and technological elements helps to better understand the interplay between them in affecting public good provision. Specially to see how individuals from both top and bottom of the caste hierarchy behave differently when they get differential returns from the public good. We are also inspired by the Nested Multitier Framework (Ostrom 2007) which demonstrates that the sustainable management of resources can be achieved by understanding not only the resource characteristics but also the user characteristics and the governance system as any resources used by humans are embedded in complex socio-ecological systems (Ostrom, 2009).

2. Framework and hypotheses

2.1. Framework

Social identity and differences in social power between groups play a crucial role in determining the success of a collective action (Miguel and Gugerty, 2005; Hoff et al., 2011; Chakravarty et al., 2016) We posit that perceived high-status identity associated with elite groups perpetuate normative behaviour that favours them. Groups with perceived low-status identity anticipate and conform to these norms. It will impair the production of public goods, perils social welfare and leads to higher inequality. It occurs through the differences in the group's ability to cooperate and to enforce social norms. Members of high-status identity utilize their social power to dominate and threaten the members from low-status identity to cooperate when it is beneficial to them. Whereas the members

[4] http://utfi.iwmi.org/

of low status anticipate it and cooperate to escape punishment due to lack of self-esteem. Hence the high-status members are able to abstract a higher portion of benefits from public goods without the need of actual (political) power and corruption (illicit actions) to distribute the public resources to their favour. It emphasizes the role of a socio-psychological relationship among groups of different social status in affecting public welfare.

Below we put forth the hypotheses which we test to understand and analyze the framework of the paper. We test this framework in the context of caste differences in India where high and low castes represent the high and low status communities in society. It represents the pathway in which the high and low caste behavior unfolds and can be understood.

2.2. Hypotheses

Hypothesis 1

Both high and low caste members contribute alike when they do not know the caste of their partners.

It suggests that among many other factors, the contribution is influenced by the identity of the partner or the group members. With respect to high and low caste, when identity is not known it shows their inherently similar behavior of contributing to the public goods in a society. If this hypothesis is accepted, it fairly crowds out the effect of differences in the underlying preferences of high and low caste members (by the virtue of their own identity) on their contribution behavior. Therefore, any difference in their contributions when the caste identity is revealed can be attributed to their caste differences with respect to partner's caste.

$$C_{HA} = C_{LA} \tag{1}$$

Where C denotes the mean contributions. HA and LA denotes High caste and Low caste members respectively with anonymous partners.

Hypothesis 2

Knowing caste of the partner, the high and low caste members change their contributions to the public good.

When the caste of the partner is revealed, it eliminates the uncertainty attached with identity of the partner. Therefore, members contribute based on their preferences on the identity of the partner. These preferences are formed by their interactions in the daily life and/or by the historical relationships between those caste groups. This provide a focal point to the members which determine their level of cooperation. We posit that the members from different caste groups (rank) possess an affirmative opinion about their own caste rank groups despite belonging to the different specific castes. Specifically, high caste members possess a stronger reciprocal relationship than the low caste members due to their superior identity which is reflected through their collective power in discrimination against low caste over centuries. On the other hand, by carrying a negative self-image, the low caste members possess lesser affinity to their identity. Thus, the social superiority may lead to a stronger in-group cooperation among the high castes. Sidanius and Pratto (1999) argues that the social power differences between social groups influences the cooperation. Therefore, the social groups having similar and dominant social power (by status) might express a stronger cooperative behavior than the groups having similar and less social power.

Collective benefit incentive due to common social identity

$$C_{LL} > C_{LA} \quad (2)$$
$$C_{HH} > C_{HA} \quad (3)$$

Collective benefit incentive is stronger for high caste members as their group identity is superior over the low caste.

$$C_{HH} - C_{HA} > C_{LL} - C_{LA} \quad (4)$$

From equation (1), we can rewrite equation 4 as,

$$C_{HH} > C_{LL} \quad (5)$$

Hypothesis 3

Contribution varies by the marginal returns obtained from the public good.

Individuals with social preferences contribute positively to the public good. These contributions reflect the normative behavior of individuals which vary by the context of the public good in question (Bicchieri, 2008). The normative behavior can be broadly grouped into three categories based on the literature on the allocation principles (Konow, 2003; Konow *et al.,* 2009; Reuben and Riedl, 2013). They are efficiency, equality and equity principles. The efficiency principle suggests the allocation of the maximum amount to the public good as it increases the collective outcome both in homogeneous and heterogeneous benefits groups. Equality principle characterizes an equal outcome from the public good. In case of homogeneous groups, an equal contribution provides an equal outcome where as in heterogeneous groups, an equal contribution provides more to the individual with high marginal benefits and less to the individual with low marginal benefits. This introduces the principle based on equity wherein the individuals contribute in proportion to the benefits (marginal benefit) they obtain from the public good. In the public goods literature, effect of differential marginal returns on contributions has obtained mixed results (Fisher *et al.*, 1995; Reuben and Riedl, 2013), in spite of theoretical (Fehr and Schimidt, 1999; Bolton and Ockenfels, 2000) and empirical (Dawes *et al.*, 2007) evidence for the presence of egalitarian motives which affect the behavior. Hence, we hypothesize that the differences in marginal returns affect contribution behavior.

$$C_{High} > C_{Equal} > C_{Low} \quad (6)$$

Hypothesis 4

Punishment behavior varies by caste of the punisher and the punished. We adopt the hypotheses formulated by Hoff et al (2011) to test the effect of caste differences on punishment. They are as follows:

4a. Caste conflict hypothesis

It says that due to the prevalence of a conflicting environment between high and low castes, the punishment will be higher in the group where both high and low caste members play i.e., in the heterogeneous caste treatment than in the homogeneous caste treatment. The hostility lead to high

caste members punishing the low caste harshly and vice versa than punishing members from the same caste status. It is represented as,

$$Pun_{HL} > Pun_{HH} \quad (7)$$

$$Pun_{LH} > Pun_{LL} \quad (8)$$

where, Pun_{HL} indicates the mean punishment conditional on contributions by the high caste member to the low caste member. Similarly, other three parameters follow the same interpretation according to the caste matching.

4b. Caste submission hypothesis

According to this hypothesis, low caste shows a submissive behavior towards high caste's norm violations due to the latter's nature of fierce retaliation against them in case of any disagreements. Therefore, this hypothesis is in contrary to the caste conflict hypothesis for the low caste and is represented as follows,

$$Pun_{LH} < Pun_{LL} \quad (9)$$

4c. Caste culture hypothesis

It implies that the high caste members punish the norm violators more severely than the low caste irrespective of the partner's caste. The lower willingness to punish by the low caste members is due to the lack of self-confidence which has resulted from their social exclusion (Sen, 2000; Rao and Walton, 2004; as in Hoff *et al.*, 2011) and also due to their restriction from the opportunities to carryout economic activities. By controlling for the individual characteristics such as wealth and education, it is shown as,

$$Pun_H > Pun_L \quad (10)$$

where, Pun_H indicates that the punisher is from high caste (HH and HL) and similarly Pun_L indicates that the punisher is from the low caste (LL and LH)

Hypothesis 5

There is a caste-specific difference in the way differential marginal returns affect contributions.

That is, high and low caste members react differently under both high and low marginal return scenarios. High caste members compared to low caste are known to have more concerns about their social status (Mukherjee et al., 2017) and therefore, receiving higher returns may imply honoring their status and vice versa therefore may affect contributions. In addition, with punishment, high caste members are more likely to enforce favorable contribution patterns for them. Anticipating stronger sanctioning levels from high caste members, the low caste members adjust their contributions.

$$(C_{High(HL)} - C_{Low(HL)}) > (C_{High(LH)} - C_{Low(LH)}) \quad (11)$$

Where, $C_{High(HL)}$ and $C_{High(LH)}$ indicates the contribution of high caste and low caste respectively in a heterogeneous setting when they are given high marginal returns. Similarly, $C_{Low(HL))}$ and $C_{Low(LH)}$ indicates the contribution of high caste and low caste respectively in a heterogeneous setting when they are given low marginal returns.

Hypothesis 6

Under caste heterogeneity, high marginal returns for high caste members lead to higher social welfare compared with high marginal returns for the low caste.

High caste members are expected to react strongly either through lower contributions or through higher punishment levels when they get an unfavorable marginal return from the public good vis-à-vis low caste. This will affect the total welfare of the community. Therefore, we posit that the social welfare will be higher when the high caste gets better marginal returns than the low caste.

$$Welfare_{High(HL)} > Welfare_{High(LH)} \quad (12)$$

Where, $Welfare_{High(HL)}$ indicate welfare when High caste member is given higher marginal return and similarly, $Welfare_{High(LH)}$ indicate welfare when Low caste member is given higher marginal return.

3. Experiment details

3.1. Experiment design

We employ a one-shot linear public good game to investigate the effect of inter-caste relationships and heterogeneous returns on contributions and willingness to punish. We incorporate caste divisions by choosing castes at the extreme top and bottom of the caste hierarchy thereby following Hoff et al (2011).

The public good game involves two members A and B who play the game privately and separately from each other in their own village. Their decisions are anonymously matched to obtain the final payoff. Each member receives an endowment E $(E > 0)$ of 250 rupees, from which he/she can choose either to contribute any of the following amounts $c_i \in [0, E]$ viz., 0, 50, 100, 150, 200 or 250 to the public good where $i = 1,2$ indexes members. Each member contributes independently to the public good. The amount contributed to the private account (kept for himself) yields the same amount to the member. Each member gets his/her share ($\alpha_i \sum_{i=1}^{2} C_i$) from the total amount contributed by the group to the public good.

Our design includes treatments based on marginal returns, punishment and caste composition of the group. We implement two types of treatments based on the type of marginal returns a member gets from the public good; equal or unequal marginal returns. In case of equal returns, both members obtain $\alpha_i = 0.75$. In case of unequal returns, one of the members gets higher returns ($\alpha_i = 0.90$) and the other one gets the lower returns ($\alpha_i = 0.60$). The choice of marginal return values is such that no member has an absolute incentive for contributions as even higher marginal returns provide 0.10 less for one-unit contribution to the public good compared to investing in private account. The difference in marginal returns between high and low treatments are such that the high return member gets 50 percent more from the public good than that of the low return member for every unit of contributions to the PG. Both under equal and unequal returns, the $\sum \alpha = 1.5$, so that the total returns from the public good remain the same across treatments. We randomly assign individuals to either high or low returns in the unequal return treatment. The pay-off function looks as follows:

$$\pi_i = E - C_i + \alpha_i \sum_{i=1}^{2} C_i \quad (13)$$

In addition to the variation in the marginal returns to the public good, we run treatments with(out) the possibility to punish the matched group member. In the punishment treatment, members can punish each other after the contribution stage. We use peer punishment/second-party punishment where the punisher is not an outsider but an actual player in the game. The choice between the use of second party punishment and third-party punishment is debatable.[5] The main reason for us to choose a design with a second party punisher is to observe behavior of high and low caste members in the absence of a formal decision-making entity (reflecting a third party). It represents the real-life situation where members in any communities tend to follow social norms commonly due to peer pressure rather than because of an authority of a single person. In the experiment, we use a neutral phrase: 'opportunity to reduce the income of partner' rather than 'punishment'. Everyone received twelve 5-rupee coins as part of their endowment, which they could use to reduce the income of their partner after the contribution stage. Each 5-rupee coin spent, reduces the income of the partner by 20 rupees. They can spend any number of coins from 0 to 12[6]. Therefore, punishment is costly, and the cost of punishment is 0.25 for every rupee punished. We used the strategy method for the punishment stage, i.e., a member must indicate how many coins he/she wants to punish for every possible contribution amount by his/her partner. We use the strategy method primarily for practical reasons as players were in different villages and unable to learn about the action of the other.[7] The payoff function for the punishment treatment is therefore as follows.

$$\pi_i = E - C_i + \alpha_i \sum_{i=1}^{2} C_i - 20p_{21} - 5p_{12} \quad (14)$$

Where p_{21} denotes the number of units of punishment given to member 1 by member 2 and p_{12} denotes the number of units of punishment given by member 1 to member 2. Each member played both games i.e., without and with the punishment and hence it's a **with-in subject treatment[8]**. We randomized the order in which the members face the game with punishment stage.

Beside the treatments where participants were anonymously matched, we included treatments where the caste composition was revealed to learn how caste divisions affects public good provision. Therefore, we regarded members from two different caste ranks. Specifically, we sampled members from both the topmost (high caste) and the bottom most (low caste/scheduled caste) castes. We designed four groups varying the composition of high and low caste in the group. If both the members are from the high caste, we call it HH. Similarly, LL indicates that both are from the low caste. Both HH and LL form homogeneous caste groups. If one of the members is from a high caste and the other is

[5] Leibbrandt and López-pérez (2012) state that the motivations behind second party punishment are inequity aversion and spite (in which spite accounted for a smaller portion [13 percent] of punishment) whereas in the case of third party punishment motivations are typically related to inequity aversion (but see Kösfeld and Rustagi (2015) for evidence of anti-social punishment).

