Can Voters be Primed to Choose Better Legislators? Experimental Evidence from Rural India

Abhijit Banerjee, Donald Green, Jennifer Green and Rohini Pande *

Preliminary and Incomplete

Abstract

Ethnic politics and political corruption are important concerns in many developing country democracies. We conducted field experiments in rural India to examine the responsiveness of voter preferences to priming about the relevance of ethnicity and politician quality for service delivery. We used vignette experiments to examine how voter preferences vary with information about politician quality. We find strong evidence that ethnic preferences are malleable. The vignette experiments show that ethnic preferences respond to politician quality and the field experiment shows that reductions in ethnic voting reduce the vote share of criminal politicians. In contrast, priming voters on the relevance of corruption left electoral outcomes unaffected (in the field experiment). We interpret these results in the context of a model where a voter's use of specific evaluative criteria is influenced by available information.

*The authors are from MIT, Yale (Green and Green) and Harvard University respectively. We are grateful to ISPS Yale and Hewlett Foundation for funding. We thank our NGO partner, Sarathi, our field coordinator Bhartendu Trivedi and Mike Mueller, Terence Leong and Felix Su for research assistance. We owe a special thanks to Jeffrey McManus for exceptional field work and research assistance.

1 Introduction

Much has been written about the hypothesis that voting based on ethnicity undermines the effectiveness of democratic institutions and leads to poor policy outcoms. Bloc voting blunts the force of electoral competition: candidates who enjoy widespread ethnic or caste support can afford to under-perform along other dimensions. Classic arguments of this kind may be found in Key (1949), who notes that the cause of maintaining white hegemony undercut two-party competition in the American South, which in turn led to rampant problems of corruption and stunted economic development (Besley et al., 2008).[1] There is a related concern about the "tyranny of the majority" – the numerically dominant ethnic group might exploit its ability to win elections without the support of others by monopolizing state power and the resources that it commands. Hence, in the long run ethnic politics can impoverish the country and lead to civil unrest and democratic instability (Horowitz, 1985; Miquel, 2006).[2]

The goal of this paper is to investigate whether citizens' tendency to vote along ethnic/caste lines is malleable, given the significant cost that they pay when legislators behave in a corrupt manner. If voters systematically differ in their policy needs and propensity to engage in ethnic voting, then it is hard to use observational data to identify the causal link between voter preferences and ethnic voting on one hand and between ethnic voting and the quality of governance on the other hand. Therefore, to gain insights into voter motivations we implemented and analyze a series of field and vignette experiments.[3].

During the 2007 UP state assembly elections, we conducted a randomized evaluation of two voter mobilization campaigns implemented by a prominent local NGO. One campaign primed voters

[1] Recently, scholars such as Dickson and Scheve (2006) have introduced social identities into formal models of campaign competition to explain how voters' group-centered preferences can diminish the role of issues in campaigns and, by extension, government outputs (Ferejohn, 1986).

[2] Although ethnic voting may ultimately be traced to policy-based motives, with some scholars arguing that it stems from voter heuristics about which party best serves their group's policy interests (Ferree, 2004), empirically ethnic voting appears to be correlated with a deterioration in government accountability (Banerjee and Pande, 2009) .

[3] While much of this literature was developed in the US context (Gerber and Green, 2000; Gerber et al., 2008), a number of recent papers extend the ambit of such studies to low income countries (Wantchekon, 2003; Vicente and Wantchekon, 2009)

to not vote on ethnic lines and the other primed voters to not vote for corrupt candidates. Both campaigns were non-partisan and simply stated that voters should vote for development. Roughly two and a half years later we conducted a series of vignette experiments in one of our study settings where we examined the extent to which ethnic preferences are influenced by information about candidates' criminal and corruption record.

The field experiments allow us to directly examine the malleability of voter preferences to priming about the *production function* for service delivery. The vignettes allow us to examine whether voter choices are sensitive to the available information about candidate quality.

Official voter turnout statistics show that our caste field experiment mobilized voters, particularly male voters. Based on electoral data, 8 percent more male voters registered in the villages where the campaign was carried out, and 11 percent more voted (raising male turnout from 60 to roughly 67 percent). The effects for women are more noisily estimated, but statistically indistinguishable from male outcomes. Because individual voters' choices are anonymous, we cannot use electoral data to examine whether the campaign reduced caste based voting. We, therefore, use survey data collected after the elections (but before the results were announced). The data show that while 57.6% of voters voted for their caste preferred party in the control villages, this figure drops approximately by 10% in the treatment villages. On the other hand the corruption campaign had no significant impact on electoral outcomes or vote share of candidates identified as being corrupt (by local journalists).

In contrast to the caste campaign, the corruption campaign had no effect on registration or overall turnout (though we see some evidence that female turnout increased). Nor did it affect the vote share of candidates who, according to a pre-election poll of local journalists that we conducted, were the most corrupt. It also did not affect the vote share of the candidates with heinous crime charges, unlike in the caste intervention.

If the theory that caste-based voting reduces legislative performance is correct, reduced votes for the caste-preferred party should mean fewer votes for low-quality candidates. Ordinarily, assessing candidate quality means invoking subjective evaluative criteria (Krasno and Green, 1988),

but Indian law requiring that candidates disclose whether they were charged with "heinous crimes" (i.e., offenses carrying a sentence of more than 5 years) provides an unusually clear objective standard. We find a strong relationship between the campaign intervention and support for candidates with criminal backgrounds. While in the control villages candidates charged with heinous crimes on average receive 65% more votes than their opponents, this gap is halved in treatment villages. This finding is paralleled in the survey data: the treatment only changes the propensity to vote for the caste-preferred candidate when he has a heinous crime on his record.

But why did the corruption camoaign not affect voter decisions? One possibility is that voters were uncertain about how to respond to the anti-corruption message, lacking any explicit guidance about which candidates were corrupt. There is some tension between this view and the fact that voters exposed to the caste intervention seemed to know whether the candidate from their caste-preferred party was a criminal. On the other hand, being corrupt and being a criminal are not the same thing–the correlation is only 0.1 in our data, and competing parties can and do routinely charge opposing candidates with corruption, while charges of past criminality are facts that can be verified independently. Further evidence that candidate quality matters comes from our vignette experiment. Citizens exhibit significant own ethnic party preference but this preference is halved when the vignette informs them that their preferred candidate is either a criminal or corrupt. The responsiveness to information is greater for criminal candidates when the basis for information is certain (i.e. the candidate had served a criminal sentence not just been charged) or when the voter is better informed about politics.

In sum, our results may be interpreted to mean that anti-caste campaigns which simultaneously emphasize the importance of voting to enable development are more effective vehicles for influencing voting patterns for two reasons: these campaigns undercut an influential decision heuristic while focusing attention on other important evaluative criteria. Nonpartisan anti-corruption campaigns, by contrast, leave the dominant evaluative criteria largely intact and do not provide the information necessary to reshape voter preferences.

One way to interpret this pattern is that there is a substantial group of voters with low levels

of information about the candidates. These voters tend to support the party that is thought to represent their caste, perhaps because it is the one clear thing they know that distinguishes the parties (on voters' use of reference groups and heuristics, see Popkin, 1991; Lupia, 1994). When told that caste is not a good reason to vote for someone, these voters may be willing to consider another strategy, such as voting for whichever candidate looks better along the other dimensions, such as past criminality. This interpretation suggests that the vote share of criminal candidates declines because voters have changed their evaluative criteria.[4]HOW SHOULD WE THINK ABOUT VIGNETTES??

The rest of the paper is organized as follows. Section 2 introduces the electoral context within which the experiment took place, noting three salient features of UP elections: the importance of caste voting, the checkered backgrounds of legislative candidates, and the low levels of information with which voters make decisions. Section 3 provides a simple model to interpret these findings. Section 4 describes our sampling and empirical strategies, including some complications that require the use of instrumental variables estimation in order to correct some potential biases. Section 5 describes the two campaigns, and Sections 6 and 7 describe the results of the caste and corruption campaigns.

