Beyond Coethnicity:

Political Influence in Ethnically Diverse Societies*

Neelanjan Sircar — CPR, Delhi

Simon Chauchard — Dartmouth College

March 14, 2017

Abstract

In most emerging democracies, influential local citizens help shape the electoral choices of other voters. This article explores the extent to which these individuals ("influencers" hereafter) direct their influence towards voters from their own ethnic group vs. other groups. Contrary to common but simplistic assumptions, we argue that influence does not always flow along ethnic lines in ethnicized societies: voters are often likely to be influenced by non-coethnics. This is because diversity at the local level, if combined with the existence of a market for political influence at the local level, provides strong incentives for relatively skilled influencers ("dominant influencers" hereafter) to add non-coethnics to their network. To test this argument, we identify the main influencers across a large sample of villages in rural Bihar (India). We then rely on a cross-referencing exercise between influencers and voters to compare the networks of "dominant influencers" vs. others. Results support our argument: when they have achieved dominance in the local market for influence (denoting that they are by far the most influential individuals at the local level), "influencers" are as likely to count coethnics as non-coethnics in their networks. These findings contribute to the empirical literature on political campaigns in India, and beyond, in ethnicized polities. They also help explain why ethnic preferences do not always transform into ethnic votes: namely, because local-level networks through which many voters are mobilized and receive assistance are often multi-ethnic.

*DRAFT. Please do not circulate without authors' permission.

1 Introduction

Political scientists have long documented elites' ability to influence voters' choices. Recognizable figures who attract media coverage influence the preferences of voters through their messages and actions in the public space, for instance framing existing information into ideological stances (Katz and Lazarsfeld 1955, Zaller 1992). When parties do not display concrete policy or ideological differences, and instead simply attempt to differentiate themselves by their ability to reach different groups of voters with benefits and public goods (Kitschelt and Wilkinson 2007), influencing voters however also entails a more indirect mechanism. Elites in this case likely need to rely on actors who are in contact with voters *at the very local level* between elections in order to influence voters during campaigns.

Depending on the context in which they operate, the methods these locally influential citizens (hereafter "influencers") rely on to sway voters vary: from legitimate discussion and canvassing (Schaffer and Baker 2015) to more problematic strategies such as gift-giving (as in the case of clientelistic "brokers", as in Stokes et al 2014), or even coercion (Wilkinson 2007). But regardless of the method (or more likely, the mix of methods) they employ, these local-level actors likely constitute important sources of influence, simply because they operate in close proximity to voters.

In developing and ethnically divided societies, these local-level actors may play a critical role in exacerbating or dampening the extent of division between ethnic groups, given their central role in social and political life. In particular, if their efforts - and their influence - largely flow along ethnic lines, then division between ethnic groups should be exacerbated. On the other hand, if local influencers preside over broader, multi-ethnic networks at the local level, then ethnic divisions should be dampened – as influencers assist and mobilize citizens *across* ethnic lines. In the context on which we focus in this manuscript (North India), this is illustrated by Krishna's (2004) study of local intermediaries in rural areas. As suggested by Krishna, a new generation of intermediaries increasingly disregarding caste lines may, more generally, contribute to the erosion of caste-based identities.

In this article, we explore whether and when local influencers reinforce or dampen ethnic divisions in ethnicized polities. Beyond their potential impact on interethnic relations, the composition of local influence networks is relevant to discussions about voting behavior and ethnic politics (Chandra, 2004; Posner, 2005). Over the past decade, empirical studies have shown

that voters have a strong preference for coethnic candidates in India (Chauchard, 2016) and in other emerging democracies (Carlson, 2015; Pepinsky, Liddle and Mujani, 2012). Yet voters' decisions at the polls are only partially determined by these individual-level ethnic preferences. This is because voting behavior is also the product of coordination mechanisms and influence by local political actors directly in touch with voters. If voters are not influenced along ethnic lines at the local level, elections are less likely to resemble "census elections" (Ferree, 2011). In such cases, voters are less likely to act on their coethnic preferences, as this implies that a non-coethnic influences their decisions. Determining whether local influence is strictly organized around ethnic lines, thus, allows us to document the extent to which ethnic preferences should be expected to transform into ethnic votes.

To grapple with these issues, we develop a theoretical model of *the market* for influence at the local level in competitive democracies. In line with much qualitative evidence from competitive polities such as India (Krishna 2004, Kruks-Wisner 2018) and with our own findings (Chauchard and Sircar 2018), we posit that a multiplicity of potential actors compete at the local level to develop a personal assistance network. In a competitive democracy, multiple parties and political actors vie for the services of these influencers, since their role on the ground presumably allows them to influence citizens around elections. Influencers also seek these relationships, as they seek to trade their local influence for rents from these political actors and parties for access to them. Since multiple actors seek them after, influencers can however renegotiate their contracts with political actors and hedge between parties. Whereas a stable and committed base of party workers, or even a clientelistic "machine" fulfills this intermediation role in dominant-party systems, these influencers are thus better thought as mercenaries in the competitive context in which we operate. In a related paper (Chauchard and Sircar 2018), we show these assumptions to be credible in the Indian case, as less than 1/4 of local "influencers" emerged as committed partisan actors, and as few of these actors provided assistance or influenced voters along partisan lines.