[6] The reason for using maximum of 12 coins is that any member who use all 12 coins (12X20 = 240) can reduce his/her partner's income close to zero (250-240=10)

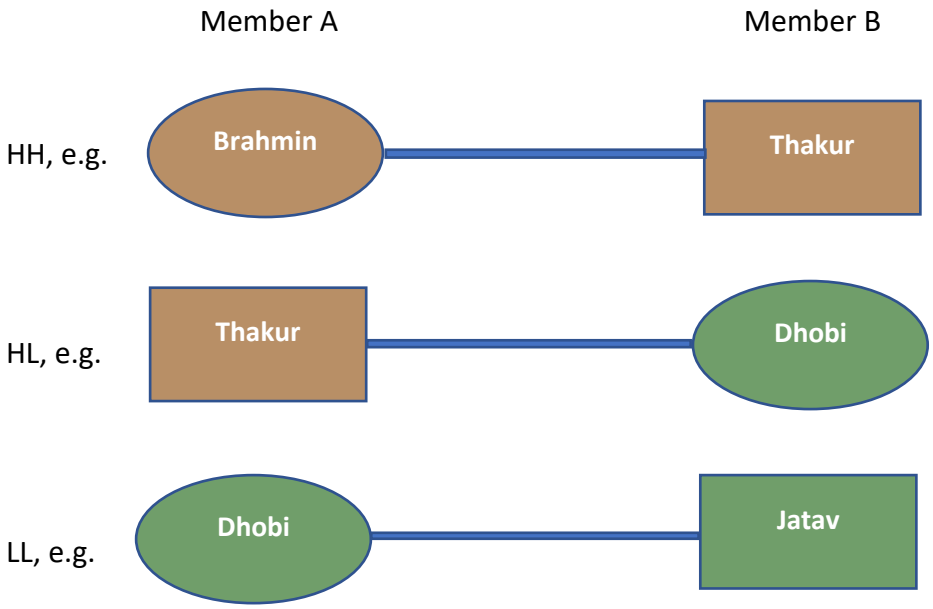
[7] Brandts and Charness (2011) show that most studies find no difference between the results obtained from the strategy method vis-a-vis real-time responses but has the advantage of having higher statistical power (Fehr and Fischbacher, 2004).

[8] The choice of with-in subject design for the punishment is to accommodate sufficient sample size in the given budget.

from a low caste, then we call it HL or LH and it is a heterogeneous (mixed) caste group. 'H' indicates that the individual belongs to the high caste (superior caste rank in the social hierarchy) and 'L' indicates that the individual is from the low caste (inferior caste rank also called untouchables, who represent the bottom of the social hierarchy). The notation also follows an order. The first letter in the notation always refer to the identity of the member and all the indications on return types and punishment treatment refers to him/her. The second letter in the notation refer to the identity of the partner.

The term caste we use in this paper refers to both status/rank and the membership in a specific caste. Each specific caste is an endogamous, mutually exclusive social group with its own set of traditions, occupations, networks that also provide goods and services to its members including for example informal insurance (Mazzocco and Saini, 2012; Mobarak and Rosenzweig, 2013), provision of loans (Munshi and Rosenzweig, 2015) etc. Caste status depends on whether a member's caste is situated in the top or bottom of the caste pyramid. There are several specific castes at every level of hierarchy. The specific castes that we observed in our sample are Brahmin, Thakur, Saxena and Gupta among high castes and Jatav, Dhobi and Valmiki among low castes. To account for both the status and association with specific caste, we designed the caste treatment following Hoff et al (2011). In case of a "homogeneous caste" treatment, both the members belong to the same caste rank but different specific caste. For example, in case of HH treatment, a member from Brahmin caste is matched with another from Thakur caste. Similarly, in case of LL, a Dhobi is matched with a Jatav. In case of heterogeneous caste treatment, it is obvious that both the members come from different specific castes but also belong to different hierarchies (caste ranks). The examples of the matchings can be seen in Figure 1.

Figure 1: Example of members in caste matching



In this figure, specific caste names in brown and green background indicate high caste and low caste respectively. Similar shapes of same color reflect same specific castes. If both the individuals would be from the same specific castes (which is possible only in case of homogeneous caste matching), we would for example have HH (Brahmin-Brahmin) or LL (Jatav-Jatav) whereas in the heterogeneous caste

matching i.e., HL (Brahmin-Dhobi), both individuals come from different ranks (high caste versus low caste) and from different specific castes confounding the effect of caste rank (status) and specific caste. To avoid these confounding effects where the difference in the treatments between homogeneous and heterogeneous matching can be attributed to ingroup favor (same specific caste), caste divide (different caste rank v/s different specific caste) or both, we have followed the design in which the matching the partner is always a member from different specific caste.

Since our experiment involves the use of social hierarchy based on caste differences, there are some ethical issues which needs to be considered. Firstly, we did ensure that the members will not know the real identity of his/her partners with whom they are matched during the experiment. Secondly, use of caste names in the experiment which is a politically sensitive issue was considered with lot of care. To address the challenge of anonymity, we considered the **strategy method** and we ensured that group members are from different villages. In order to make the issue of caste less salient (to avoid any potential experimenter demand effect and also because caste is a politically sensitive issue), we use the procedure followed by Hoff et al (2011), in which individual's last names were presented (with fictitious first names) which mostly indicates the specific caste to which the partner member belong to. This was used to convey information on caste of the partners to the members without explicitly mentioning it to them. The last names were selected based on our discussions with members from high and low castes and village heads during the pilot survey[9]. Given this common knowledge of members to identify the specific caste, we used last names as a less intrusive method to convey the caste information of the partner.

Table 1: Treatment conditions

Marginal Returns	Caste structure				
	HH	**LL**	**HL**	**LH**	**Anonymous**
No Punishment					
Equal	0.75_0.75	0.75_0.75	0.75_0.75	0.75_0.75	0.75_0.75
Unequal	0.9_0.6	0.9_0.6	0.9_0.6	0.9_0.6	0.6_0.9
With Punishment					
Equal	0.75_0.75	0.75_0.75	0.75_0.75	0.75_0.75	0.75_0.75
Unequal	0.9_0.6	0.9_0.6	0.9_0.6	0.9_0.6	0.6_0.9

Note: In the caste structure type (eg: HL) first letter refers to the caste of the player and the second letter indicate the caste of the partner. In the marginal returns (eg: 0.9_0.6), first value refers to the marginal return of the player and the second value indicate the marginal return of the partner.

The treatments are summarized in Table 1. In the homogeneous caste groups HH and LL, unequal marginal returns are implemented by randomly assigning high (0.9) and low (0.6) marginal returns to the members in the group. In case of heterogenous caste matching structure (HL), where both the members are distinct by their social hierarchy (High caste and Low caste), it is in our interest to analyze the effect of heterogeneous returns in a heterogeneous caste setting. To do this, we randomly assigned high returns and low returns to both high and low caste members. This essentially created two groups in case of heterogenous returns in a heterogeneous caste matching setting. The first group HL describes the condition where high caste member receives higher marginal returns and the low

[9] From the pilot survey, we gathered some common last names which was mostly indicating the specific caste of an individual. We tested the use of last names in the pilot experiment and asked all the individuals in the end whether it reflects anything in their mind. Almost all the individuals responded with the name of the caste that last name was indicating.

caste member receives lower marginal returns from the public good. The second group LH describes the condition of matching where the low caste member was assigned higher marginal returns whereas the high caste member was assigned with lower marginal returns from the public good. Finally, to see if there are any underlying differences in the behavior of high and low caste members, we designed another matching called 'Anonymous' in which the last names of the partner members were not presented during the experiment.

The treatments based on castes and marginal returns are **between subject treatments**. That is, if one of the members gets equal returns, another member also gets the equal returns, and both play with and without punishment games with the same type of returns. In case of unequal returns, if one of the members (ex: H) gets higher returns, then the other member (ex: L) gets the lower returns and vice versa. Similarly, they play both the games with the fixed returns type.

3.2. Common factors

The common factors in the experiment are as follows. First, the members were told about the homogeneity of endowments given to all, i.e., INR 250. Second, they were informed about the marginal returns from the public good assigned to them and to their partners i.e., equal, high or low. Third, all subjects interacted anonymously, and the partner's real identity was never revealed. Fourth, everyone had the chance to punish their partners in one of the games and to carryout punishment they were given twelve 5-rupee coins. Fifth, the partners were always chosen from a different village. Sixth, every subject received an amount of INR 100 as a participation fee irrespective of what they earned from the experiment. Seventh, the members could lose all their endowment amount and even their payoff could become negative in some extreme cases. For example, if both the members contribute nothing but punish each other severely (12 coins) then the member's payoff becomes -50 rupees. In such cases, they were told during the instructions that they are not required to make up for the losses, however they get nothing from the game. Eighth, one of the games were picked randomly to pay for the experiment. Ninth, the participation fee was given at the end of the experiment and the payoff from the experiment was paid after 3-4 days of participating in the experiment. The delay was due to the time required for matching members from different villages.

3.3. Experiment procedure

The lab in the field experiment was conducted in 22 villages of Rampur district, Uttar Pradesh with 735 individuals. The selection of villages was based on the presence of both high and low caste members or either of them in the village. In each village we announced our experiment and invited members from the village to voluntarily participate in it with a limit of one person per household. This was done to give equal opportunity for everyone to be able to participate. Once members are gathered, they were given general instructions about the study and recruited to the experiment after they satisfactorily answered the test questions. Before the start of the experiment, the instructions were read aloud by the enumerator in local language (Hindi) and the opportunity to ask questions was provided. An arrangement was made for each subject to sit with an enumerator in separate rooms where the experiment was conducted and thus maintaining the condition of lab where there is an absence of noise and external influence. Upon the arrival of the subject, a lottery was conducted to select the partner from a bowl of chits in which first names (fictitious) and last names of partner members were presented. After the selection of the partner, the type of returns (equal, high or low) to be assigned to the subject was randomly selected by the enumerator. At the same time, the subject was informed about the type of returns his/her matched partner assigned with. All the subjects were

paid their earnings confidentially after 3-4 days and the participation fee on the same day after the experiment. After playing the games, the subjects were asked to respond to a survey consisting of questions on their socio-economic background and other related information. The average earnings from the experiment was INR 280.3 (≈3.7 Euros), in addition to the participation fee of INR 100 (≈1.3 Euros). The standard labor wage in the region was about INR 250 (for 8 hours of work) and therefore, our payments were fair enough to incentivize their participation in the experiment. The experiment (individual sessions) lasted for about 90-120 minutes.

3.4. Sample and village characteristics

The subjects come from 22 villages which are diverse in terms of social composition with respect to the caste and religion. Table A1 in the appendix exhibit the proportion of high caste, low caste(SC's), OBC (Other Backward Castes) and Muslim members living in the sample villages. Among high and low castes, 82% of the villages are characterized by the higher proportion of low caste members. However, the OBC's form the dominant caste group in most of the sample villages. Rampur district is known for the dominance of Muslim religion and it is also reflected by a considerable proportion of Muslims in many villages.

To test the balance among high and low caste individuals with respect to the socio-economic characteristics, t test is conducted, and the result is presented in Table A2 in the appendix. The average age group is in the range of 41-50 years for the high caste subjects and 31-40 years for the low caste members. Men participated in an overwhelmingly larger proportion than women among both high (89.5%) and low castes (84.8%). It may be because woman in villages are restricted by the men only to carryout household chores and to help in agricultural activities. Therefore, their participation in any public activity is very limited. The high caste members are more educated with an average of 9.5 years of schooling than the low caste individuals with 5.2 years of schooling. A similar trend is also observed with respect to wealth. The annual income of high caste members is in the range INR 50000-60000 and that of low caste individuals is at INR 30000-40000. It is also reflected in their land holdings. The average land holding of high caste members is 2.2 acre and that of low caste is 0.76 acre. The distribution of their annual Income is depicted in Figure 2. Although the high caste members are an average richer than the low caste members, there is a considerable amount of overlap in their income distribution. The significant difference is mainly due to the presence of higher proportion of poor among low castes than that of high castes. The Kolmogorov-Smirnov test for the differences in the distribution indicated a significant difference ($p=0.00$) in the distribution of the annual income among high and low caste individuals.

Table A3 presents the test for balance for some important socio-economic indicators across the treatments. Overall the samples look balanced except for age in comparison across marginal returns and land area in comparison across caste treatment in which they showed some modest significant differences.

4. Results

Table 2 presents the summary statistics of the experiment parameters. The average public good (PG) contribution without punishment by pooling the data across all caste matching groups and return types is INR 67.14 which is 27% of the endowment amount. Although free riding could be a better option in a one-shot game due to absence of any credible threat or tit-for-tat strategy by other players which is the case in repeated games, we observe a positive contribution. The average contributions

are less than what is normally found in the previous literature for lab experiments which is 40-50% of the endowment (Chaudhuri, 2011). In the field experiments, the contribution is known to vary from very high to low due to many contextual factors which are absent in the lab (Ostrom, 2000). When the punishment is introduced, the mean contribution increased to INR 107.96 (43% of endowment). The Wilcoxon rank-sum test of the difference between the contributions before and after the punishment showed a significant (p=0.00) change in the contributions. Although the mean contributions jump close to double after punishment, it is still far away from the efficient level or social optimum.

Table 2: Summary statistics of the experiment parameters

Variable	Categories	Obs	Mean	Std. Dev.
Contribution	Without Punishment	735	67.14	41.50
	With Punishment	735	107.96	51.22
Return type (Dummy)	Equal=1 (35.8%) High=2 (33.5%) Low=3 (30.7%)	735	1.95	0.81
Caste matching (Dummy)	Low – Low=1 (32%)			
	High – High=2 (21.1%)			
	Low – High=3 (15.1%)			
	High – Low=4 (15.5%)			
	Low – Ano=5 (10.2%)			
	High – Ano=6 (6.1%)	735	2.69	1.59
Punishment level	C0*	735	5.23	3.23
	C50	735	3.47	2.88
	C100	735	2.24	2.43
	C150	735	1.27	1.97
	C200	735	0.53	1.48
	C250	735	0.19	1.05

Note: *C0 indicates the number of coins punished when the contribution is 0. Similar explanation holds for all the punishment levels viz., C50, C100, C150, C200, and C250. 'Ano' indicates anonymous partner.