2 Electoral context: Uttar Pradesh

UP is arguably one of the most important subnational political entities. Its population of more than 190 million makes it the most populous subnational unit; indeed, if UP were a country, it would be the sixth most populous in the world. It is also one of the poorest states in India, with a total GDP of 69 billion dollars in 2006-07. More than two-thirds of the population is engaged in agriculture. Literacy rates are 70% among men and 43% among women. Elections in UP therefore occur on a

[4] The share of the criminal candidates would go down even if people who switch away from voting on caste vote at random for one of the other parties, if, as seems to be the case, candidates supported by the majority caste group are more likely to be criminals. However in order to explain the fact that the switch away from the caste-preferred candidate happens only when the caste-preferred candidate is a criminal, we do need to assume that the voters have some idea of whether their caste-preferred candidate is a criminal.

massive scale and within a context in which large segments of the electorate have limited access to information and material resources.

2.1 The Rise of Caste Politics

More than 80% of the population is Hindu, and UP politics are closely tied to the Hindu caste system. Historically, the caste system has divided Hindu society into a hierarchically ordered set of endogamous groups, with groups lower in that hierarchy categorized by the present political system as Scheduled Castes (SC) and Other Backward Castes (OBC). At Independence, the Congress Party, which drew its leadership from the upper castes, dominated UP politics. Although non-Congress parties briefly came to power during the late 1960s and again during the late 1970s, the Congress hegemony in UP went largely unchallenged until the mid-1980s. Prior to the mid-1980s, the main opposition parties were dominated by upper caste leaders and candidates, and low caste legislators were mainly confined to reserved jurisdictions, where only SC candidates could stand for election. In 1984, an explicitly SC party, the Bahujan Samaj Party (BSP), was formed. A second low caste party which mainly targeted OBC voters, the Samajwadi Party (SP), was formed in 1992. Since the early 1990s, one or both of these two parties have formed the majority coalition in the state legislature, enabling party leaders to be named Chief Minister.[5]

Surveys of the UP electorate show the prevalence of caste voting in the wake of this party realignment. In 1996, for example, over 70% of voters reported voting for the party that represented their caste group (the Congress and the Bharatiya Janata Party, or BJP, for upper caste voters, and SP and BSP for lower caste voters). Explanations for caste voting both in UP and India more generally abound. Yadav (2000) argues that ethnicity was made politically salient in the 1980s by the growth of popular low caste movements spearheaded by charismatic individuals who went on to form low caste parties. Chandra (2004) and Jaffrelot (2003) argue that the political salience of caste

[5] Muslims, who account for approximately 18% of the UP electorate, are not directly subject to the caste system, and their relationship to the party system is complex, with allegiances shifting over time in response to the ways in which the various parties appeal to Muslims' economic interests and aspirations for social equality.

identities increased in the wake of affirmative action policies in favor of low castes. These policies created a large group of middle-class, low-caste citizens who demanded political recognition and social change, and caste-based quotas led them to organize politically along these lines. Conversely, opposition to affirmative action may have contributed to the hardening of the political stances of upper caste Hindus along both caste and religious lines and was reflected in the growing political influence of BJP. The caste composition of electoral districts, which was a weak predictor of electoral returns in the late 1970s, became a strong predictor by the late 1980s and continues to be the most important determinant of election results (Banerjee and Pande, 2009).

2.2 Criminality and Corruption in UP politics

Accompanying the realignment of the party system has been a striking increase in prominence of criminals among the ranks of elected officials. Although the upward trend was clear to those close to UP politics, the proportion of candidates with criminal backgrounds could not be determined with precision until relatively recently. In the wake of a 2002 Supreme Court ruling declaring the right of voters to know candidates' backgrounds, election officials now require that candidates disclose their criminal records and financial assets. According to an affidavit filed by the Government of UP in 2005, on the eve of the 2007 election 206 of the 403 sitting members of the legislative assembly (MLAs) had pending criminal charges against them. Remarkably, 159 sitting MLAs, distributed more or less evenly among political parties, were charged with what are termed heinous crimes. The legal definition of heinous crimes in the context of elections is provided in Section 8 of the Representation of People Act 1951: a criminal charge which attracts more than 5 years of punishment. Although voters were presented with ample opportunities to vote for criminals (882 candidates had pending charges against them, and at least six campaigned from jail), the 2007 election saw a decline in the proportion of elected officials with criminal backgrounds: 155 elected MLAs had criminal backgrounds, 91 of whom were charged with heinous crimes.

Criminal candidates are arguably advantaged in the current environment in two ways. First,

in contrast to the old Congress Party, which carefully monitored candidate recruitment and advancement within the party amid relatively limited inter-party competition, the current array of parties have relatively few checks on the advancement of ambitious and well-funded candidates. Candidates who in effect buy their way onto a party's ticket provide an important public good for the party, subsidizing other candidates' campaigns. Second, voters tend to have little information about the candidates other than caste-related cues. In 2007, a representative survey of the UP electorate found that 68% of men and 89% of women reported that they "never read the newspaper", and similar proportions reported never listening to news on TV or radio. In this kind of environment, elections are shaped either by reference group cues such as caste and by campaign events (processions, rallies, gift-giving) that cost money, which in turn amplifies the importance of self-financed candidates.

3 An elementary model

3.1 Set up

Consider a voter who has two decisions–whether to vote and who to vote for. His cost of voting, c, is randomly drawn from a distribution $G(c)$ on $[c_0, c_1]$, where c_0 is negative (some people actually like voting) but $c_1 > 0$. If the voter votes he decides beween two candidates, A and B. If voter does not vote his utility from the entire electoral process is set to 0. If he votes, he pays the cost of voting but has the satisfaction of having voted for the best person: If the expected value he puts on candidate i is v_i,then his payoff from having voted is $\max\{v_A - v_B, v_B - v_A\}$.

The value of a candidate comes from what he is expected to deliver to the voter. Candidates differ along two dimensions which we will call P and Q. Q is quality of the candidate, that is his ability to get things done, his probity, his commitment to his job, etc. Assume that Q takes its values in $\{-1, 1\}$, with 1 representing higher quality. P represents the ethnic affiliation of the candidate (P for parochial interest) and takes the value 1 if he is affiliated with the same group as

the voter and -1 otherwise.

We assume that P is known to voters and that candidate A has $P = 1$, while candidate B has $P = -1$. However the voter does not know the exact value of Q: He has a prior that the vector $(Q_A\ , Q_B\ )$ can be either (1,-1) or (-1,1) with probabilities p and $1 - p$. In particular p need not be equal to 1/2, allowing for the possibility that ex ante one candidate is believed to be higher quality than the other. In addition he gets a signals, s_A or s_B, about the quality of the candidates: $\Pr\{s_A|Q_A = 1, Q_B = 1\} = \Pr\{s_B|Q_A = -1, Q_B = 1\} = \theta \geq 1/2$.

The value of candidate i is given by the expression

$$v_i = \alpha Q_i^e + \beta P_i + \gamma D_i^e, \alpha, \beta, \gamma \geq 0$$

where Q_i^e is *expected* quality and D_i^e (delivery) is a measure of how much governmental resources candidate i is expected to deliver to the voter. We assume that the voter believes that

$$D_i^e = \mu Q_i^e + (1 - \mu) P_i, \mu, \nu \geq 0.$$

In other words, the voter values candidate quality and (the right) ethnicity potentially both in themselves and as a means to get stuff from the government.

We assume that α, β, γ are known and fixed parameters–the voter knows what he likes. However he does not necessarily know the true production function for D: μ and v are just his priors and additional information can influence them. WE SHOULD EXPAND THIS TO ALSO CONSIDER THE CASE WHERE INFORMATION IS VARIED BUT NOT PARAMETERS TO COVER THE VIGNETTE CASE Within this framework the NGO campaigns are naturally interpreted as providing informative signals about μ. The caste campaign provides a signal that makes the voter less optimistic about ethnicity as a delivery mechanism: The voters who were exposed to the campaign have a posterior $\mu' > \mu$. The corruption campaign emphasized the importance of quality in terms of getting delivery, which, in this somewhat abstract formulation also amounts to an increase in μ. Of course, in reality,

quality has multiple dimensions and the two campaigns probably emphasized different dimensions of quality–while corruption campaign focused on one specific dimension, honesty, the caste campaign left people free to decide the specific aspect of quality that they wanted to emphasize.