To render a complete picture of the effect pathways through which elite groups may perpetuate normative behavior in their interest, we follow the steps of the framework in the presentation of the empirical results. In each of the following six sub-sections, we detail the empirical strategy to test the corresponding hypothesis and present the estimation results. In the main document, we report predicted averages and marginal differences to describe conditional differences between groups and refer to the Appendix for the complete regression output. All our estimates are derived from linear OLS regressions by controlling for a set of socio-economic factors such as education, age, gender, land area, house type, family size etc. and fixed effects for villages, enumerators and the order in which the members faced punishment treatment.

4.1. Caste identity and PG contributions under anonymity (Hypothesis 1)

In the first step, we test if there exist any systematic difference in the PG contribution behavior of high and low caste members when they are anonymously matched (LA and HA). A difference in PG contributions could be indicative of unobserved differences between high and low caste participants that could lead to confounding effects in our analysis.

Therefore, we regress PG contributions C on a binary variable $High\ caste$ indicating whether participant i belonged to a high caste or not. We also control for a set of individual characteristics *Z* which may have an effect on the behavior. The regression model is the following:

$$C_i = \alpha + \beta High\ caste_i + kZ_i + u_i \quad (15)$$

This regression is restricted only to the sample of anonymous treatment. Table 3 shows the conditional difference in PG contributions between high and low caste members. From the results, it is evident that difference (pooled) is not statistically significant, which suggest that the contribution behavior of both high and low caste group is comparable in cases where the identity of the matched partner was unknown. Punishment significantly increased overall contributions but did not have any significantly different effect between both the caste groups. It confirms our first hypotheses of no underlying difference in the contribution behavior of high and low caste members.

Table 3: Margins for contributions under anonymity

Caste matching	Predicted average	Marginal difference
		LA
Pooled		
HA	**74.54**	*14.06*
No punishment		
HA	**62.24**	*3.75*
With punishment		
HA	**86.85**	*24.37*

Note: Refer to Table A4 for the regression output. ***, ** and* indicates the significance at 1%, 5% and 10% respectively. HA and LA indicate High and Low caste members respectively matched with anonymous partners

4.2. Caste identity and contributions of homogeneous caste groups (Hypothesis 2)

In the next step, we test whether differences emerged in treatments where the caste affiliation was implicitly announced and whether there were differences in the effect depending on the own caste and the one of the matched partner. In particular, our focus is on testing the difference in the level of cooperation among the members from the same social rank between high (HH) and low caste (LL). To test if these differences among homogeneous caste groups is robust, we employ the previous estimation strategy only to the sample of homogeneous caste treatments (ignoring anonymous treatments). Therefore, in the estimation model we control whether members of the matched group j are from high caste ($High\ caste = 1$) along with the set of socio-economic variables *Z*.

$$C_i = \alpha + \beta High\ caste_i + kZ_i + u_i \quad (16)$$

The low caste homogeneous (LL) forms the base category which is represented by the coefficient a. The difference in contribution between high and low caste homogeneous groups is given by the coefficient β. We first run the regression on the pooled sample, and thereafter on the sub-sample of treatments with and without punishment. The average predictions and marginal differences are presented in Table 4.

As expected, the estimation results show significantly higher PG contributions among homogeneous high caste groups (HH) compared to homogeneous low caste groups (LL). On average, the model suggests that PG contributions of high caste members is INR 20 higher than that of low caste. Results from the estimations segregated by punishment also shows a similar pattern. The high caste members have contributed higher among homogenous groups irrespective of the presence of norm enforcement, although the presence of punishment has increased the magnitude of difference (INR 24.84). The results indicate lack of both strong norm enforcement ability (social preference) and collective benefit incentive among low castes compared to high caste members.

The results could suggest that creating a common identity by providing the caste information led to significantly higher PG contributions for high caste groups compared to low caste indicating the differences in their abilities to cooperate. For now, we only discuss the results for the case of homogeneous groups and will put our focus on the case of mixed (heterogeneous) caste groups from hypothesis 4 onwards.

Table 4: Margins for contributions in homogeneous caste groups

Caste matching	Predicted average	Marginal difference
		LL
Pooled		
HH	**97.37**	*-20.11****
No punishment		
HH	**75.03**	*-15.37***
With punishment		
HH	**119.71**	*-24.84****

Note: Refer to Table A5 for the regression output. ***, ** and * indicates the significance at 1%, 5% and 10% respectively. HH and LL indicate High and Low caste homogeneous groups respectively.

4.3. Heterogeneous marginal returns and PG contributions (Hypothesis 3)

In the first two steps we analyzed the effect of caste information on experiment behavior. After finding support for our first two hypotheses, we now leave the anonymous treatments aside and consider only observations of the caste treatments. In particular, we analyze how heterogeneity in marginal returns to the PG affect contributions. From a utilitarian perspective, the social optimum would be to fully contribute to the PG regardless of whether marginal returns are high or low as long as it is positive. However, the way in which normative behavior in PG games evolves may also be guided by fairness considerations (Reuben and Riedl, 2013). For example, besides pure efficiency concerns, normative behavior could follow equality concerns (same contributions to the PG regardless of the

marginal returns to the PG) or equity concerns (contributions to the PG are adjusted so that everyone benefits equally).

To get a better understanding if and how changes in marginal returns affect behavior, we compare the contributions under the conditions of equal, high and low marginal returns from the PG. Therefore, we regress contributions on the marginal return regime R (Equal, High, Low) controlling for $Caste$ accounting for the four different caste composition possibilities (LL, HH, LH, HL) and Z controlling for a set of individual characteristics, which can be formulated as follows:

$$C_i = \alpha + \beta R_i + \gamma Caste_i + kZ_i + u_i \quad (17)$$

The model predictions and marginal effects are presented in Table 5. We present the model predictions of PG contributions under equal marginal returns to the PG (numbers in bold) and next to it we show the predicted difference if a member received high or low marginal returns respectively. The estimates are based on the pooled sample of all caste treatments and we additionally present the results for the sub-sample with and without punishment.

We find that the marginal returns have a highly significant effect on contributions in the pooled sample. Compared to the equal returns, obtaining higher returns from the PG resulted in an increase of INR 8 in contributions on average whereas lower returns from the public good reduced contributions by INR 13. It shows that members were strongly responsive to the differences in marginal returns relative to those of others in the group rejecting our hypothesis 3. Since we observe a stronger reduction in contributions under lower marginal returns as compared to an increase in contributions under higher marginal returns, the overall contributions with heterogeneous marginal returns is lower than in the case of homogeneous returns, but the difference is not statistically significant (Kruskal-Wallis test H_o: $C_{Het} = C_{Equal}$; p=0.31).

However, especially in the presence of punishment, members' absolute increase in contributions to high marginal returns is smaller than the decrease after being assigned low marginal returns. Thus, providing higher returns did not increase PG contributions unless there was a punishment threat. Without punishment, members in the high marginal return did not increase their contributions compared to the equal marginal returns regime. Once sanctioning was possible, we observe an increase in the gap between contributions of the participants with high and low marginal returns. These differences could indicate a shift towards more equity concerned normative behavior specially once the members received low marginal returns to the PG.[10]

Looking into the effect of caste composition on contributions from the regression (Table A6), we find that Comparing to homogeneous low caste (LL), the contributions under HH and LH are significantly higher to the tune of INR 11.63 and 11.32 respectively. Whereas, the contributions under HL is also higher but weakly significant. It also shows that presence of high caste members has always increased contributions by themselves or from their partners compared to the group with no high caste members. Segregating by punishment shows that the difference in contributions among these caste

[10] It may also because of pessimistic beliefs about the partner's contribution when given lower marginal returns leading to lower contributions (Fischbacher et al. 2014).

groups are mainly driven by the presence of punishment. Therefore, it is the sanctioning which reinforced the caste differences in contributions.

Table 5: Margins for contributions under different marginal return types

	Predicted average	Marginal difference	
	Equal Returns	High Returns	Low Returns
Pooled	**89.67**	*8.25***	*-13.10****
No Punishment	**68.24**	*6.09*	*-8.50***
With Punishment	**111.10**	*10.42***	*-17.70****

Note: Refer to Table A6 for the regression output. ***, ** and * indicates the significance at 1%, 5% and 10% respectively.

4.4. Punishment behavior and caste (Hypothesis 4)

As we have witnessed from the previous sections, PG contributions differ depending on the PG regime and the caste of the matched partner. The effects are most salient if members were able to sanction their partner. In this sub-sesction, we emphasize the role of punishment and analyze whether there are caste differences in the actual punishment behavior. In addition, we open the discussion up for the cases of mixed caste groups, which we haven't touched upon in the previous sections.

Following the framework, we explore the punishment behavior by group composition to test the caste conflict, caste submission and the caste culture hypothesis. According to the caste conflict hypothesis, both high and low caste members should punish stronger under the heterogeneous caste treatments compared to the homogeneous caste treatments. The caste submission hypothesis states that low caste members punish less in the heterogeneous caste treatment than in the homogeneous caste treatment. Lastly, the caste culture hypothesis states that high caste members punish more irrespective of the partner's caste.

To test these hypotheses empirically we regress the amount of punishment *P* (number of coins) of member i on a binary variable indicating whether i belongs to a $High\ caste$ or not, a binary variable indicating $Heterogeneous$ caste groups, the interaction of both variables, and the same vector of control variables as in the previous regressions. In addition, we also control for the members' contributions to PG. The regression model is as follows:

$$P_i = \alpha + \beta High\ caste_i + \gamma\ Heterogeneous_j + \delta\ (High\ caste * Heterogeonous)_i + kZ_i + u_i \quad (18)$$

To test the caste conflict hypothesis, we need to measure the punishment levels of HH, HL, LL and LH, which are measured by the coefficients of the punishment regression. $\alpha + \beta$ provides the level of punishment in HH. Similarly, $\alpha + \gamma$ indicates the punishment level in LH and $\alpha + \beta + \gamma + \delta$ indicate the punishment level in HL. While the coefficient γ indicate if punishment level at LH>LL, F test $(\gamma + \delta > 0)$ indicates if HL>HH. Table 6 shows the predictions and marginal differences based on the estimation model. The numbers in bold show the predicted punishment for HH and LL and L. The numbers to the right indicate the predicted marginal differences if participants were matched differently respective to the benchmark in column 2.

The model predictions suggest that participants of homogeneous high caste groups invested about 2.32 coins to punish their matched partner. This is 0.44 more compared to participants of

homogeneous low caste groups, but the difference is not statistically significant. However, punishment increased significantly in heterogeneous groups as compared to homogenous caste groups and the effect is similar in size for low and high caste members (the predicted marginal difference is 0.43 and 0.54 respectively). Therefore, our results clearly support the caste conflict hypothesis and reject the caste culture hypothesis. Comparing the punishment behavior of high and low caste participants irrespective of the caste of the matched partner shows that high caste player tended to punish 0.54 on average. The difference is significant on the 10% level additionally providing modest support for the caste submission hypothesis.

Among other control variables, PG contributions have a small but positive effect indicating that participants who contribute more also exercise more punishment (see Table A7 in the Appendix). As expected, we observe that participants punish less the more the partner contributes. We find the individuals who obtain higher returns to the PG to punish 0.36 coins less than the ones in the equal returns regime.

Table 6: Margins for punishment behavior

Caste matching	Predicted average	Marginal difference			
		HL	LL	LH	H
HH	**2.32**	0.54**	-0.44		
LL	**1.88**			0.43**	
L	**2.01**				0.54*

Note: Refer to Table A7 for the regression output. ***, ** and * indicates the significance at 1%, 5% and 10% respectively. HH and LL indicate High and Low caste homogeneous groups respectively. HL and LH indicate High and Low caste heterogeneous groups respectively. H and L indicate High and Low caste members respectively.

4.5. Contributions under heterogeneity (Hypothesis 5)

After we have discussed differences in the punishment behavior, we now turn to the combined effects of heterogeneous caste matching and marginal returns on PG contributions. Thereby we focus on mixed caste groups and do not consider homogeneous caste groups.

Following the framework, we assume that high caste members use the sanctioning mechanisms (and the threat thereof) to enforce contributions that lead to equal returns (payoff) to the PG when they receive low marginal returns and equal PG contributions if they are assigned high marginal returns. In other words, we hypothesize that high caste enforces normative behavior in their interest through differences in their contribution and punishment behavior.

To test this, we compare contributions among high and low caste members between high and low marginal returns in heterogeneous caste treatments. Since we are interested in analyzing if the difference of these differences in contributions among high and low caste individuals under heterogeneous returns are significant in the presence of punishment, we run the following regression:

$$C_i = \alpha + \beta\,(High\ caste) + \gamma\,(High\ returns) + \delta\,(High\ caste * High\ returns) + kZ_i + u_i \quad (19)$$

In the regression, we account for the caste of the member whether they belong to $High\ caste$ or not and for marginal returns if they have assigned $High\ returns$ from PG and the interaction of both the terms. We also control for a set of socio-economic variables Z for each member. Significance of the interaction term (parameter δ) confirms our hypothesis 5.

The model predictions and marginal differences are presented in Table 7. The numbers in bold refer to the predicted contributions of high and low caste members when assigned low marginal returns to the PG. Next to it, we show the marginal differences if high marginal returns to the PG is assigned. In the last column, we present the p-value of an F-test on the difference in the difference between high and low caste members. We first present the results for the pooled sample of heterogeneous treatments and thereafter split the sample by punishment and no-punishment treatments.

The results show that high caste members contributed INR 74 on average when they were assigned low marginal returns, which is about INR 18 less than low caste individual receiving low marginal returns. However, when high caste members were assigned high marginal returns, their contributions increased strongly by INR 33. On the other hand, low caste members did not increase their contributions significantly when assigned high marginal returns. It shows clearly that low caste members were less sensitive to changes in the marginal PG returns. While low caste members with high marginal returns contributed INR 22 more on average than the high caste members with low marginal returns, the gap plunged to only INR 15 in the opposite case where high caste members received high returns. This difference in difference is statistically significant. The sub-sample analysis shows that the effect is driven mainly by punishment treatment although we found a modestly significant difference even in the absence of punishment.

The results support our hypothesis that high caste members influence PG behavior in their interest if sanctioning instruments are available. An alternative interpretation could be that low caste members were less sensitive to changes in marginal returns because of confounding effects such as lacking numerical literacy needed to understand the implications of changes in marginal returns. We control for the number of years of education in all estimations and used control questions and tables to ensure a clear understanding of payouts, but numerical literacy may still influence choices. To ensure that differences are the result of the provision of the caste information, we re-estimate the model for anonymous treatments in which the caste information was not provided, but the de-facto matching was the same. The results are shown in Table A9 in the appendix and shows non-significance of the interaction term indicating the difference in difference of contributions among high and low caste members under heterogeneous returns is driven by the caste of the partner. Therefore, it indicates that, under heterogeneity it's not only the differences in marginal returns that matters but also the contributions depend on who receives what based on their caste identity. This shows that our results are driven by the caste identity of members and their perceptions about the other members' identities.

Table 7: Margins of PG contributions under both caste and returns heterogeneity

Caste-Return matching	Predicted Average	Marginal difference		Difference of marginal differences (p value)
		High caste-High return	Low caste-High return	
Pooled				
High caste-Low return	**74.08**	*33.22****		
Low caste-Low return	**92.22**		*3.50*	***0.01***
No punishment				
High caste-Low return	**56.63**	*28.33****		
Low caste-Low return	**69.88**		*1.64*	*0.09*
With punishment				
High caste-Low return	**91.53**	*38.11****		
Low caste-Low return	**114.56**		*5.35*	***0.05***

Note: Refer to Table A8 for the regression output. ***, ** and * indicate the significance at 1%, 5% and 10% respectively.

4.6. Welfare effects of heterogeneity (Hypothesis 6)

The findings indicate that high caste members influenced group's behavior in mixed (heterogeneous) groups in their interest mainly if sanctioning was possible. In the last step, we test whether this has implications for provision of the public good and the group welfare. Following the framework, we hypothesize that contributions to the PG are highest if high caste participants receive high marginal returns and lowest if low caste participants receive high marginal returns. In an extreme case, low caste participants might even be better off in terms of their earnings (payoff) if they receive low marginal returns as compared to high marginal returns. This could have important welfare implications. The sum of experiment earnings would be highest if high caste members are assigned high marginal returns, however, at the same time the distribution of earnings would be more unequal leaning towards high caste members compared to the scenario with high marginal returns for low caste members. That is, the increase in total earnings (social welfare) comes with increased inequality of earnings.

To test this, we compare the welfare of high and low caste members and the amount of public good generated under heterogeneous returns and heterogeneous caste scenario. Thereby, we define welfare as the sum of individual payoffs in a group and public good as the sum of individual payoffs without including the cost of punishment both to the punisher and the punished and the investment in the private account. As we are considering outcomes at the group level, we do not condition the comparison on individual characteristics and instead look at simple mean differences using Kruskal-Wallis test. Table 8 compares the group welfare of high and low caste members and the amount of public good generated under heterogeneous returns and heterogeneous caste scenario.

The mean comparison test shows that the returns to the investments in the public good are INR 45 higher in cases where the high caste member received high marginal returns and not the low caste group member. This results in a significant difference in group welfare to the tune of INR 32 when the high caste gets higher returns compared to when low caste gets higher returns. Splitting the sample shows that the difference is driven by treatments with punishment (welfare gap INR 53) and plunges

to a difference of INR 10 in treatments without punishment. These results support our hypothesis 6 suggesting welfare implications of the allocation of marginal PG returns in heterogeneous caste groups.

While from the utilitarian welfare perspective, the high caste receiving higher marginal returns prove to be a better option than otherwise, it is important to know if the benefits received (payoff) from the public good are significantly different for both high and low caste. Are low caste members better off receiving lower marginal returns from the public good?

We answer this question using a linear regression as specified in equation 19 with payoff as the dependent variable. The model predictions and marginal differences are presented in Table 9. The results show that the high caste members always appropriate significantly higher payoffs in the range of INR 32 to INR 63, when they are given higher marginal returns both in the presence and absence of punishment. On the other hand, although the low caste members receive higher payoff when they are given higher marginal returns, the difference is not significant and hence they remain indifferent in the way heterogeneous marginal returns affect their payoff.

Table 8: Welfare comparison under caste and returns heterogeneity

	Caste matching				**Kruskal-Wallis test (p values)**
	H*L	**n**	**HL***	**n**	
Public good					
Pooled	297.11 (128.17)	78	251.64 (114.04)	76	**0.01**
No Punishment	226.92 (102.50)	39	195.39 (77.08)	38	0.17
With Punishment	367.31 (112.56)	39	307.89 (117.99)	38	**0.01**
Welfare					
Pooled	542.95 (77.01)	78	511.05 (88.87)	76	**0.03**
No Punishment	575.13 (60.65)	39	565.13 (38.53)	38	0.54
With Punishment	510.77 (78.83)	39	456.97 (92.30)	38	**0.00**

Note: H*L indicates that the high caste gets higher marginal returns and the low caste gets lower marginal returns. Similarly, HL* indicates that the low caste gets higher marginal returns and the high caste gets lower marginal returns. Standard deviations in the parenthesis.

This shows that the higher welfare levels achieved when high caste receives higher marginal returns is mainly reflected in the increase in payoffs of high caste members. Answering the above question, we do not find that low caste members are better off in terms of their experiment payoffs if assigned low instead of high marginal returns to the PG. However, in contrast to high caste members the marginal returns modalities didn't make a significant difference for their payoffs.

Table 9: Margins of Payoff under caste and return heterogeneity

Caste-Return matching	Predicted margins	Marginal difference	
		High caste-High return	Low caste-High return
Pooled			
High caste-Low return	246.66	48.00***	
Low caste-Low return	250.87		14.27
No punishment			
High caste-Low return	278.25	32.47***	
Low caste-Low return	271.84		10.83
With punishment			
High caste-Low return	215.07	63.53***	
Low caste-Low return	229.91		17.72

Note: Refer to Table A10 for the regression output. ***, ** and * indicate the significance at 1%, 5% and 10% respectively

Since we observe differences in payoffs obtained under heterogeneous marginal returns, we also attempt to test if this leads to inequality in earnings from public good. We test this by comparing the significance of differences in returns from public good under heterogeneous marginal returns for both high and low castes using Kruskal-Wallis test. From Table 10, it is evident that when high caste members are assigned with higher marginal returns from PG, they achieve higher benefits from it compared to low caste members to the tune of INR 9. Comparing these differences by splitting sample by punishment shows that these differences are driven by the presence of sanctioning. Therefore, we assert that high caste members getting higher marginal returns from public goods increases both total welfare and inequality at the same time.

Table 10: Inequality under caste and return heterogeneity

	Caste matching				Kruskal-Wallis test (p values)
	H*L	n	HL*	n	
Public good difference					
Pooled	59.42 (25.63)	78	50.33 (22.81)	76	**0.01**
No Punishment	45.38 (20.50)	39	39.08 (15.41)	38	0.17
With Punishment	73.46 (22.51)	39	61.58 (23.60)	38	**0.01**

Note: H*L indicates that the high caste gets higher marginal returns and the low caste gets lower marginal returns. Similarly, HL* indicates that the low caste gets higher marginal returns and the high caste gets lower marginal returns. Standard deviations in the parenthesis.

5. Mechanism

Above results indicate lack of collective ability among low caste members compared to high caste. It could be due to lack of self-esteem among low caste members. Hoff et al (2011) showed that low caste members fail to enforce norms and argues that its due to their lack of self-confidence. Even though our results show low caste members strongly punishing high caste norm violators, it did not reflect in their contribution behavior among fellow low caste members. In order to test if self-depreciation and negative stereotype is the cause of their behavior, we employed a priming strategy (Dijksterhuis et al, 2007; Shariff and Norenzayan, 2007) where the members after playing the first two games are briefed with the success stories of a few role models from their caste group. After briefing, all the members played both the games again in the same order that was followed previously, with their partners and type of marginal returns unchanged. Therefore, each member played four games in total (2 games before and 2 games after the role model priming). We hypothesize that introducing members to the stories of their role models (Pleiss and Feldhusen, 2010; Hurd et al., 2011) breaks the chain of self-depreciation, instill positivity about their identity and boost self-confidence which will positively influence both individual and collective wellbeing. This effect is expected to be higher among low caste members as they seem to be suffering from the lack of it compared to high castes.

To examine the effect of priming, we performed analysis for the data after role model priming, similar to what we have presented above. Here we present only some of the important results and the other results can be shared upon request. In this part, we restrict our explanations to only compare the difference between high and low caste members cooperation behavior before and after the role model priming as our primary interest is to see if this has any effect on the behavior of low caste vis-à-vis high caste members.

5.1. Contribution behavior

We compare the difference in contributions among homogeneous high and low castes after role model priming. We find that the low caste contributing INR 10.4 less than the high caste with modest significance and is driven by punishment (Table A12). That shows a moderately strong response of high caste members compared to low caste for punishment even after priming. This could be due to the general nature of high caste who are known for their sanctioning behavior. Whereas, the low caste have showed some improvement after priming through their better adherence to the cooperation norms. Comparing it with our previous results shows that there is a considerable improvement in the contribution (difference reduced by half and lower significance) of low caste members after priming. We find education to have a significant effect in increasing contributions (Table A11). It shows that members with higher number of years of education have significantly and positively responded to the priming.

Further, we investigate if this improved cooperation among low castes in homogeneous groups is also reflected in their contributions under caste and returns heterogeneity. Table A14 shows a significant improvement in contribution behavior of low caste members. That is, the difference of differences in contributions among high and low caste when given high and low marginal returns is modestly significant although the difference in contributions of low castes between high and low returns is not significant.

5.2. Punishment behavior

As important as the contributions, the norm enforcement behavior also is an indicator of the nature of cooperation among the members. Table A16 shows the prevalence of caste conflict behavior with an increased intensity of punishment. Both low and high caste have increased their punishment on each other by 0.33 coins. It shows that boosting morale using the role models from their castes not only increases cooperation among them but also increase the punishment against the other caste. It might be due to the strong feeling of belonging to their group and showing hostility toward the other groups. We also find a strong evidence for caste culture hypothesis (against to the weak evidence before role model priming) indicating the high castes showing strong punishment behavior overall compared to low castes.

5.3. Welfare effects after priming

Table 11 presents the welfare comparison among high and low caste members under heterogeneity. After priming, we find no significant difference in the welfare among high and low castes receiving different marginal returns. On the other hand, the return from the public good is still higher (INR 31) when high castes are given higher marginal returns with a modest significance. Therefore, it shows that irrespective of who gets high or low returns, the overall welfare remains unchanged. Comparing the individual payoffs (Table A18) shows both high and low caste members managed to obtain significantly higher payoffs of INR 54.59 and INR 35.30 respectively when given higher marginal returns compared to the case of lower marginal returns. Comparing it to the results before the priming, it indicates a significant improvement in the low caste's payoff.

Table 11: Welfare comparison under caste and returns heterogeneity after priming

	Caste matching				Kruskal-Wallis test (p values)
	H*L	n	HL*	n	
Public good					
Pooled	379.81 (108.34)	78	348.35 (121.68)	76	0.09
No Punishment	326.92 (81.59)	39	292.10 (98.31)	38	0.09
With Punishment	432.69 (106.71)	39	404.60 (117.67)	38	0.28
Welfare					
Pooled	578.72 (45.19)	78	563.03 (55.00)	76	0.23
No Punishment	608.72 (43.55)	39	598.95 (59.73)	38	0.33
With Punishment	548.72 (65.52)	39	527.11 (76.33)	38	0.39

Note: H*L indicates that the high caste gets higher marginal returns and the low caste gets lower marginal returns. Similarly, HL* indicates that the low caste gets higher marginal returns and the high caste gets lower marginal returns. Standard deviations in the parenthesis.

Table A19 compares the difference in inequality from the public good returns when high and low caste gets unequal marginal returns after priming. It shows an improvement (compared to the results from

before priming) in the level of inequality with a modest significant difference in the level of inequality (INR 6.3) when high caste gets higher marginal returns compared to low caste getting higher marginal returns. In the presence of punishment, the difference is not significant indicating the importance of sanctioning institutions along with priming to ensure equitable distribution of benefits from the public good.

6. Discussion

In this section we discuss the results by following our design of framework with subsections for each hypothesis.

6.1. Anonymity

No difference in contributions when the identity is not revealed is an indicator of inherently similar pro-social behavior among both high and low castes irrespective of their position in the social hierarchy. Bohnet and Frey (1999) argues that, any individual display fairness considerations with only intrinsic motivations when the partner's identity is anonymous. That is, in the absence of any external influence, we assume that the intrinsic motivations are similar for the members irrespective of their caste identity.