The timing of the model is assumed as follows: First s_A and s_B are revealed. Then the NGO campaign takes place (or not). Finally the voter decides whether to go to vote and if he votes, who to vote for.

3.2 Analysis

Given a signal $s = s_A, s_B$, if the voter votes, he will vote for candidate A if and only if

$$v_A(s) - v_B(s) = 2[(\beta + \gamma(1-\mu)) - (\Pr\{Q_B = 1|s\} - \Pr\{Q_A = 1|s\})(\alpha + \gamma\mu)] \geq 0$$

Moreover a voter will vote as long as

$$|v_A(s) - v_B(s)| = 2\,|(\beta + \gamma(1-\mu)) - (\{\Pr\{Q_B = 1|s\} - \Pr\{Q_A = 1|s\})(\alpha + \gamma\mu)| \geq \ c.$$

Clearly, no one would ever vote for candidate B unless $\Pr\{Q_B = 1|s_B\} - \Pr\{Q_A = 1|s_B\}$is positive enough that

$$(\Pr\{Q_B = 1|s_B\} - \Pr\{Q_A = 1|s_B\})(\alpha + \gamma) > \beta.$$

Therefore assume this condition holds and also that

$$\Pr\{Q_B = 1|s_A\} - \Pr\{Q_A = 1|s_A\} < 0,$$

i.e. that an A signal does make A the high quality candidate.. Let $\mu*$ be the value of μ such that

$$[(\beta + \gamma(1-\mu)) - (\Pr\{Q_B = 1|s_B\} - \Pr\{Q_A = 1|s_B\})(\alpha + \gamma\mu)] = 0.$$

For $\mu < \mu*$, people vote A both when they get signal $s_A$and when they get signal s_B. Once μcrosses $\mu*$, those who get signal $s_B$start to vote for B. Therefore turnout from those who have the signal is s given by

$$T(s) = \int^{2|(\beta+\gamma(1-\mu))-(\{\Pr\{Q_B=1|s\}-\Pr\{Q_A=1|s\})(\alpha+\gamma\mu)|} dG(c).$$

Expected turnout is given by

$$T = (p\theta + (1-p)(1-\theta))T(s_A) + ((1-p)\theta + p(1-\theta))T(s_B).$$

The expected vote share of candidate B is zero as long as $\mu \leq \mu*$, but once μ exceeds $\mu*$it is given by

$$p\frac{(1-\theta)T(s_B)}{\theta T(s_A) + (1-\theta)T(s_B)} + (1-p)\frac{\theta T(s_B)}{\theta T(s_B) + (1-\theta)T(s_A)}$$

Given these expressions it is easily checked that as long as $\mu \leq \mu*$, an increase in μwill only lead to a fall in turnout–everybody who votes will vote for A. Once $\mu > \mu*$, an increase in μ will increase the vote share of candidate B. The effect on turnout remaimns ambiguous since it continues to go down among those who get s_A but now increases among those who get s_B.

Result 1: An increase in μ(weakly) increases the share of votes that go to the ethnically non-favored candidate. It may also increase turnout.

Corollary: If turnout goes up, the vote share of the ethnically non-favored candidate must go up.

The increase in turnout is more likely the larger is $\Pr\{Q_B = 1|s_B\} - \Pr\{Q_A = 1|s_B\})$ and the weaker is pure ethinic preference β. $\Pr\{Q_B = 1|s_B\} - \Pr\{Q_A = 1|s_B\})$, in turn depends on how much information voters have.

Result 2: If the voters have little information about the quality of the candidates an increase in $\mu$is more likely to reduce or fail to increase turnout and the vote share of the ethnically non-favored candidate. This is also true for voter groups that have very strong ethnic preferences and those that believe that their ethnically favored candidate is better ex ante ($p > 1/2$).

Does a campaign that pushes for an increase in quality of the electors necessarily increase expected quality? In this model, an increase in μalways weakly increases the vote share of candidate B. This is irrespective of whether B is the better candidate ex ante ($p < 1/2$).Therefore there is no presumption that the campaign will increase the vote share of the higher quality candidate. The best case scenario of course is when signals are very precise (θclose to 1)

Result 3: An increase in μmay reduce the vote share of the higher quality of the electors unless the higher quality elector is also the ethnically non-favored candidate for the majority of the population

These results are quite stark because of the simplicity of our model. However the effects they rely on are much more general.

3.3 Empirical Implications

We will consider a situation where individuals are primed about the salience of either the caste or corruption dimension of their voting decision. The caste-preferred party of a voter is clearly observed, and the model predicts that citizens will be able to respond to the message by reducing the incidence of caste-preferred voting. Interpreting a generalized corruption message in terms of which candidate is corrupt is harder, and as discussed in this section we would expect significant heterogeneity in voter response.

4 Study Design

The UP state assembly elections were held in April-May 2007. During March and April 2007, i.e. over the two months leading up to the election, we conducted two voter campaigns in three districts of central UP – Lalitpur, Bahraich and Sitapur. The caste campaign was conducted in the eighteen jurisdictions in these districts while the corruption campaign occurred in 14 of these jurisdictions.[6] Below, we describe our sampling and empirical strategies.

[6] Our corruption intervention excluded four jurisdictions (Behta, Charda, Laharpur, and Nanpara) where the four major party candidates received similar corruption rankings from journalists (on a 1-10 scale).

4.1 Sampling Strategy

Prior to sampling villages we used the polling station coverage maps (generated by the election commission) to conduct two mapping exercises. First, within each jurisdiction we generated clusters of proximate polling stations which covered, on average, fifteen villages (the range is 12 to 20 villages).[7] Second, within each cluster, we identified polling stations which either served the same village or very nearby villages; we defined these as *village units*. The average cluster had XX village units, and the average village unit had 1.585 polling stations and 1.75 villages.

Our sampling strategy proceeded as follows: first, we stratified the sample of clusters by jurisdiction.Within each jurisdiction, we randomly chose a (predetermined number of) clusters for the caste and corruption interventions.This gave us a sample frame of 36 caste clusters and 32 corruption clusters. Within a cluster, we then defined our sampling frame. Here, the unit of observation was a polling station, unless one polling station served multiple villages. In those cases it was the village. We term this unit of observation as the *randomization* unit.

Within each cluster, we generated and assigned each randomization unit a random number. Within a caste cluster we then ordered *randomization* units in ascending value of the random numbers.

For corruption clusters we followed a slightly different procedure. This was prompted by our desire to implement different corruption treatments in villages where the incumbent had spent money from his discretionary fund (called the MLA fund). These MLA fund village units, we felt, might be particularly interesting as a site for the corruption treatment, since it is often alleged that the MLA fund money is misused. We therefore stratified our treatment by MLA fund status. Within each corruption cluster, we first sorted villages by whether the incumbent had reported spending any money from the MLA fund in that village. XX of the ZZ clusters had at least one village where such spending had been reported. To do this, we first ordered all randomization units

[7] Each polling station is supposed to cover 1500 electors who are no more than two kilometers from the polling station. The average polling station in our sample covered 1.38 villages. We dropped clusters with less that 12 villages.

where some MLA fund spending had occurred in ascending order of their random number. We then did the same for non-MLA fund unit clusters.

While randomization units were ordered by their random number, this number did not perfectly determine a unit's treatment status. Rather, concerns of spillovers in nearby units (especially when one village was served by two polling stations) led the field team to adopt the following decision rule: conditional on a randomization unit being selected for treatment, the entire village unit associated with that randomization unit was designated for treatment. In other words, all the other randomization units associated with the village unit associated with the original randomization unit get pulled into the treatment. Within a cluster the randomization unit with the lowest randomization number (and any other randomization units associated with the same village unit) was assigned to treatment. This procedure was then repeated for the unit with the next lowest randomization number that had not already been pulled into treatment. This procedure was repeated until six village units were selected for treatment within each cluster.

This decision rule, however, may have led to non-random assignment. Specifically, village units associated with a larger number of randomization units have a higher probability of selection into treatment (as they had more entries in the data-set) . To address this concern we implemented a Instrumental Variable Strategy which made use of the original random number assignment. [8]

4.1.1 Instrumental Variable Strategy

The different stratification rules for treatment followed by the caste and corruption intervention led us to implement slightly different IV strategies for the two cases (full details in Appendix). However, the basic spirit of the IV strategy is very similar, and for expositional ease we describe

[8] Finally, once the campaign began it became clear that the NGOs did not have the manpower to cover all treatment units. We therefore, dropped villages and polling stations symmetrically from both control and treatment groups. The specific drop rules are in the Appendix – essentially, we dropped very large villages and polling stations which covered a relatively large number of villages. This drop rule partially accounted for the bias in selection, since all villages with multiple polling stations were dropped. However, it was also the case that the drop rule led to a larger number of polling stations and villages being dropped from the treatment group (treatment units were more likely to have the characteristics targeted by the drop rule, such as being small villages).

the IV procedure for the caste intervention below.

We take the original randomization file (before the drop rule is applied) and within each cluster assign an indicator *assigned to treatment* that equals 1 if and only if the randomization unit belongs to one of the first six randomization units in the original randomization file (which includes the pre-dropped observations). In the electoral data we then assigned all remaining randomization units as controls. Figure 1 shows the relationship between assigned to treatment and actual treatment.

We also evaluate our caste campaign using survey data. The survey only occurred in a subset of villages in a cluster – specifically, the first twelve village units in a cluster. For this sample we identified the *assigned to treatment* variable as above. We assigned the randomization units with the next six lowest randomization numbers as the control units and discarded the rest of the sample.

In both cases we applied the drop rule after this, and used *assigned to treatment* as an instrument for treatment in the sample with the drop rule applied.

4.2 Data

Three data sources were used for this study. Polling station-level data on voter registration, turnout and candidate vote shares come from the Election Commission of India. As background, all Indians above the age of eighteen and resident in the state are eligible to vote in state elections. Voting is not mandatory, and in order to vote an individual's name should be listed on the voter rolls. Voter rolls are generally verified by election commission officers who visit residential areas and verify the presence of eligible voters at individual locations and residences. The rolls are generally updated until about ten days before the election, when they are frozen and no more registration is permitted. Our second source of data on voter behavior is survey-based, and was collected for treatment villages and a subset of the control villages covered by the caste campaign. In these villages a household survey conducted in the X week between the election and when the results are declared. The survey was conducted, as part of a larger state-wide election survey, by a well known

Indian research organization that specializes in election related surveys.[9]

In each caste cluster, all villages which received the caste treatment and villages falling in the next six village units were surveyed. Twelve households were surveyed per village, where we stratified by gender and sought to survey an equal number of male and female respondents. The survey recorded demographic and socio-economic characteristics of the respondent, including age, gender, education, caste, income and occupation. To get her to reveal her voting choice, a respondent received a ballot sheet and was asked to (privately) mark her choice and put in a ballot box (provided by the surveyor). The ballot sheet identifier allowed a respondent's voting choice to be linked to her socio-economic information. An important innovation of the survey was the use of a vignette in which the respondent was asked which of two hypothetical candidates from different parties she would prefer to vote for: an incumbent widely believed to misappropriate public funds, and a challenger with no political experience running on an anti-corruption platform. The surveyors assigned each vignette candidate to one of the three most popular parties in UP: BJP, BSP, or SP. The vignette text is provided in the Appendix. The likelihood of assigning a respondent an incumbent from the party which represented her ethnic group was the same in both treatment and control units.

Ex post, we observe that the survey company did not sample randomly within the village. Rather, since their typical goal is to predict electoral outcomes, they oversampled likely voters. Reported turnout among those surveyed was one and half times higher than actual turnout (79% versus 53%). The survey data, therefore, will not provide a reliable estimate of turnout effects. On the other hand the survey seems to have been effective in getting individuals to reveal who they voted for. For the four major parties the survey predictions of vote share were within three percentage points of the right number. [10]

[9] This organization, the Center for Study of Developing Societies (CSDS) is one of the best-known polling organizations in India and has conducted numerous election-related surveys, often on behalf of some media organization that wants to predict the election outcome.

[10] Electoral vote shares of BJP, BSP, INC and SP, were 12%, 43%, 5% and 30% of the vote, while the survey vote shares were 11%, 40%, 7% and 28%.

Our third data set provides measures of politician quality. Since 2004, candidates are required by law to file an affidavit before the election where they declare, among other things, their criminal record. This is public information and is available from the Election Commission website. We collated these data for the candidates from the four main parties–BJP, BSP, INC and SP in each of the sampled jurisdictions. A quarter of these candidates (15.37%) face violent crime charges (murder, rape, kidnapping and armed robbery). In comparison, only 0.015% of citizens in these three districts faced similar charges. In other words, candidates in the 2007 UP elections were over a 1000 times more likely to have committed a violent crime than the average citizen in these districts. Figures 2 and 3 provide a graphical representation of the types of crime charges faced by the the 13 criminals with heinous crime charges in our sample.

To supplement these data, we interviewed journalists to obtain information on politician's perceived corruption record (Banerjee and Pande, 2009 discuss this methodology in more detail). From the list of prominent journalists who were covering the election in each jurisdiction we randomly selected three. Each journalist was interviewed about candidates from all four major parties, giving us three reports per politician. These interviews were conducted roughly X days prior to the election. On a scale of 1-10, the average politician was ranked as having a corruption rank of 5. In Table 1 we use these interview data to describe some salient candidate characteristics for the 70 candidates in our sample jurisdictions, broken down by the criminal antecedents of the candidates.

4.3 Randomization Check and First Stage

We start by providing a randomization check. For both the caste and corruption campaigns we check balance on a number of village level statistics, obtained from the 2001 census.[11] In addition, for the caste campaign we examine whether household survey data is balanced along a set of demographic characteristics. Tables 2a and 2b provide the results. For both campaign samples, no census co-

[11] Unfortunately, data on electoral outcomes in the previous election cannot be easily matched to polling stations in 2007 as there was a new delimitation of polling stations. Therefore, we cannot check balance in terms of previous electoral outcomes

variates are individually significant (Table 2a). The same is true for the household survey sample for the caste intervention (Table 2b).

In Table 3 we provide the first stage regressions for our IV estimates. In columns (1) and (2) we report estimates using the polling station sample. The unit of observation is defined at the level of the randomization unit.[12] The empirical specification for the caste sample is of the form

$$t_{ij} = \alpha_j + \beta T_{ij} + \epsilon_{ij} \tag{1}$$

where i is the randomization unit and j denotes jurisdiction. The outcome is t_{ij} which is an indicator for whether the polling station was treated by the NGO. The regressions include a jurisdiction fixed effect (α_j) and T_{ij} is an indicator variable which equals one if the unit was among the first six randomization units in the cluster. Standard errors are clustered by polling station. Table 3, column (1) provides the results and we see that assignment to treatment is significantly correlated with being treated. Roughly 30% of polling stations that were not assigned to treatment were also treated for reasons described above. Figure 1 provides a graphical illustration.

For the corruption sample, the regression specification needs to account for the additional stratification (within a cluster) by whether the village had received any money from the MLA fund. The specification, therefore, takes the form

$$t_{icj} = \alpha_{cj} + \delta_1 M_{icj} + \delta_2 M_{icj} \times \alpha_{cj} + \beta T_{icj} + \epsilon_{icj} \tag{2}$$

The outcome once again is actual treatment status of the polling station. Since stratification occurred within a cluster, we include a cluster fixed effect (c denotes cluster), a dummy for whether the village received MLA funds and the interaction of the cluster fixed effect with this dummy. Therefore, in essence we are only comparing units within a cluster which have the same MLA fund status. Standard errors are clustered by polling station. Table 3, column (2) shows that assignment

[12] A polling station enters the data-set twice if it covers two randomization units. This typically occurs when two villages share the same polling station.

to treatment is a very significant predictor of having been treated.