6.2. Homogeneous caste and contributions

Our comparison of homogeneous caste groups is to understand and test if there exist difference in collective ability among high and low castes. Revelation of social identity increased contributions only for high caste members. Identity revelation may reduce the uncertainty attached to the identity of the person or provide a focal point based on the identity of the person. This may also provide social approval (Rege and Telle, 2004) for the members who then may be influenced and contribute. In our case, it worked for the high caste members as they had strong social approval incentives to contribute under punishment whereas it seemed absent for the low caste who might have expected lower level of adherence to the social norm from their peers. Revelation of identity activates any members' identity related preferences for their partners based on their previous life experiences. In villages, individuals commonly interact with people from their own specific caste be it with kin or any other relatives who follows common values and customs. It also forms an important part of their social network for sharing information, lending loans etc. (Maertens and Barrett 2012). Also, the social norm of marrying within the specific caste has been followed by the majority of population leading to a better understanding of the members from their specific castes. On the other hand, there has been a segregation of population based on the hierarchy of castes within the villages[11]. This develops an affinity not only to their own specific castes but also towards their own caste groups or caste ranks, viz., high caste or low caste.

[11] For instance, in most of the Indian villages, the low caste individuals live together in the designated localities. Also, there are differential rights that are possessed by the individuals from different caste hierarchies. High caste individuals mainly involve in the activities which are seen superior and project power such as land lords, priests and decision makers in any village issues. Whereas, the low caste individuals mostly take up the inferior or low skilled tasks such as agriculture labourers and other menial jobs. During many village festivals, the high and low caste individuals always form separate groups to celebrate and worship.

The high caste group's better cooperation may explain how these groups managed to dominate and discriminate against low caste over centuries. A very strong threat of punishment among high caste is an important factor that explains a strong compliance to norms among them. One of the explanations why the high caste members comply to norms under punishment is that they aim to maintain the supremacy over others with a superior social identity. A study by Mukherjee et al (2016) shows that the households with high status collectively come together for any activity which are status driven and status maintaining. Therefore, the public good contributions also can be considered as a kind of charitable and status driven activity for the high caste individuals to retain their social power in a multiethnic society. This may induce high caste members to comply strongly in the presence of punishment as they expect a strong retaliation from the other high caste members to showcase the collective superiority due to strong social identity. In addition, their strong collective ability even in the absence of punishment is an indicator of the incentives of their identity compared to that of low castes who carry a negative stereotype. Dasgupta et al (2016) showed that there is a difference in the level of aspirations among high and low caste members. That is, they showed that the low caste members possess less behavioral preferences and personality traits which positively affect their overall wellbeing when compared to high caste members. Therefore, it may be the lack of these elements leading to a weak group cohesion for the collective benefit among low caste individuals. This point us towards the lack of self-image among low caste members. It may be due to the burden of deprivation, discrimination and oppression from high caste leading to a stigmatic identity with lack of self-respect (Bros, 2014).

6.3. Heterogeneous marginal returns and contributions

Our results are supported by Fisher et al (1995) and Dawes et al (2007). Although the dominant strategy even in case of high returns is to free ride as compared to the privileged group types (Reuben and Riedl, 2009) where it is optimum to contribute positive amount, we still witness a positive contribution by the members. Similarly, in case of low marginal returns although an individual may view it as an unfair distribution of benefits from the public good, he contributed positively but less to the public good. The differences in contribution behaviour at high and low marginal returns exhibits presence of self-serving bias (Babcock and Loewenstein, 1997) among high and low caste members. That is, under high returns, the members appear to be concerned about equality of contributions and under lower returns their concerns are towards equality of payoffs (inequity aversion). Therefore, their contributions significantly dropped under lower returns compared to increase in contributions under higher returns only in the presence of punishment.

The differences in contributions across different caste matching groups reflect some important characteristic features of caste heterogeneity. The presence of high caste has resulted in an increase in contributions in both homogeneous and heterogeneous groups. It signals at the strong threat of punishment to the group members when there is a high caste member in the group. Whereas it also shows a lack of norm enforcement threat among low caste members in homogeneous caste groups. This shows how the punishment mechanism works differently based on the identity of the punisher and the punished. In case of homogeneous group of low castes, the lower contributions reflect the ineffectiveness of the norm enforcement institutions among them.

6.4. Caste and Punishment behaviour

Observing differences in contributions among the members of different caste groups reflected by the presence of punishment, we test if there exist a difference in their actual punishment behavior. We

compare our results with Hoff et al (2011) who provided evidence for the caste culture hypothesis and conducted their experiments in the same region over a decade ago. Our evidence on caste conflict hypothesis indicates the raise of low caste members in their power to fight against norm violations by high caste members. This may be the result of opportunities (eg: affirmative policies) given to them over time to express their concerns and to come to the mainstream leading to improved economic and social wellbeing (Kapur et al., 2010). It also signifies an improved confidence among low caste members to retaliate high castes violations. The Bhima Koregaon incident in India is one of the very recent and significant assertions of low caste against high caste's discrimination, where more than 300,000 Dalits gathered which threatened high caste leading to violence (Thakur and Moharana, 2018). Jaoul (2017) points to the increasing political mobilization of Dalits in Uttar Pradesh state by Non-governmental organizations (NGO'S) using western funds.

In addition, our evidence (modest) on caste culture hypothesis is an indication of persistence of higher levels of punishment levels by high caste members. This can be linked to various recent caste-based discriminations by high caste members over low castes in India. As reported in a newspaper, a Dalit boy was recently murdered by high caste men in Gujarat, India for owning and riding horse because it was seen as a symbol of royalty and therefore the low caste men are not entitled to have one (Kateshiya, 2018). In another similar incident, the Dalit groom was forced off the horse by high caste men during the marriage procession[12]. Our results show that these caste differences in punishment is mainly driven by caste specific behavior (hostility) towards the other caste groups. At the same time, we see no significant difference in the punishment between homogeneous high and low castes (HH & LL), although high castes' level of punishment is higher. Linking the punishment with contribution behavior for the homogeneous caste groups, it is evident that the high caste members perceive strong threat of punishment from other high caste members where as their actual punishment behavior is not different from that of low caste in the homogeneous caste groups.

6.5. Contributions under heterogeneity

Presence of caste conflict behaviour makes it important to learn how it unfolds specifically under heterogeneous marginal returns. That is how both caste and returns heterogeneities interplay in creating a public good. The contributions under caste and returns heterogeneity indicate the differential behaviour of both high and low caste members depending on the type of returns given to them and their partners. High caste members' higher contributions when given higher marginal returns compared to their lower contributions under low marginal returns indicate their sensitivities to the difference in marginal returns from the public good. It shows that high caste members react positively when they face advantageous inequality and on the other hand they retaliate strongly with lower contributions under disadvantageous inequality. Low caste's indifferent contribution behaviour under both type of marginal returns reflects a condition of submissiveness. That is, when high caste members receive higher marginal returns (therefore low caste with lower marginal returns), they use their strong social sanctioning behaviour to enforce higher contributions from low castes so as to increase their returns from the public good. On the contrary, under low marginal returns high caste's lower contributions and unchanged contributions of low caste (given high marginal returns) indicate a tacit understanding among these castes. It can be explained through the concept of culture of honour (Brooks et al., 2016). It may well be considered that high caste considers getting lower

[12] https://timesofindia.indiatimes.com/city/indore/madhya-pradesh-dalit-groom-beaten-forced-off-horse-during-marriage-procession/articleshow/63789554.cms

marginal return from public good than that of low caste as a challenge to their honour in the society as they have occupied the top of social (caste) hierarchy and always expects the highest respect. This results in a strong and negative emotion which leads to lower cooperation levels. Given higher marginal returns, it does not only influence the high caste positively as they may consider it as a token of respect and thus increase contributions, they also subject the low caste members to their punishment pressure and make them contribute more so that the high caste can gain more from the public good and their contributions will not be wasted.

Since we also observe similar behaviour in the absence of punishment (with modest significance), it indicates that the low castes might contribute better when they are mixed with high castes compared to their contributions under homogeneous caste group (LL). It could be due to their lack of incentives in contributing towards their own caste groups due to stigma and expecting similar behaviour from their peers leading to lower collective ability. On the other hand, they may expect higher contributions from high castes when they are given higher returns and therefore they contribute more in order to benefit from the public good (although lesser than high castes but better than that of contributing to their homogeneous group).

6.6. Welfare under heterogeneity

Higher contributions under high marginal returns for the high caste members resulted in higher welfare levels compared to the case of higher marginal returns for low caste members. The increase in welfare is also coupled with higher payoff for high caste compared to low caste whose payoff remain unchanged under higher returns. This reflects on how even in the absence of any formal power to influence the distribution of benefits (as the case of typical elite capture in political economy), the high caste members still manage to get higher benefits from the public good. It shows that high caste members use their social power and elite behaviour to influence (manipulate) the behaviour of low caste members to conform to the norms which is beneficial to them. That is, the low caste to contribute more (or to not contribute less) even when an unfavourable (low) marginal return is given to them. We call this as an elite capture effect where higher benefits from the public resources are abstracted by the high caste members by manipulating the low caste's behaviour using their social power and creating a mental dilemma as low castes' benefits seem to be unchanged in any scenario. Along with it, the lower collective ability among low castes also makes them easily subjected to this effect. Although it improves overall welfare from the public good, it also creates higher inequality in terms of distribution of benefits.

6.7. Priming

Our results after priming reinforces the effect of caste identity on cooperation. Specifically, low caste members with the burden of negative self-image have shown considerable changes after being motivated with the success stories of members from their own caste rank. The reduced significance of the difference in cooperation between high and low caste homogeneous groups indicate an increase in social preference towards their group for the low castes. It could be due to the development of positive self-image and confidence by priming. On the other hand, the increase in punishment levels in the heterogeneous groups compared to homogeneous groups shows an augmented conflicting behavior across groups with an improvement in cooperation within caste rank groups. Comparing our results with other studies where reminding Roma (Vecci and Zelinský, 2016) and low caste (Hoff and Pandey, 2014) students with their identity reduced their performance shows how deeply the social framing of certain identities with negative stereotypes can result in lack of

confidence leading to subdued development. Nevertheless, reminding low caste members with role models who belong to the same social group improved their cooperation and similar results are found by Vecci and Zelinský (2016) for Roma students' academic performance. The low caste's behavior in the heterogeneous group also shows an improvement as their difference in contributions compared to high caste between high and low returns is less significant. With this, low castes are able to get significantly higher payoffs under higher returns as similar to the high castes leading to similar welfare levels irrespective of who received high returns. Therefore, creating positive self-image resulted in ceasing elite capture effect.

What we observed before priming shows that the low castes are negatively influenced by the context in which they make their decisions which is supported by the theories of context dependence (Tversky and Kahneman, 1983). That is, the members of the low caste usually live and socially interact with high caste members in their real life and that is very similar to our design. On the other hand, the caste identities are very salient in their daily societal transactions whereas it was less salient in our case. Therefore, the effect of role models on the behavior and welfare of low caste members may be is an underestimate. Since priming is a within subject treatment, we expect some learning effects in the decisions after priming. As both high and low caste members are primed, and we are interested only in the differences in behavior between them after priming, the learning effect may matter less as it gets neutralized when we make comparison among them.

7. Conclusions

Collective action for public goods is determined by the differences in marginal returns, differences in social identity and the presence of sanctioning institutions. The differences in marginal returns (unequal returns) imparts differential effect on social welfare when the social identity has a significant influence on the behavior of individuals. This depends on both who gets what proportion (high or low) of benefits (marginal returns) from public good and the differences in returns from the public good. It is crucial aspect of field experiments as the subjects bring their real-life experiences to make decisions in the game. The different interpretations of unequal benefits by individuals are may be influenced by the nature of the society that they live in (cooperative or rivalry) and the social interactions.

High caste members have shown to successfully infuse certain behavior of obedience on low caste through their persistent dominance over many decades. This helps them to abstract higher benefits from the public resources or common goods by making low caste members to conform to the norms which benefits high caste. It is mainly due to the low self-esteem (due to stigma) that the low caste possesses and the high caste's stronger sanctioning behavior. The stigma attached with their identity not only affect them when they are involved with high caste but also when they cooperate among themselves. The punishment behavior of low caste reflects an interesting situation where they seem to have gathered courage to punish high caste norm violators and hostility may have played a bigger part on it. On the other hand, showing lower level of cooperation among their own caste group indicate a lack of social approval among them. The positive image building through role model priming improved low caste's abilities to cooperate mainly through breaking the negative stereotype associated with their identity. This also enhanced their welfare vis-à-vis high caste. Since the annihilation of caste system is impossible in a near future, our study shows how the identities can be constructively used to improve wellbeing of the oppressed, although the longevity of the effect of these exercises remain a concern.

Affirmative policies which have mainly focused on the improvement of economic wellbeing may have fallen short in eradicating the negative effects of identity on the oppressed groups. Therefore, the policymakers need to consider the role of social identities (social power differences) while making decisions and bringing the projects which yield unequal benefits in the heterogeneous societies. An improvement of capacity building is necessary among low caste members to successfully address the negative effects of their identities. This can influence the collective ability and the behavior of individuals to increase participation in public projects and thus effectiveness of self-governing institutions. It reduces inequality from public good benefits in addition to improving welfare of the society.

References

Akerlof, G.A. and Kranton, R.E., 2000. Economics and identity. *The Quarterly Journal of Economics*, *115*(3), pp.715-753.