In column (3) we provide the first stage regressions for the household survey sample. These data are only available for the caste intervention. The regression specification is similar to equation (1). The main difference is that the unit of observation is a survey respondent, and the treatment variable t_{ij} refers to whether she lives in a village which was treated by the NGO. Similarly, the assignment to treatment variable refers to whether her village was among the first six randomization units. In this, and all subsequent, regressions we restrict the sample to Hindu respondents (column (3)).

5 Campaign

Our campaigns were designed in conjunction with, and implemented by, a local NGO partner with experience in conducting rural village-level campaigns in UP, Sarathi. A two-person Sarathi team (composed of one man and one woman) spent one day in each treatment village to deliver the campaign messages. A typical day started at 10 am and ended in the evening. Table 4 describes the entire day schedule for the caste and corruption treatments.

There were three main ways in which the campaign message was delivered in the village: meetings, posters and a puppet show. A recurrent theme throughout the day was that the representation of the community and being responsiveness to its development needs were in the legislator's job descriptions, and that voting works to hold politicians accountable. Early in the day, villagers were encouraged to identify and discuss the issues that matter to them. Field notes show that the top three issues identified by men were roads, political issues and sewage/toilet. For women, the top three issues were pensions, toilets/sewage and health. The core campaign message was delivered during the puppet show and reiterated in the evening after the puppet show. The scripts of the puppet show are summarized in the Appendix. Examples of the caste and corruption posters are provided in Appendix Figures 1 and 2.

For the caste campaign the core message was *Development issues affect everyone in your*

village, not just the members of one caste. Vote on development issues rather than along caste lines.

For the corruption campaign the message was *Corrupt politician steal money set aside for development funds and do nothing for you. Vote for clean politicians that care about your development needs.*

An additional element of intervention occurred in villages where the incumbent had spent some of his MLA fund. In these villages, the campaign team was supposed to read out a list of the MLA fund projects that had been completed.

Table 5 uses survey data (only available for the caste campaign) to provide evidence on exposure of villagers to the treatment. The second stage equation for the IV regressions is

$$y_{ij} = \alpha_j + \beta t_{ij} + \epsilon_{ij} \tag{3}$$

where y_{ij} is the outcome of interest (attendance in NGO meetings, turnout, vote share of candidates with particular characteristics, etc.). To get the IV estimates $t_{ij}(t_{icj})$ is instrumented by $T_{ij}(t_{ij})$. The first stage regression is given by column (3), Table 3.

Column (1) of Table 5 shows that treatment and control villages were as likely to receive party campaigning. Overall, the level of party campaigning is high, with 87% of respondents in the control sample reporting that parties had campaigned in their village. Exposure to a NGO campaign is significantly lower at 10%. Respondents in treated villages were 20% more likely to report seeing a NGO campaign. Increased exposure to NGO campaigns does not translate into a difference in the number of speeches, discussions on political campaigns, slogan chanting or children rallies. We interpret this as suggesting that these features were common to both NGO and party campaigns. However, respondents in treated villages are significantly more likely to state that they saw posters being put up. The largest difference is on puppet shows - respondents in treated villages were 16% more likely to observe a puppet show.

6 Did the Caste Campaign Change Behavior?

We start by examining the impact of the campaign on voter turnout. Given the twin concerns that the survey data was not representative of the entire population *and* that voters may respond to the treatment not by changing voting behavior but rather by simply changing what they report, we focus on electoral data. The results are in Table 6.

In column (1) we evaluate the campaign impact of voter registration. As our campaign occurred while voter registration was still ongoing, the campaign may have mobilized voter efforts to ensure that their names feature on the electoral rolls. In column (1) the dependent variable is log registered voters. Overall, we observe no significant effect of the campaign on voter registration. In columns (2) and (3) we look for evidence of gender differential effects. While we cannot reject the effect being the same across both gender, we observe a more precisely estimated registration effect of 7% for male voters (significant at 10%). In columns (4)-(6), we evaluate voter turnout. Here, the dependent variable is log voters. Column (4) suggests a large effect of the campaign on turnout – overall, the campaign increased turnout by 9%. Once again, when we examine the effects separately for men and women we observe a larger and more precisely estimated effect of 11% for men. That said, we cannot reject the effects being equal sized across the two genders.

Next, we turn to evaluating the central question of whether the caste campaign influenced the electoral salience of caste. Here, we have to rely on the survey data where we can identify the caste of the respondent and have information on the party she voted for. We restrict the sample to Hindu respondents and use as the dependent variable a dummy which equals one if the respondent voted for the party which represents her caste-preferred party. The regressions control for the respondent's caste (specifically, whether she belongs to Scheduled Caste (SC) or Other Backward Caste (OBC)).

In column (1) of Table 7a we start by examining whether the caste campaign affected respondent's voting preferences. In the control sample, over 57% of respondents state that they voted for the party that represented their caste group. Overall, we see a noisily estimated reduction in the likelihood that a respondent voted for her caste-preferred party. In columns (2) and (3) we

disaggregate this into lower and upper castes. We observe a significant reduction in the propensity to vote for the caste-preferred party only among the lower castes.

Next we examine whether an individual's responsiveness to the campaign is mediated by the quality of the candidate from her caste-preferred party. We consider two measures of quality – whether the caste-preferred candidate is charged with a heinous crime and whether the caste-preferred candidate is ranked as the most corrupt candidate in the jurisdiction.

Table 7a, column (4) shows that respondents are willing to move away from the caste-preferred party only when the candidate for their caste-preferred party has been charged with a heinous crimes. We find that the entire reduction in caste-preferred voting comes from a move away from voting for candidates who have a heinous crime charge. In column (5) we show this result holds for the low castes. To check against the concern that this is a mechanical effect driven by the largest population group's preferred party in column (6) we control for a differential treatment effect among OBCs and find a similar effect. In column (7) we consider upper castes and see no such effect. Columns (8)-(10) consider as the quality measure whether the caste-preferred candidate was the most corrupt. We observe an insignificant effect.

A concern with using survey data is the veracity of responses: respondents in treated villages may have changed the response they give surveyors without changing their actual voting behavior in the polling booth. In the electoral data we cannot directly examine the decline in caste-preferred voting, as party vote shares are not disaggregated by voter caste. We therefore focus on checking whether the vote share of candidates charged with heinous crimes is lower in treated villages. In Table 7b we see that a candidate who face heinous charges has a lower vote share in treatment areas. This effect is absent when we measure quality by whether the candidate is the most corrupt.

In Table 8 we provide ancillary evidence from the vignette fielded during the survey. The full script of the vignette is provided in the Appendix. A respondent was asked to state her preference between a corrupt incumbent who was also known for local development and a relatively unknown challenger who spoke up against corruption. We varied the party assigned to the incumbent and the party assigned to the challenger across respondents. Column (1) shows that the overall willingness

to vote for the incumbent is unaffected by the treatment. In column (2) we examine whether this is affected by the party assigned to the incumbent and challenger. We observe significant caste preference – a respondent is much more likely to vote for the candidate (either incumbent or challenger) who belongs to his/her caste-preferred party. Overall, we see no significant treatment effect. However, once we consider low caste respondents we see that the treatment reduced caste preference. These results are consistent with the survey data on actual voting preferences.

7 Did the Corruption Campaign Change Behavior?

Next we turn to the corruption campaign. Here we always examine whether the treatment effect differs across villages where the incumbent spent money from the MLA fund and other villages. The results are in Table 9. We observe that the campaign had no impact on voter registrations (columns (1)-(3)). We also observe no significant impacts on voter turnout columns (4)-(6).

In Table 10 we examine whether the corruption campaign altered the vote share of candidates. We measure candidate quality in two ways: by the corruption rank, which was assigned to them by journalists, and by whether the candidate had a criminal record. In neither case do we observe any impact of the campaign.