Alatas, V., Banerjee, A., Hanna, R., Olken, B.A., Purnamasari, R. and Wai-Poi, M., 2013. *Does elite capture matter? Local elites and targeted welfare programs in Indonesia* (No. w18798). National Bureau of Economic Research.

Babcock, L. and Loewenstein, G., 1997. Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic perspectives*, *11*(1), pp.109-126.

Benjamin, D.J., Choi, J.J. and Strickland, A.J., 2010. Social identity and preferences. *American Economic Review*, *100*(4), pp.1913-28.

Bicchieri, C., 2008. The fragility of fairness: an experimental investigation on the conditional status of pro-social norms. *Philosophical Issues*, *18*(1), pp.229-248.

Bohnet, I. and Frey, B.S., 1999. Social distance and other-regarding behavior in dictator games: Comment. *The American Economic Review*, *89*(1), pp.335-339.

Bolton, G.E. and Ockenfels, A., 2000. ERC: A theory of equity, reciprocity, and competition. *American economic review*, *90*(1), pp.166-193.

Brooks, B., Hoff, K. and Pandey, P., 2016. *Can the culture of honor lead to inefficient conventions? experimental evidence from india*. The World Bank.

Bros, C., 2014. The Burden of caste on social identity in India. *The Journal of Development Studies*, *50*(10), pp.1411-1429.

Chakravarty, S., Fonseca, M.A., Ghosh, S. and Marjit, S., 2016. Religious fragmentation, social identity and cooperation: Evidence from an artefactual field experiment in India. *European Economic Review*, *90*, pp.265-279.

Charness, G., Rigotti, L. and Rustichini, A., 2007. Individual behavior and group membership. *American Economic Review*, *97*(4), pp.1340-1352.

Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, *14*(1), pp.47-83.

Cinyabuguma, M., Page, T. and Putterman, L., 2005. Cooperation under the threat of expulsion in a public goods experiment. *Journal of public Economics*, *89*(8), pp.1421-1435.

Coffey, D., Hathi, P., Khurana, N. and Thorat, A., 2018. Explicit Prejudice. *Economic & Political Weekly*, *53*(1), p.47.

Dasgupta, A. and Beard, V.A., 2007. Community driven development, collective action and elite capture in Indonesia. *Development and change*, *38*(2), pp.229-249.

Dasgupta, U., Mani, S., Sharma, S. and Singhal, S., 2016. *Caste differences in behaviour and personality: Evidence from India* (No. 2016/60). WIDER Working Paper.

Dawes, C.T., Fowler, J.H., Johnson, T., McElreath, R. and Smirnov, O., 2007. Egalitarian motives in humans. *nature*, *446*(7137), p.794.

Deshpande, M. & Kerbo, H., 2010. History of the Indian caste system and its impact on India today (Unpublished senior project). San Luis Obispo: California Polytechnic State University.

Deshpande, A., 2008. Quest for equality: affirmative action in India. *Indian Journal of Industrial Relations*, pp.154-163.

Dijksterhuis, A., Chartrand, T. L., and Aarts, H., 2007. Effects of priming and perception on social behavior and goal pursuit. *In J. A. Bargh (Ed.), Social psychology and the unconscious: The automaticity of higher mental processes New York, NY: Psychology Press*, pp. 51–132.

Dirks, N.B., 2011. *Castes of mind: Colonialism and the making of modern India*. Princeton University Press.

Dumont, L., 1980. *Homo hierarchicus: The caste system and its implications*. University of Chicago Press.

Fehr, E. and Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), pp.817-868.

Fehr, E. and Fischbacher, U., 2004. Third-party punishment and social norms. *Evolution and human behavior*, *25*(2), pp.63-87.

Fehr, E., Hoff, K. and Kshetramade, M., 2008. Spite and development. *The American Economic Review*, *98*(2), pp.494-499.

Fisher, J., Isaac, R.M., Schatzberg, J.W. and Walker, J.M., 1995. Heterogenous demand for public goods: Behavior in the voluntary contributions mechanism. *Public Choice*, *85*(3), pp.249-266.

Fritzen, S.A., 2007. Can the design of community-driven development reduce the risk of elite capture? Evidence from Indonesia. *World Development*, *35*(8), pp.1359-1375.

Habyarimana, J., Humphreys, M., Posner, D.N. and Weinstein, J.M., 2007. Why does ethnic diversity undermine public goods provision?. *American Political Science Review*, *101*(4), pp.709-725.

Hoff, K. and Kshetramade, M., E. Fehr, 2011, Caste and Punishment: the legacy of caste culture in norm enforcement. *Economic Journal*, *121*, pp.449-475.

Hoff, K. and Pandey, P., 2014. Making up people—The effect of identity on performance in a modernizing society. *Journal of Development Economics*, *106*, pp.118-131.

Hurd, N.M., Zimmerman, M.A. and Reischl, T.M., 2011. Role model behavior and youth violence: A study of positive and negative effects. *The Journal of Early Adolescence*, *31*(2), pp.323-354.

Jaoul, N., 2017. Politics Against the Grain: The Dalit Movement of Uttar Pradesh in the Throes of NGOization. *Critical Sociology*, p.0896920516688756.

Kateshiya, G.B., 2018. 'For riding horse, upper caste men kill Dalit youth in Gujarat'. *The Indian Express*, 31 March, viewed 6 December 2018 https://indianexpress.com/article/india/for-riding-horse-upper-caste-men-kill-dalit-youth-in-gujarat-5117872/

Kapur, D., Prasad, C.B., Pritchett, L. and Babu, D.S., 2010. Rethinking inequality: Dalits in Uttar Pradesh in the market reform era. *Economic and Political Weekly*, pp.39-49.

Konow, J., 2003. Which is the fairest one of all? A positive analysis of justice theories. *Journal of economic literature*, *41*(4), pp.1188-1239.

Konow, J., Saijo, T. and Akai, K., 2009. Morals and mores: Experimental evidence on equity and equality. *Levine's Working Paper Archive 122247000000002055, David K. Levine*.

Köszegi, B., 2006. Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, *4*(4), pp.673-707.

Ledyard, J., 1994. *Public Goods: A Survey of Experimental Research*. EconWPA.

Leibbrandt, A. and López-Pérez, R., 2012. An exploration of third and second party punishment in ten simple games. *Journal of Economic Behavior & Organization*, *84*(3), pp.753-766.

Lund, J.F. and Saito-Jensen, M., 2013. Revisiting the issue of elite capture of participatory initiatives. *World development*, *46*, pp.104-112.

Maertens, A. and Barrett, C.B., 2012. Measuring social networks' effects on agricultural technology adoption. *American Journal of Agricultural Economics*, *95*(2), pp.353-359.

Martinez-Bravo, M., Mukherjee, P. and Stegmann, A., 2017. The Non-Democratic Roots of Elite Capture: Evidence From Soeharto Mayors in Indonesia. *Econometrica*, *85*(6), pp.1991-2010.

Miguel, E. and Gugerty, M.K., 2005. Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of public Economics*, *89*(11-12), pp.2325-2368.

Mondal, A., 2014. Dalit Reservation and the Issue of Social Upliftment: An Overview. *International Journal of Scientific and Research Publications*, p.635.

Mukherjee, P., Ray, B. and Bhattacharya, R.N., 2017. Status differences in collective action and forest benefits: evidence from joint forest management in India. *Environment, Development and Sustainability*, *19*(5), pp.1831-1854.

Munshi, K. and Rosenzweig, M., 2015. *Insiders and outsiders: local ethnic politics and public goods provision* (No. w21720). National Bureau of Economic Research.

Musgrave, M.K. and Wong, S., 2016. Towards a More Nuanced Theory of Elite Capture in Development Projects. The Importance of Context and Theories of Power. *Journal of Sustainable Development*, *9*(3), p.87.

Ostrom, E., 2000. Collective action and the evolution of social norms. *Journal of economic perspectives*, *14*(3), pp.137-158.

Ostrom, E., 2007. A diagnostic approach for going beyond panaceas. *Proceedings of the national Academy of sciences*, *104*(39), pp.15181-15187.

Ostrom, E., 2009. A general framework for analyzing sustainability of social-ecological systems. *Science*, *325*(5939), pp.419-422.

Platteau, J.P. and Abraham, A., 2002. Participatory development in the presence of endogenous community imperfections. *Journal of Development Studies*, *39*(2), pp.104-136.

Pleiss, M.K. and Feldhusen, J.F., 1995. Mentors, role models, and heroes in the lives of gifted children. *Educational Psychologist*, *30*(3), pp.159-169.

Rao, V and Walton, M., 2004. Culture and Public Action: Relationality, Equality of Agency, and Development, in Rao and Walton, eds.*, Culture and public action.* Stanford, CA: Stanford University Press, pp. 3-36.

Rege, M. and Telle, K., 2004. The impact of social approval and framing on cooperation in public good situations. *Journal of public Economics*, *88*(7), pp.1625-1644.

Reuben, E. and Riedl, A., 2009. Public goods provision and sanctioning in privileged groups. *Journal of Conflict Resolution*, *53*(1), pp.72-93.

Reuben, E. and Riedl, A., 2013. Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, *77*(1), pp.122-137.

Rodrik, D., Subramanian, A. and Trebbi, F., 2004. Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of economic growth*, *9*(2), pp.131-165.

Sen, A., 2000. *Social exclusion: Concept, application, and scrutiny*. Asian Development Bank.

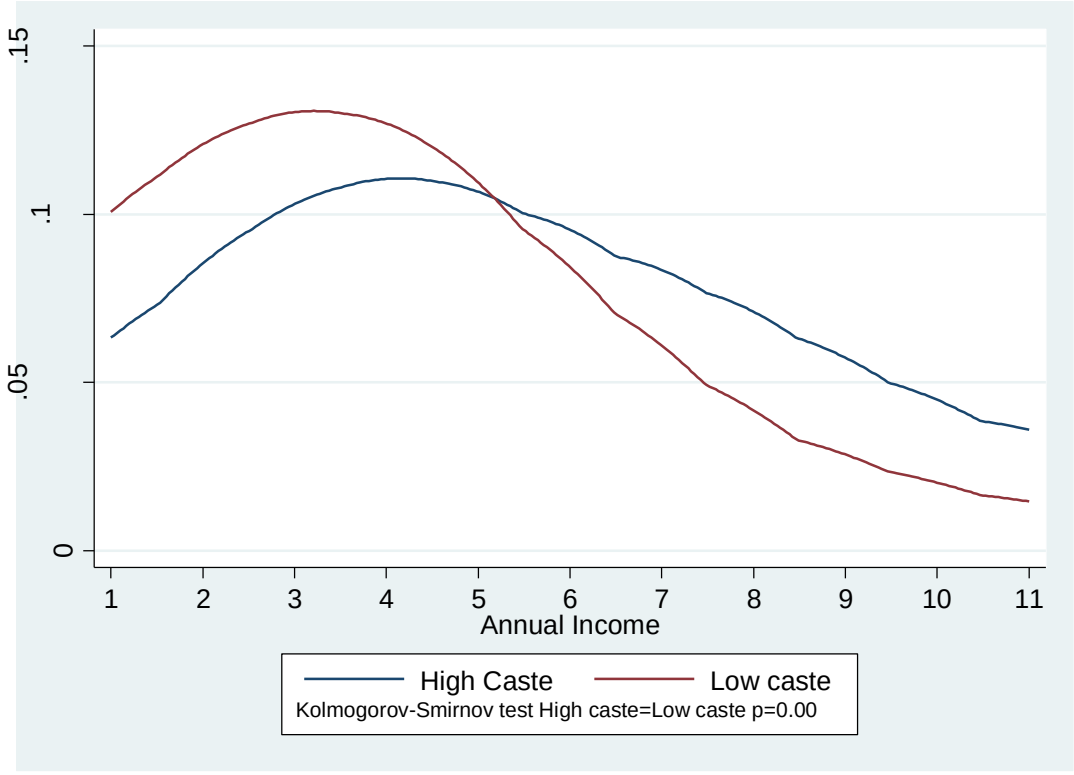
Sidanius, J. and Pratto, F., 2001. *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

Thakur, S. and Moharana, B., 2018. Bhima Koregaon and Politics of the Subaltern. *Economic & Political Weekly*, *53*(7), p.13.

Tversky, A. and Kahneman, D., 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, *90*(4), p.293.

Vecci, Joseph & Zelinsky, Tomas, 2016. "Social Identity and Role Models," Working Papers in Economics 672, University of Gothenburg, Department of Economics.

Figure 2: Income comparison between high and low castes



Appendix

Table A1: Social composition of sample villages

Village	Block	Scheduled/ Low caste (SC) (percent)	General/ High caste (percent)	Other Backward Castes (OBC) (percent)	Muslims (percent)	Total
Bhonakpur	Milak	33.3	22.2	44.5	0	1800
Emi	Milak	25.1	6.4	50.2	18.3	1095
Janu Nagar	Milak	17.8	5.5	76.7	0	1825
Jiwai Jadid	Milak	12.8	69.8	17.4	0	1360
Jyora	Milak	21.2	12.1	18.2	48.5	1650
Karimpur Sarki	Milak	28.1	22.5	44.9	4.5	890
Ladpur	Milak	18.5	5.3	59.6	16.6	1510
Nangla Udai	Milak	16.7	5	47.5	30.8	1790
Narkera	Milak	13.4	1.1	74.9	10.6	1870
Pipal Gao	Milak	16.7	6.7	76.6	0	1045
Rahsaina	MIlak	12.2	8	37.2	42.6	1880
Rajpura	Milak	20.8	9.7	48.6	20.9	720
Ratonda	Milak	38.7	3.1	50.4	7.8	2580
Sihari	Milak	20	0.8	79.2	0	2500
Shadi Nagar	Milak	0	50	50	0	500
Suhag Nangla	Milak	21.7	17.4	60.9	0	1150
Dilpura	Sahid Nagar	36.4	2.4	51.4	9.8	2530
Bagad Khan	Sahid Nagar	9.1	2.3	81.8	6.8	4400
Gujroula	Bilaspur	46.7	10.8	9.2	33.3	1200
Rampura	Bilaspur	8.9	8.9	44.8	37.4	670
Patti Ashokpur	Chamraua	76.3	2.3	21.4	0	885
Baadli	Swar	45	55	0	0	500

Table A2: Mean comparison of socio-economic characteristics of high and low caste individuals

Characteristics	High caste	Low caste	T test (p values)
Age (years)	41-50	31-40	**0.00**
Gender			
Male	281 (89.5%)	357 (84.8%)	0.06
Female	33 (10.5%)	64 (15.2%)	
Family size	6.21	6.78	**0.01**
Marital status			
Married	258 (82.2%)	342 (81.2%)	0.75
Unmarried	56 (17.8%)	79 (18.8%)	
Annual income (INR)	5.4	3.8	**0.00**
< 12000=1			
12001-20000=2			
20001-30000=3			
30001-40000=4			
40001-50000=5			
50001-60000=6			
60001-70000=7			
70001-80000=8			
80001-90000=9			
90001-100000=10			
>100000=11			
Education (years)	9.50	5.21	**0.00**
Land area (acre)	2.24	0.76	**0.00**
House type			
Brick house	202 (64.3%)	174 (41.3%)	**0.00**
Mud/Thatched house	112 (35.7%)	247 (58.7%)	

Table A3: Balance test across treatments

Characteristics	**Marginal returns[1]** F test (p value)	**Caste[2]** T test (p value)
Age (years)	0.08	0.41
Education (years)	0.32	0.40
Land area (acre)	0.12	0.06
Annual Income (INR)	0.22	0.27
Family size	0.72	0.56

Note: [1] Mean comparison between Equal, High and Low returns. [2] Mean comparison between Caste treatment and Anonymous treatment

Table A4: PG contributions in anonymous groups

	(1) Contributions	(2) Contributions	(3) Contributions (No punishment)	(4) Contributions (With Punishment)
Constant	85***	101.3***	81.07**	159.1***
	(3.954)	(32.59)	(39.12)	(56.90)
High caste (HA)[1]	-4.444	-14.06	-3.749	-24.37
	(5.998)	(21.48)	(27.73)	(35.82)
Return[2]				
High		5.037	-0.993	11.07
		(6.880)	(7.674)	(11.97)
Low		-18.87***	-14.88*	-22.86**
		(6.620)	(7.829)	(11.04)
Punishment		37.50***		
(1=yes, 0=no)		(5.291)		
Education (years)		0.319	0.473	0.166
		(0.731)	(0.811)	(1.295)
Male		-0.218	9.255	-9.690
		(12.72)	(9.521)	(25.26)
Age (years)[3]				
31-60		-13.03*	-7.641	-18.42
		(6.647)	(7.606)	(11.35)
>60		-18.12	-21.97*	-14.26
		(12.33)	(12.81)	(22.28)
Family size		-1.931*	-2.080*	-1.781
		(1.169)	(1.242)	(2.073)
House type		9.383	7.624	11.14
(1=Brick house		(6.643)	(8.056)	(11.07)
0=Thatched/mud house)				
Land area (Acre)		1.793	2.879	0.706
		(1.531)	(1.934)	(2.406)
Order of Punishment	No	Yes	Yes	Yes
Village FE	No	Yes	Yes	Yes
Village FE	No	Yes	Yes	Yes
Observations	240	240	120	120
R-squared	0.002	0.280	0.165	0.158

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: [1] LA is the base category. [2] Equal return is the base category. [3] 18-30 years is the base category.

Table A5: PG contributions in homogeneous caste groups

	(1) Contributions	(2) Contributions	(3) Contributions (No Punishment)	(4) Contributions (With Punishment)
Constant	81.17***	54.78***	69.55***	78.98***
	(2.245)	(10.45)	(13.78)	(14.98)
High caste (HH)[1]	10.28***	20.11***	15.37**	24.84***
	(3.701)	(5.696)	(7.643)	(8.499)
Return[2]				
High		14.08***	8.766	19.40***
		(4.440)	(5.783)	(6.646)
Low		-10.24***	-5.952	-14.52***
		(3.802)	(5.162)	(5.445)
Punishment		38.97***		
(1=yes, 0=no)		(3.107)		
Education (years)		0.225	-0.452	0.903*
		(0.365)	(0.484)	(0.542)
Male		-1.489	-0.505	-2.473
		(4.646)	(6.127)	(7.035)
Age (years)[3]				
31-60		0.476	1.057	-0.106
		(3.825)	(4.929)	(5.825)
>60		-8.748	-3.798	-13.70
		(6.783)	(9.589)	(9.310)
Family size		1.328**	0.823	1.834**
		(0.665)	(0.915)	(0.892)
House type		4.878	3.821	5.936
(1=Brick house		(3.464)	(4.664)	(5.065)
0=Thatched/mud house)				
Land area (Acre)		-2.058**	-0.500	-3.616***
		(0.863)	(1.054)	(1.312)
Order of Punishment	No	Yes	Yes	Yes
Village FE	No	Yes	Yes	Yes
Enumerator FE	No	Yes	Yes	Yes
Observations	780	780	390	390
R-squared	0.010	0.286	0.134	0.241

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

Note: [1] LL is the base category. [2] Equal return is the base category. [3] 18-30 years is the base category.

Table A6: PG contributions under types of marginal returns and caste matching groups

	(1) Contributions	(2) Contributions	(3) Contributions	(4) Contributions (no punishment)	(5) Contributions (with punishment)
Constant	89.07***	61.82***	50.74***	68.38***	74.57***
	(2.157)	(2.904)	(8.713)	(11.41)	(12.56)
Return[1]					
High	8.188**	9.228***	8.253**	6.089	10.42**
	(3.263)	(3.255)	(3.250)	(4.236)	(4.950)
Low	-13.85***	-11.62***	-13.10***	-8.501**	-17.70***
	(3.054)	(3.131)	(3.074)	(4.085)	(4.563)
Punishment		41.46***	41.46***		
(1=yes, 0=no)		(2.644)	(2.515)		
Caste matching[2]					
High-High		8.945***	11.63**	7.532	15.73**
		(3.354)	(4.794)	(6.404)	(7.104)
High-Low		11.64***	8.929*	3.057	14.80*
		(3.798)	(5.040)	(6.663)	(7.615)
Low-High		10.69***	11.32***	7.914	14.72**
		(3.921)	(4.146)	(5.793)	(5.919)
Education (years)			0.547*	0.0587	1.035**
			(0.311)	(0.422)	(0.456)
Male			1.789	0.346	3.231
			(4.105)	(5.435)	(6.208)
Age (years)[3]					
31-60			4.085	3.401	4.770
			(3.038)	(3.980)	(4.604)
>60			-5.239	-1.926	-8.552
			(5.156)	(7.031)	(7.467)
Family size			0.0503	-0.393	0.494
			(0.491)	(0.641)	(0.732)
House type			-0.459	0.396	-1.313
(1=Brick house			(2.923)	(3.971)	(4.278)
0=Thatched/mud house)					
Land area (acres)			-0.788	-0.100	-1.475*
			(0.586)	(0.746)	(0.832)
Order of punishment	No	No	Yes	Yes	Yes
Village FE	No	No	Yes	Yes	Yes
Enumerator FE	No	No	Yes	Yes	Yes
Observations	1,470	1,230	1,230	615	615
R-squared	0.031	0.199	0.294	0.141	0.210

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: [1] Equal return is the base category. [2] Low-Low is the base category. [3] 18-30 years is the base category.

Table A7: Caste and punishment behavior

	(1) Punishment	(2) Punishment	(3) Punishment	(4) Punishment
Constant	1.958***	4.543***	4.643***	5.116***
	(0.0995)	(0.563)	(0.558)	(0.677)
Caste matching[1]				
High Caste	0.0935	0.439	0.544*	
	(0.174)	(0.307)	(0.281)	
Heterogeneous caste	0.704***	0.431**		0.411**
	(0.215)	(0.207)		(0.205)
High caste and Heterogeneous	-0.0261	0.110		0.171
Caste	(0.355)	(0.388)		(0.387)
Caste[2]				
Dhobi				-1.162*
				(0.600)
Gupta				1.066**
				(0.517)
Jatav				-0.283
				(0.311)
Saxena				-0.467
				(0.528)
Thakur				-0.139
				(0.350)
Valmiki				-0.770*
				(0.439)
Contribution		0.00548***	0.00573***	0.00532***
		(0.00168)	(0.00170)	(0.00168)
Contingent contributions[3]				
50		-1.802***	-1.802***	-1.802***
		(0.0644)	(0.0644)	(0.0645)
100		-3.088***	-3.088***	-3.088***
		(0.0876)	(0.0876)	(0.0876)
150		-4.086***	-4.086***	-4.086***
		(0.104)	(0.104)	(0.104)
200		-4.816***	-4.816***	-4.816***
		(0.118)	(0.118)	(0.118)
250		-5.195***	-5.195***	-5.195***
		(0.129)	(0.129)	(0.129)
Return[4]				
High		-0.366**	-0.394**	-0.340**
		(0.172)	(0.172)	(0.171)
Low		0.176	0.133	0.195
		(0.170)	(0.171)	(0.170)
Age (years)[5]				
31-60		0.135	0.139	0.128
		(0.163)	(0.165)	(0.167)
>60		-0.223	-0.236	-0.205
		(0.274)	(0.274)	(0.279)
Male		0.0936	0.0934	0.0923
		(0.197)	(0.197)	(0.198)

Education (years)		0.00990	0.0156	0.0111
		(0.0163)	(0.0163)	(0.0164)
Married		0.215	0.229	0.176
		(0.213)	(0.216)	(0.209)
Family size		0.0368	0.0378	0.0311
		(0.0348)	(0.0355)	(0.0338)
Land area (acres)		-0.00624	-0.00730	-0.00494
		(0.0268)	(0.0271)	(0.0269)
Order of punishment	No	Yes	Yes	Yes
Village FE	No	Yes	Yes	Yes
Enumerator FE	No	Yes	Yes	Yes
Observations	3,690	3,690	3,690	3,690
R-squared	0.013	0.472	0.468	0.477
F stat: $pun^{HL}=pun^{HH}$	8.25	3.23		
P value	0.00	0.04		

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: [1] Low-Low is the base category; [2] Brahmin is the base category; [3] 0 contribution is the base category; [4]Equal return is the base category; [5] 18-30 years is the base category. Standard errors are clustered at individual level

Table A8: PG contributions under heterogeneous caste and returns

	(1) Contributions	(2) Contributions (no punishment)	(3) Contributions (with punishment)
Constant	55.93***	79.51***	75.77***
	(15.84)	(21.82)	(19.99)
High caste[1]	-18.14**	-13.25	-23.03*
	(8.713)	(11.50)	(13.12)
High return[2]	3.498	1.642	5.354
	(8.378)	(11.46)	(12.50)
High caste and High return	29.72**	26.69*	32.76*
	(11.60)	(15.83)	(17.39)
Punishment	43.43***		
(1=yes, 0=no)	(5.209)		
Education (years)	0.868	0.915	0.822
	(0.630)	(0.842)	(0.951)
Male	6.921	-6.379	20.22
	(10.29)	(14.96)	(13.69)
Age (years)[3]			
31-60	18.12***	11.43	24.81***
	(5.703)	(7.742)	(8.484)
>60	-1.436	-6.985	4.113
	(12.09)	(15.75)	(17.45)
Family size	-2.193**	-1.915*	-2.471*
	(0.868)	(0.985)	(1.454)
House type	-7.570	-6.157	-8.983
(1=Brick house	(7.527)	(9.442)	(11.97)
0=Thatched/mud house)			
Land area (acre)	-0.745	-2.614	1.124
	(1.875)	(2.215)	(3.086)
Order of punishment	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes
Observations	274	137	137
R-squared	0.401	0.306	0.318

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

Note: [1] Low caste is the base category; [2] Low return is the base category; [3] 18-30 years is the base category. We did not control for village fixed effects as controlling it due to issues with degrees of freedom. Instead, using a dummy variable for the villages based on proportion of high and low caste, we performed the above regressions (not reported here) and found similar results.

Table A9: PG contributions of high and low caste under heterogeneous returns and anonymity

	(1) Contributions	(2) Contributions (no punishment)	(3) Contributions (with punishment)
Constant	77.05***	66.65**	125.1**
	(28.51)	(31.42)	(51.39)
High Caste[1]	-18.25	-14.69	-21.82
	(11.11)	(11.17)	(20.34)
High Return[2]	-31.42***	-19.57**	-43.27***
	(8.379)	(9.700)	(14.07)
High Caste and High Return	10.47	11.10	9.827
	(12.59)	(13.58)	(21.99)
Punishment	37.66***		
(1=yes, 0=no)	(6.062)		
Education (years)	0.402	0.540	0.264
	(0.818)	(0.826)	(1.523)
Male	6.738	11.59	1.888
	(12.73)	(10.17)	(23.91)
Age (years)[3]			
31-60	-5.906	-1.762	-10.05
	(8.072)	(9.071)	(13.91)
>60	-21.92**	-27.71*	-16.12
	(9.913)	(14.59)	(14.25)
Family size	-3.720***	-2.792*	-4.649*
	(1.380)	(1.459)	(2.460)
House type	15.68*	6.901	24.47
(1=Brick house	(8.478)	(9.441)	(15.21)
0=Thatched/mud house)			
Land area (acre)	1.516	3.128	-0.0961
	(1.424)	(2.157)	(1.807)
Order of punishment	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes
Observations	154	77	77
R-squared	0.351	0.205	0.256

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: [1] Low caste is the base category; [2] Low return is the base category; [3] 18-30 years is the base category.