8 Vignette Treatment

TBA

9 Conclusion

TBA

10 Appendix

10.1 Vignette Description

Investigator flag: Randomly select and circle one of the following combinations: [1: Incumbent (SP) Opponent (BJP)

2: Incumbent (BSP) Opponent (BJP)

3: Incumbent (BJP) Opponent (SP)

4: Incumbent (BSP) Opponent (SP)

5: Incumbent (BJP) Opponent (BSP)

6: Incumbent (SP) Opponent (BSP)]

Use selected combination in the question given below): As part of our survey, we have been talking with people from all over Uttar Pradesh, and we'd like to get you opinion about some candidates in other parts of UP, so imagine that you lived in a different constituency, far from here. Consider two candidates in this constituency. The incumbent candidate belongs to party [ ]. He is a forceful advocate for economic development and has secured government funding for local projects, but he is believed to divert public funds in order help his friends and family. His main opponent belongs to party [ ], who speaks out against corruption but otherwise has no experience in politics. If you were a voter in that district, which of the two candidates would you be more likely to support? (Record the name of party).

10.2 Brief Synopses of Puppet Shows

10.2.1 Caste Puppet Show

This puppet show encourages the audience to prioritize local infrastructure and development needs when casting their ballots rather than vote strictly on the basis of traditional ethnic affiliations. This is accomplished by juxtaposing a "good" MLA, who wins elections by appealing to development issues, with a "bad" MLA, who wins elections by appealing to caste allegiances. The puppet show

also reminds the audience that the new voting process is entirely secret, and so they should not fear retribution from repressive candidates or their "goons."

The production opens with a brief comic act that captures the audience's interest and exhorts them to consider development issues in their village during the puppet show and when they cast their vote. The first scene of the main act introduces Chandar Bhaiya Jindabad, a politician from a nearby village who is noted for his honesty and his commitment to local development issues. These traits are emphasized in a song between scenes that urges the audience to, "elect a politician who will build roads, an electrical grid and a sanitation system, who will never make excuses, and who will never break promises." Chandar Bhaiya is challenging the incumbent MLA of Tendwa village, Lalan, a corrupt and overconfident politician who has failed to deliver on his promises of local infrastructural improvement. Lalan is introduced in Scene 3 when he bribes a village elder, Bharose Lal, to arrange a meeting with members of their caste in order to secure their votes in the upcoming election. His dishonest character is highlighted in the following scene while discussing strategy with a supporter when he declares, "I will win this election by hook or by crook. Old strategies are still working."

Unfortunately for Lalan, Bharose Lal runs into Ramua, who has recently arrived from a nearby community that has prospered under the leadership of a development-focused MLA. Ramua gives Bharose Lal and other villagers the courage to vote out of a concern for the development needs of their village rather than out of caste-based loyalties or fear. The final scene shows a regretful Lalan who, after losing the race, ruefully acknowledges that "in order to win an election now, we must work for the village." The production closes with a song that restates major themes in the script, including a warning to the audience: in a democratic election, voters are ultimately responsible for outcomes, for better or for worse.

Throughout the script, the audience is encouraged to trust the voting process and not to fear retributive candidates, because "If you (villagers) join together, nobody will be able to stop you." Additionally, the first and last scenes remind the audience that since the new voting process relies on push-button machines rather than stamped ballots, votes cannot be traced back to voters, and

so they do not need to fear retributive politicians.

10.2.2 Corruption Puppet Show

This puppet show encourages the audience to be politically informed and to hold politicians accountable for the promises they make during campaigns and the actions they take while in office. This is accomplished by defining corruption, giving an example of a corrupt official, and outlining the steps that the audience can take to avoid electing crooked MLAs. The puppet show also reminds the audience that the new voting process is entirely secret, and so they should not fear retribution from repressive candidates or their "goons."

The main act opens with a song that presents major themes and warns the audience to "be prepared for the coming election." The first scene introduces Bharose Lal, who is described to the audience as "a villager like you, in a village like yours," and who is initially naive about political corruption but is gradually informed through the course of the play. Bharose Lal's village suffers under the ineffectual leadership of a corrupt MLA, who has not returned to the village since winning the election. After running into fellow villagers Rahim, Radhey, and Pandit Ramkhelawan, Bharose Lal calls a meeting to discuss Radhey's recent trip to the capital (Lucknow) to meet with their MLA.

Radhey reports how he extended an invitation to the MLA to visit their village but was, unsurprisingly, turned down. He further describes how the MLA, though as impoverished as other villagers prior to the election, is now very wealthy from conducting shady business dealings, taking bribes, maintaining an active role in the black market, and pocketing funds slated for the village's development needs. The contrast between the extravagant lifestyle of the MLA and the destitute lifestyles of the villagers he represents is highlighted in a song between scenes. The villagers then discuss the direct implications that this behavior has had for their village: an undersupply of grain and kerosene and shoddy workmanship on a school that led to the collapse of a roof and the near-death of several students. Rahim declares that such corrupt leaders are cruelly "playing with the lives of the poor."

In the last scene, the villagers consider practical measures for avoiding electing corrupt politicians. Their final strategy consists of the MLA taking a sacred oath (using water from the Ganges) to guarantee honest motives, followed by constant monitoring of the MLA's actions by the villagers once he is elected. The villagers promise that, "we are not going to sleep" to ensure that they always hold politicians accountable. The script even suggests that the best MLAs may not share the same religious or ethnic background as the majority of voters: when Radhey describes his plan to use water from the Ganges to hold the MLA accountable, Rahim points out that the candidate could be Muslim, in which case they should use a Qur'an for the oath-taking process. The show concludes with a battle-cry against crooked politics: "We have the right to be informed. This is our weapon against corruption."

References

Banerjee, A. and R. Pande (2009). Parochial politics: Ethnic preferences and politician corruption. mimeo.

Besley, T., D. Sturm, and T. Persson (2008). Political Competition and Economic Performance: Theory and Evidence from the United States. mimeo, LSE.

Chandra, K. (2004). *Why Ethnic Parties Succeed: Patronage and Ethnic Headcounts in India*. Cambridge: Cambridge University Press.

Dickson, E. and K. Scheve (2006). Social identity, political speech, and electoral competition. *Journal of Theoretical Politics 18*(1), 5–39.

Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice 30*, 5–25.

Ferree, K. (2004). The Micro-Foundations of Ethnic Voting: Evidence from South Africa. *Afrobarometer*.

Gerber, A. and D. Green (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review 94*(3), 653–663.

Gerber, A., D. Green, and C. Larimer (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review 102*(1), 33–48.

Horowitz, D. L. (1985). *Ethnic Groups in Conflict*. Berkeley: University of California Press.

Jaffrelot, C. (2003). *India's Silent Revolution: The Rise of the Lower Castes in North India*. London: Hurst and Company.

Key, V. O. (1949). *Southern Politics in State and Nation*. New York: A.A. Knopf.

Krasno, J. and D. Green (1988). Preempting quality challengers in house elections. *Journal of Politics 50*, 920–936.

Lupia, A. (1994). The effect of information on voting behavior and electoral outcomes: An experimental study of direct legislation. *Public Choice 78*, 65–86.

Miquel, G. P. (2006). The control of politicians in divided societies: The politics of fear. mimeo.

Popkin, S. (1991). *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: University of Chicago Press.

Vicente, P. and L. Wantchekon (2009). Clientelism and Vote Buying: Lessons from Field Experiments in African Elections. *Oxford Review of Economic Policy*. Forthcoming.

Wantchekon, L. (2003). Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin. *World Politics 55*(3), 399–422.

Yadav, Y. (2000). *Understanding the Second Democratic Upsurge: Trends of Bahujan Participation in Electoral Politics in the 1990s*. Transforming India: Social and Political Dynamics of Democracy.