Table A10: Payoff under caste and return heterogeneity

	(1) Payoff	(2) Payoff	(3) Payoff (no punishment)	(4) Payoff (with punishment)
Constant	256.3***	269.0***	261.8***	234.1***
	(7.502)	(20.90)	(18.45)	(37.62)
High caste[1]	-17.34*	-4.211	6.413	-14.83
	(10.34)	(18.30)	(13.97)	(36.36)
High return[2]	17.89*	14.28	10.83	17.72
	(9.720)	(11.76)	(12.12)	(21.52)
High caste and High return	31.60**	33.73**	21.64	45.81
	(13.71)	(16.05)	(17.49)	(28.58)
Punishment		-42.19***		
(1=yes, 0=no)		(6.351)		
Education (years)		0.00300	1.149	-1.143
		(0.903)	(0.874)	(1.606)
Male		13.75	-6.826	34.32
		(15.57)	(14.88)	(29.76)
Age (years)[3]				
31-60		4.739	3.138	6.341
		(9.033)	(9.412)	(16.25)
>60		-1.581	-9.074	5.911
		(20.10)	(12.63)	(41.50)
Family size		0.568	2.357*	-1.222
		(1.463)	(1.236)	(2.603)
House type		12.06	-0.914	25.04
(1=Brick house		(10.27)	(9.124)	(18.56)
0=Thatched/mud house)				
Land area (acre)		-1.499	-0.695	-2.304
		(2.270)	(2.130)	(4.195)
Order of punishment	No	Yes	Yes	Yes
Village FE	No	Yes	Yes	Yes
Enumerator FE	No	Yes	Yes	Yes
Observations	274	274	137	137
R-squared	0.102	0.302	0.281	0.271

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: [1] Low caste is the base category; [2] Low return is the base category; [3] 18-30 years is the base category.

Table A11: PG contributions in homogeneous caste groups after priming

	(1) Contributions	(2) Contributions	(3) Contributions (No Punishment)	(4) Contributions (With Punishment)
Constant	131.6***	83.08***	78.38***	112.8***
	(2.519)	(10.56)	(13.80)	(16.15)
High caste (HH)[1]	4.211	10.43*	3.123	17.73*
	(4.009)	(6.275)	(8.052)	(9.760)
Return[2]				
High		19.71***	17.51***	21.91***
		(4.733)	(6.468)	(7.074)
Low		-25.06***	-24.42***	-25.71***
		(4.537)	(6.195)	(6.752)
Punishment		25***		
(1=yes, 0=no)		(3.455)		
Education (years)		1.060**	0.815	1.304**
		(0.423)	(0.554)	(0.657)
Male		6.215	7.527	4.903
		(5.245)	(7.160)	(7.930)
Age (years)[3]				
31-60		-4.573	-5.688	-3.457
		(4.415)	(5.888)	(6.673)
>60		-1.342	-4.087	1.404
		(7.040)	(9.733)	(10.54)
Family size		2.449***	2.789***	2.109**
		(0.713)	(1.006)	(1.043)
House type		2.972	1.984	3.961
(1=Brick house		(3.940)	(5.289)	(5.990)
0=Thatched/mud house)				
Land area (Acre)		-1.249*	0.401	-2.899***
		(0.700)	(0.888)	(1.039)
Order of Punishment	No	Yes	Yes	Yes
Village FE	No	Yes	Yes	Yes
Enumerator FE	No	Yes	Yes	Yes
Observations	780	780	390	390
R-squared	0.001	0.258	0.230	0.237

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: [1] LL is the base category. [2] Equal return is the base category. [3] 18-30 years is the base category.

Table A12: Margins for contributions in homogeneous caste groups after priming

Caste matching	Predicted average	Marginal difference
		LL
Pooled		
HH	**139.55**	*-10.43**
No punishment		
HH	**122.65**	*-3.12*
With punishment		
HH	**156.45**	*-17.73**

Note: Refer to Table A11 for the regression output. ***, ** and * indicates significance at 1%, 5% and 10% respectively. HH and LL indicate High and Low caste homogeneous groups respectively.

Table A13: PG contributions under heterogeneous caste and returns after priming

	(1) Contributions	(2) Contributions (No Punishment)	(3) Contributions (With Punishment)
Constant	90.32***	100.2***	116.2***
	(15.74)	(18.84)	(24.78)
High Caste[1]	-6.899	-5.200	-8.598
	(8.822)	(11.50)	(13.47)
High Return[2]	9.767	10.98	8.555
	(8.051)	(10.78)	(11.98)
High caste and High Return	20.50*	22.93	18.08
	(12.23)	(17.04)	(18.05)
Punishment	35.77***		
(1=yes, 0=no)	(5.406)		
Education (years)	0.188	0.102	0.274
	(0.628)	(0.875)	(0.946)
Male	-4.166	-1.157	-7.176
	(10.12)	(12.27)	(16.36)
Age (years)[3]			
31-60	5.205	3.017	7.393
	(6.101)	(8.220)	(9.150)
>60	-17.96	-13.21	-22.72
	(12.21)	(11.78)	(21.17)
Family size	-0.801	-0.931	-0.670
	(0.971)	(1.144)	(1.584)
House type	-16.24**	-10.36	-22.13**
(1=Brick house	(6.992)	(9.124)	(10.94)
0=Thatched/mud house)			
Land area (acre)	0.691	-0.744	2.126
	(1.896)	(2.357)	(3.021)
Order of Punishment	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes
Observations	274	137	137
R-squared	0.326	0.233	0.288

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

Note: [1] Low caste is the base category. [2] Low return is the base category. [3] 18-30 years is the base category. We did not control for village fixed effects as controlling it due to issues with degrees of freedom. Instead, using a dummy variable for the villages based on proportion of high and low caste, we performed the above regressions (not reported here) and found similar results.

Table A14: Margins of PG contributions under both caste and returns heterogeneity after priming

Caste-Return matching	Predicted Average	Marginal difference		Difference of marginal differences (p value)
		High caste-High return	Low caste-High return	
Pooled				
High caste-Low return	**108.71**	*30.27****		
Low caste-Low return	**115.61**		*9.78*	*0.09*
No punishment				
High caste-Low return	**90.27**	*33.90****		
Low caste-Low return	**95.47**		*10.98*	*0.18*
With punishment				
High caste-Low return	**127.14**	*26.63***		
Low caste-Low return	**135.74**		*8.55*	*0.32*

Note: Refer to Table A13 for the regression output. ***, ** and * indicate the significance at 1%, 5% and 10% respectively.

Table A15: Caste and punishment behavior after priming

VARIABLES	(1) Punishment	(2) Punishment	(3) Punishment	(4) Punishment
Constant	2.040***	4.024***	4.286***	4.326***
	(0.0809)	(0.410)	(0.410)	(0.497)
Caste matching[1]				
High caste	0.153	0.349	0.508**	
	(0.130)	(0.218)	(0.213)	
Heterogeneous caste	0.821***	0.762***		0.760***
Treatment	(0.169)	(0.164)		(0.162)
High caste and Heterogeneous	0.0307	0.108		0.135
caste treatment	(0.295)	(0.325)		(0.327)
Caste[2]				
Dhobi				-0.813*
				(0.421)
Gupta				-0.0965
				(0.407)
Jatav				-0.259
				(0.237)
Saxena				-0.193
				(0.518)
Thakur				-0.0465
				(0.303)
Valmiki				-0.342
				(0.331)
Contribution		0.00823***	0.00779***	0.00825***
		(0.00150)	(0.00159)	(0.00149)
Contingent contributions[3]				
50		-2.159***	-2.159***	-2.159***
		(0.0695)	(0.0695)	(0.0696)
100		-3.519***	-3.519***	-3.519***
		(0.0861)	(0.0860)	(0.0861)
150		-4.654***	-4.654***	-4.654***
		(0.101)	(0.101)	(0.101)
200		-5.470***	-5.470***	-5.470***
		(0.114)	(0.114)	(0.114)
250		-5.928***	-5.928***	-5.928***
		(0.123)	(0.123)	(0.123)
Return[4]				
High		-0.337**	-0.374***	-0.332**
		(0.134)	(0.138)	(0.135)
Low		0.0884	-0.00247	0.0965
		(0.136)	(0.141)	(0.135)
Age (years)[5]				
31-60		0.0756	0.0810	0.0827
		(0.135)	(0.140)	(0.137)
>60		-0.133	-0.164	-0.124
		(0.251)	(0.257)	(0.254)

Male		0.164	0.166	0.172
		(0.149)	(0.152)	(0.149)
Education (years)		0.00411	0.0148	0.00589
		(0.0137)	(0.0136)	(0.0141)
Married		0.218	0.229	0.204
		(0.165)	(0.170)	(0.162)
Family size		0.0494*	0.0517*	0.0491*
		(0.0293)	(0.0305)	(0.0287)
Land area (acres)		-0.0182	-0.0215	-0.0176
		(0.0220)	(0.0232)	(0.0216)
Order of punishment	No	Yes	Yes	Yes
Village FE	No	Yes	Yes	Yes
Enumerator FE	No	Yes	Yes	Yes
Observations	3,690	3,690	3,690	3,690
R-squared	0.021	0.598	0.586	0.599
F stat: $pun^{HL}=pun^{HH}$	18.00	13.84		
P value	0.00	0.00		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: [1] Low-Low is the base category; [2] Brahmin is the base category; [3] 0 contribution is the base category; [4]Equal return is the base category; [5] 18-30 years is the base category. Standard errors are clustered at individual level.

Table A16: Margins for punishment behavior after priming

Caste matching	Predicted margins	Marginal difference			
		HL	LL	LH	H
HH	**2.31**	*0.87****	*-0.35*		
LL	**1.96**			*0.76****	
L	**2.19**				*0.51***

Note: Refer to Table A15 for the regression output. ***, ** and * indicate the significance at 1%, 5% and 10% respectively

Table A17: Payoff under caste and return heterogeneity after priming

	(1) Payoff	(2) Payoff	(3) Payoff (No punishment)	(4) Payoff (With punishment)
Constant	265.3***	282.4***	269.0***	264.0***
	(5.659)	(21.27)	(23.49)	(36.67)
High caste[1]	-9.799	8.963	3.548	14.38
	(8.291)	(10.55)	(15.14)	(15.94)
High return[2]	40.26***	35.30***	25.85**	44.75***
	(8.234)	(8.954)	(10.65)	(14.76)
High caste and High return	18.44	19.29	16.02	22.56
	(11.36)	(12.67)	(15.97)	(21.36)
Punishment		-31.75***		
(1=yes, 0=no)		(5.372)		
Education (years)		0.252	1.318	-0.814
		(0.740)	(0.839)	(1.167)
Male		3.072	8.684	-2.539
		(12.13)	(17.33)	(18.88)
Age (years)[3]				
31-60		-3.834	3.384	-11.05
		(7.470)	(8.532)	(12.22)
>60		13.28	14.36	12.20
		(14.08)	(13.27)	(25.67)
Family size		-0.441	-0.334	-0.547
		(1.369)	(1.497)	(2.379)
House type		7.787	6.340	9.233
(1=Brick house		(7.425)	(8.780)	(12.10)
0=Thatched/mud house)				
Land area (acre)		-0.568	-1.077	-0.0580
		(1.732)	(2.332)	(3.042)
Order of punishment	No	Yes	Yes	Yes
Village FE	No	Yes	Yes	Yes
Enumerator FE	No	Yes	Yes	Yes
Observations	274	274	137	137
R-squared	0.225	0.384	0.343	0.420

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

Note: [1] Low caste is the base category; [2] Low return is the base category; [3] 18-30 years is the base category.

Table A18: Margins of Payoff under caste and return heterogeneity after priming

Caste-Return matching	Predicted margins	Marginal difference	
		High caste-High return	Low caste-High return
Pooled			
High caste-Low return	**267.05**	*54.59****	
Low caste-Low return	**258.09**		*35.30****
No punishment			
High caste-Low return	**286.56**	*41.86****	
Low caste-Low return	**283.02**		*25.85***
With punishment			
High caste-Low return	**247.54**	*67.32****	
Low caste-Low return	**233.16**		*44.75****

Note: Refer to Table A17 for the regression output. ***, ** and * indicate the significance at 1%, 5% and 10% respectively.

Table A19: Inequality under caste and return heterogeneity after priming

	Caste matching				Kruskal-Wallis test (p values)
	H*L	n	HL*	n	
Public good difference					
pooled	75.96 (21.67)	78	69.67 (24.33)	76	0.09
Without pun	65.38 (16.32)	39	58.42 (19.66)	38	0.09
With pun	86.54 (21.34)	39	80.92 (23.53)	38	0.28

Note: H*L indicates that the high caste gets higher marginal returns and the low caste gets lower marginal returns. Similarly, HL* indicates that the low caste gets higher marginal returns and the high caste gets lower marginal returns. Standard deviations in the parenthesis.