Table 1: Characteristics of Candidates in Four Major Parties by Criminality

				Two-Group Mean-Comparison Test
	All	Non-Criminals	Criminals	Difference
	(1)	(2)	(3)	(4)
Corruption Rank (Normalized Scale, 1-7)	3.508	3.351	3.969	0.618
				(0.322)
				[0.059]
Candidates with Family Improvement (%)	46.032	38.298	68.750	30.452
				(14.133)
				[0.035]
Candidates with Personal Gain (%)	44.444	38.298	62.500	24.202
				(14.284)
				[0.095]
Gross Assets Held by Candidates (Average, in Rs)	6011947	4569233	10509820	5940589
				(2934252)
				[0.047]
Net Assets Held by Candidates (Average, in Rs)	5567134	4174953	9907462	5732509
				(2898515)
				[0.052]
Candidates with a College Degree (%)	41.667	49.057	23.529	-25.527
				(13.648)
				[0.066]
Observations	Heinous Charges	70	57	13
	All Charges	70	53	17

Notes

1. The category Corruption Rank is calculated from the mean response of three journalists reporting on the corruption of the candidate using a normalized scale from 1 (least corrupt) to 7 (most corrupt), where values are normalized by rankings of three hypothetical candidates. The category Candidates with Family Improvement is a dummy variable calculated from three journalist reponses to whether the candidate's family's economic status improved after the election; the dummy=1 if all three journalists agreed that improvement occurred. The category Candidates with Personal Gain is calculated in a similar way. The category Assets has two variables, Gross and Net, as reported by Election Watch (EW), and Candidates with a College Degree is a dummy variable=1 if the candidate reported having a college degree during an interview.

2. Column (1) reports summary statistics for all observations, whereas columns (2)-(3) report results separately for criminal and non-criminal candidates. Criminality was determined by whether or not the candidate had a criminal record, as reported by EW. Column (4) gives the results of two-group mean-comparison test to determine if there is a significant difference between criminals and non-criminals for each characteristic Standard errors are given in parentheses, and p-values are given in square brackets.

Table 2a: Randomization Check, Village-Level Data

	Caste Experiment			**Corruption Experiment**		
	Average		Diff (1) and (2): p-value	Average		Diff (4) and (5) :p-value
	Assigned to Control	Assigned to Treatment		Assigned to Control	Assigned to Treatment	
	(1)	(2)	(3)	(4)	(5)	(6)
Land Area (hectares)	402.26	430.34	0.322	380.58	447.32	0.351
	(313.80)	(324.00)		(405.65)	(375.89)	
Total Households	265.83	283.38	0.141	260.31	250.82	0.537
	(153.09)	(137.86)		(136.16)	(149.95)	
Total Males	854.03	901.97	0.182	840.70	816.85	0.592
	(495.54)	(432.25)		(435.95)	(490.63)	
Total Females	737.08	775.86	0.213	721.24	707.40	0.662
	(427.28)	(376.86)		(376.80)	(419.48)	
Scheduled Caste (SC) Population Share	0.33	0.32	0.265	0.28	0.27	0.380
	(0.17)	(0.16)		(0.18)	(0.17)	
Male Literacy Rate	0.46	0.47	0.553	0.44	0.45	0.259
	(0.13)	(0.13)		(0.11)	(0.13)	
Female Literacy Rate	0.23	0.23	0.667	0.20	0.21	0.210
	(0.11)	(0.11)		(0.09)	(0.09)	
Public Good Index	2.65	2.57	0.586	2.47	2.27	0.170
	(1.60)	(1.48)		(1.45)	(1.44)	
Village received MLA fund	0.05	0.06	0.669			
	(0.22)	(0.23)				
Joint Significance Test			0.160			0.820
Observations	497	104		427	102	

Note:

1. The unit of observation is the randomization unit (i.e. minimum of village and polling station). The village data are from the 2001 census: the Public Good index is the equally weighted average of schools, medical facilities and banks in the village. The MLA fund data is from XX.

2. Columns (1), (2) and (4),(5) report means with standard deviations in parentheses. Columns (3) and (6) report p-values of tests of differences in means across columns (1) and (2) and columns (4) and (5) respectively. Column (3) tests are based on regressions with constituency fixed effects and column (6) tests are based on regressions with fixed effects and each combination of clustercode and MLA fund status. Standard errors are clustered by polling stations.

Table 2b: Randomization Check, Household Data

	Caste Experiment		
	Average		
	Assigned to Control	Assigned to Treatment	Diff (1) and (2): p-value
	(1)	(2)	(3)
Female	0.408	0.428	0.420
	(0.49)	(0.50)	
Scheduled Caste (SC)	0.251	0.259	0.910
	(0.43)	(0.44)	
Other Backward Castes (OBC)	0.471	0.457	0.753
	(0.50)	(0.50)	
Male Literates	0.334	0.336	0.965
	(0.47)	(0.47)	
Female Literates	0.117	0.125	0.841
	(0.32)	(0.33)	
Low income (less than Rs 2000 per month)	0.264	0.273	0.625
	(0.44)	(0.45)	
Very low income (less than Rs 1000 per month)	0.48	0.45	0.395
	(0.50)	(0.50)	
Below 36 years	0.486	0.495	0.403
	(0.50)	(0.50)	
Joint Significance Test			0.700
Observations	960	791	
Received vignette where incumbent is	(0.42)	(0.41)	0.500
from caste preferred party	0.49	0.490	
Received vignette where opponent is	0.77	0.750	0.310
from caste preferred party	(0.42)	(0.43)	
Observations	796	684	

Notes

1. See the Appendix for a description of the vignette text.

2. Household data is from CSDS. Standard errors are clustered by randomization unit.

Table 3: First Stage Results

	Dependent Variable: Treated		
	Electoral Data		
Sample	Caste	Corruption	Caste (Hindu respondents)
	(1)	(2)	(3)
Assigned to Treatment	0.718***	0.762***	0.633***
	(0.029)	(0.030)	(0.053)
Treatment*MLA fund		0.038	
		(0.156)	
MLA fund		-0.102	
		(0.206)	
F-test		0.23	
		[0.63]	
Control mean	0.264	0.248	0.330
Observations	601	529	1538

Notes

1. The dependent variable Treated is an indicator variable=1 if the NGO campaigned in that village. Assigned to Treatment=1 if the village was among the first six villages selected by randomized assigment. MLA fund=1 if the

2. Columns (1) and (2) use data at the polling station level. Column (3) uses survey data, where the dependent variable is whether the NGO campaigned in the respondent's village.

3. Standard errors are clustered by polling station. Column (1) and (3) regressions include constituency fixed effects, and the column (2) regression includes clustercode*MLA fund fixed effects. The column (3) regression also includes (i) dummies for whether the respondent is SC or OBC and (ii) a dummy for being female and its interaction with treatment.

4.The F-test tests whether the effect in MLA fund villages differs from that in non-MLA fund villages.

Table 4: Caste and Corruption Campaigns

Time of day	Process	
Morning	Village mapping	Conducted with villagers to identify lower caste areas and to structure the day.
	Door-to-door visits in each ward	Informed about meetings and delivered message to elderly, sick, and marginalized. Door-to-door drive was conducted in each ward directly prior to meeting.
	Posters	Placed in public areas around village.
	Ward-wise meetings (3 on average)	Warm up meeting to identify village problems and discuss the importance of voting.
Afternoon	Children's rally	Children were taught the campaign message and led on a rally where they chanted slogans.
	Women’s meeting	Meetings with pre-existing self-help groups and all village women.
	Village-level meeting	Main meeting in which the central message was delivered.
	Wall-writing	Wall-writing with villagers of campaign slogans.
Early Evening	Puppet show	Trained puppeteers delivered the message in dramatic format.
	After-puppet-show meeting	Follow-up and summary of the day's message.
Night	Meetings	Meeting with those that worked outside the village during the day.

Table 5: Exposure to NGO Campaign: Survey Data from Caste Campaign

				Self-Reported Knowledge of Elements of NGO Campaign					
	Party Campaign	NGO Campaign	Campaign Exposure Index	Speeches	Political Issue Discussions	Chanting Slogans	Children's Rally	Posters/Wall Writing	Puppet Shows
Sample: Hindu respondents	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treated	0.007	0.206***	0.079*	0.039	0.048	0.071	0.066	0.087*	0.161***
	(0.036)	(0.062)	(0.043)	(0.046)	(0.041)	(0.044)	(0.047)	(0.051)	(0.060)
Control Mean	0.870	0.100	0.080	0.037	0.026	0.024	0.006	0.170	0.270
Observations	1538	1538	1538	1538	1538	1538	1538	1538	1538

Notes

1. The regressions use CSDS survey data. Each column represents a separate regression. Party campaign (NGO campaign)=1 if respondent states political parties campaigned in her village. Dependent variables in column (4)-(9) are indicator variables. An indicator variable=1 if the respondent states that she attended the NGO campaign and the specified activity occurred as part of the campaign. The campaign exposure index is the equally weighted index of indicator variables defined in columns (4)-(9).

2.We report IV regressions, where the first stage is reported in Table 2, column (3). All regressions include (i) constituency fixed effects (ii) SC and OBC dummies and (iii) female dummy and its interaction with treatment. Standard errors are clustered by polling station.

3. The Control Mean is the average for the randomization units not assigned to treatment.

Table 6: Caste Campaign and Voter Turnout: Electoral Data

	Log Registered			Log Votes		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.045	0.070*	0.036	0.092*	0.111**	0.070
	(0.039)	(0.037)	(0.043)	(0.049)	(0.048)	(0.057)
Observations	601	601	601	601	601	601

Notes

1.The regressions use polling station electoral return data. Each column represents a separate regression. Columns (1) and (4) report regression results for all observations, whereas columns (2), (3), (5), and (6) report regression results separately by gender. The dependent variable in columns (1)-(3) is log (number of registered voters) and in columns (4)-(6) is log (number of valid votes).

2. In all cases we report IV regressions where treatment is instrumented by assignment to treatment (Table 2, column (1) provides the first stage). All regressions include constituency fixed effect. Standard errors are clustered by polling station.

Table 7a: Caste Campaign and Voting Preferences: Survey data

	Dependent Variable: Voted for Candidate belonging to Caste-Preferred Party									
Quality measure				Caste-Preferred Candidate faces heinous crime charge				Caste-Preferred Candidate ranked as most corrupt		
	All Hindus	Lower Castes	Upper Castes	All Hindus	Lower Castes	Lower Castes	Upper Castes	All Hindus	Lower Castes	Upper Castes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treated	-0.065	-0.120*	0.043	0.009	-0.022	-0.023	-0.012	-0.002	-0.046	0.056
	(0.062)	(0.068)	(0.120)	(0.070)	(0.076)	(0.093)	(0.141)	(0.077)	(0.080)	(0.133)
Treated*Quality measure				-0.177*	-0.223**	-0.224**	0.045	-0.132	-0.138	-0.129
				(0.091)	(0.099)	(0.105)	(0.209)	(0.101)	(0.105)	(0.221)
Quality measure				0.274***	0.237***	0.237***	0.456**	0.159**	0.085	0.235
				(0.075)	(0.084)	(0.085)	(0.178)	(0.078)	(0.080)	(0.241)
OBC*Treated	N	N	N	N	N	Y	N	N	N	N
Observations	1538	1171	367	1538	1171	1171	367	1538	1171	367

Notes

1. The regressions use CSDS survey data for Hindu respondents. Each column represents a separate IV regression, where the first stage is as given in column (3) of Table 3. The dependent variable is an indicator variable=1 if the respondent states that she voted for the party which represents her caste (the control mean is 0.576). All regressions include (i) constituency fixed effects (ii) SC and OBC dummies (iii) female dummy and female dummy interacted with treatment. Standard errors are clustered by randomization unit.

2. Three sets of regressions are reported. Columns (1)-(3) report results from regressions of the dependent variable on the previously-defined variable Treated. The regressions in columns (4)-(7) include a quality measure=1 if the caste-preferred candidate faces a heinous crime charge, as well as an interaction variable between Treated and that quality measure. The regressions in columns (8)-(10) use a different quality measure=1 if the caste-preferred candidate is ranked as the most corrupt candidate by EW, as well as an interaction variable between Treated and that quality measure.

Table 7b: Caste Campaign and Voting Preferences: Electoral Data

	Dependent Variable: Log Candidate Votes	
Quality measure	Candidate has heinous crime charge	Candidate ranked most corrupt in jurisdiction
	(1)	(2)
Treated	0.128*	0.073
	(0.068)	(0.070)
Treated*Quality measure	-0.233*	-0.008
	(0.125)	(0.108)
Quality measure	0.502**	0.967***
	(0.116)	(0.097)
Observations	2293	2293

Notes

1.These IV regressions use candidate outcomes aggregated at the polling station level. The dependent variable is the log of votes received by a candidate in a polling station, and the sample is restricted to candidates of the four caste preferred parties. Treated and quality measure are as defined in previous tables.

Table 8: Party-Quality Trade-Off: Evidence from Vignettes

	Dependent Variable: Voted for Incumbent			
	All		Lower Castes	
	(1)	(2)	(3)	(4)
Treatment	-0.007	-0.040	0.012	-0.077
	(0.051)	(0.106)	(0.042)	(0.102)
Incumbent from Caste Preferred Party*Treatment		-0.092		-0.205***
		(0.080)		(0.079)
Incumbent from Caste Preferred Party		0.383***		0.468***
		(0.058)		(0.057)
Challenger from Caste Preferred Party*Treatment		0.106		0.242**
		(0.112)		(0.106)
Challenger from Caste Preferred Party		-0.141*		-0.254***
		(0.084)		(0.081)
Control Mean	0.545		0.540	
Observations	1480	1480	1128	1128

Notes

1. The regressions use CSDS survey data. Each column represents a separate regression, where the dependent variable is an indicator variable=1 if the respondent states that s/he would vote for the incumbent in the vignette. Incumbent from Caste Preferred Party is an indicator variable=1 if the incumbent in the vignette was from the party preferred by the respondent's caste (challenger from caste preferred party is similarly defined)

2. We report IV regressions where we instrument Treated by assignment to treatment (first stage is in Table 2). All regressions include dummies for the party of the incumbent in the vignette. Standard errors are clustered by randomization unit.

Table 9: Corruption Campaign and Voter Turnout: Electoral Data

Dependent Variable:	Log Registered			Log Votes		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.007	0.009	-0.002	0.052	0.039	0.097
	(0.047)	(0.048)	(0.047)	(0.049)	(0.048)	(0.060)
Treatment*MLA fund	-0.152	-0.172	-0.088	-0.251	-0.233	-0.302
	(0.133)	(0.151)	(0.123)	(0.124)	(0.132)	(0.122)
MLA fund	0.285	0.307	0.222	0.347	0.300	0.426
	(0.191)	(0.203)	(0.186)	(0.169)	(0.175)	(0.170)
F-test	0.82	0.86	0.82	0.55	0.26	0.85
	[0.36]	[0.35]	[0.36]	[0.45]	[0.60]	[0.36]
Observations	526	526	526	527	527	527

Notes

1.The regressions use polling station electoral return data. Each column represents a separate regression. The dependent variable in columns (1)-(3) is log (number of registered voters) and in columns (4)-(6) is log (number of valid votes).

2. In all cases we report IV regressions where treatment is instrumented by assignment to treatment (Table 3, column (2) provides the first stage). All regressions include clustercode*MLAfund fixed effects.

3. The F-test tests whether the effect in MLA fund villages differs from that in non MLA fund villages.

Table 10: Corruption Campaign and Voting Preferences

	Dependent Variable: Candidate Vote Share	
Quality measure	Candidate has heinous crime charge	Candidate ranked most corrupt in jurisdiction
	(1)	(2)
Treatment	0.029	0.030
	(0.081)	(0.080)
Treatment*Quality	-0.022	0.103
	(0.562)	(0.113)
Quality	0.602**	1.270**
	(0.127)	(0.067)
Treatment*Quality*MLA fund	0.114	-0.081
	(0.177)	(0.186)
Treatment*MLA fund	-0.175	-0.155
	(0.229)	(0.151)
Quality*MLA fund	0.16	0.241
	(0.523)	(0.354)
MLA fund	0.035	0.027
	(0.249)	(0.173)
F-test	0.00	0.01
	[0.99]	[0.92]
Observations	2028	2028

Notes

1.The regressions use candidate outcomes aggregated at the polling station level. The dependent variable is the log of votes received by a candidate in a polling station, and the sample is restricted to candidates of the four caste preferred parties.