In this context, what then determines whether these actors concentrate their efforts on their coethnics or diversify their network? We argue that the response to this question has to do with the level of competition these actors face from their peers (that is, from other influencers) at the local level. In our model, we envision a market composed of influencers with varying levels of "quality". When an influencer distinguishes himself from others by demonstrating superior quality, then he is able to use his appeal across ethnic groups to build a broad-based, multieth-

nic coalition. If, however, the market for influencers becomes highly competitive, that is, if there exist multiple influencers of similar quality at the local level, then influencers can no longer differentiate themselves simply through the quality of their work. In such a scenario, influencers fall back on coethnic appeals – which may be effective due to ethnic divisions in society – to attempt to expand their personal networks. We thus hypothesize that the greater the competition between influencers at the local level, the more influence flows along ethnic lines[1], which more generally suggests that local influence networks should not *always* be expected to be structured along ethnic lines.

We explore this hypothesis in the context of the Northern Indian state of Bihar, where much of party politics has revolved around caste for the past 30 years (Witsoe, 2013; Vaishnav, 2017). Making inferences on influence in this context is important for a number of reasons. Insofar as our Bihar sample resembles much of rural India, it thus allows us to better understand who candidates and party elites rely on to influence over 600 million rural Indian voters with extremely diverse ethnic identities. Whether or not the local networks maintained by these actors are mono-ethnic is in addition *substantively* important, since it tells us about the strength of caste boundaries and about the ability of local political actors to reach out to and mobilize members of different caste groups.

In order to ascertain the identity of influencers, we rely on a multi-stage survey to identify the individuals the most likely to influence voters at the local level across a large sample of villages. To do this, we rely on attitudinal measures. Practically, we ask a random sample of voters to select the two local-level individuals that would be the most likely to influence them ahead of elections. Relying on this data, we show that voters far from systematically report a coethnic as being most likely to influence them ahead of elections. We further use these voter responses to understand the ethnic composition of the personal networks assembled by influencers. To test our hypothesis, we focus on characterizing the relationship between competitiveness in the market for influencers and ethnic division, i.e., the extent to which personal networks assembled by these locally influential individuals focus on coethnics. To measure competitiveness, we rely on the difference in the popularity of the top two most influential people at the local level. To characterize the ethnic division in an influencers' personal networks, we measure the difference between the proportion of coethnics and the proportion of non-coethnics naming the indi-

[1]We are agnostic in this paper on the causes of different levels of competition between influencers. NS: Need more here.

vidual as influential (a larger gap implies greater ethnic division and a smaller one multiethnic networks).

Results support our argument that greater competition between influencers is associated with greater ethnic division. In particular, when the most popular influencer is being challenged, the most popular influencer has a high share of coethnics in his personal network. In less competitive places, where the most popular influencer has separated himself from the pack, the most popular influencer displays a more multiethnic network. Furthermore, subsequent interviews of these locally influential individuals confirm our findings based on voters' self-reports.[2] We ask these influencers to report how easily they would be able to mobilize each of the voters we previously interviewed for a political event, and how ready they would be to help each of these voters. These are two tasks that these actors commonly engage in. While we think of their willingness to help a voter as predictive of their ability to influence her, we think of their ability to mobilize her as a *manifestation* of this influence. Taken together with voters' own perceptions of influencers, this allows us to characterize the personal networks constructed by influencers, and how these networks vary with the local characteristics of the market for influencers. Results on these additional analyses reinforce our main result, as they confirm that dominance in the local market for influence leads influencers to become more inclusive.

2 Theory: Influence Networks in Ethnicized Societies

To what extent are local intermediation networks built along ethnic lines? The literature on ethnic politics has so far remained relatively agnostic on this issue. In highly ethnicized societies, where ethnic ties strongly structure social relations and politics, it is however often *assumed* that these local influence networks strictly follow the boundaries of ethnic groups. Prominent theoretical works in the comparative literature on ethnic politics frequently assume that voters are influenced by leaders from their own group (Shepsle and Rabushka, 2008; Horowitz, 1985; Fearon and Laitin, 1996). Several arguments potentially justify this assumption in the literature on ethnic conflict and ethnic politics. Non-coethnics tend to be harder to reach (Habyarimana et al., 2007). Voters tend not to trust non-coethnic leaders (Gay, 2002). More generally speaking, in many contexts, ethnic boundaries are too thick and too costly to be crossed (Chandra, 2006). In other contexts, ethnic homogeneity at the local level simply makes it implausible that voters

[2] As part of this interview, we ask these influencers to participate to a cross-referencing exercise allowing us to evaluate their ability to influence voters.

could enter the network of a non-coethnic.

Where and when political intermediation takes place along ethnic lines, one should expect voters to be influenced and mobilized by a member of their own group, who herself only mobilizes coethnics - what we call "ethnic influencers" in this article. With regards to electoral politics, this would broadly suggest that voters strictly coordinate on preferences within their ethnic group, since their only potential sources of influence are coethnics. Baldwin (2013) illustrates this logic, as she shows that tribal chiefs in Zambia are particularly influential with their fellow co-ethnics. In India, caste leaders have similarly been described as playing a central role in political brokerage during elections. Recent electoral successes in the country have for instance been attributed to "social engineering strategies" that explicitly lead party higher-ups to recruit caste leaders ahead of campaigns in an effort to attract the vote of specific caste groups.[3]

In this article, we contend that a second model exists and that local-level influence networks need not be organized around ethnic lines in places in which ethnicity is politically salient. We argue that some influencers, in fact the most prominent ones, are likely to be "inclusive influencers" and reach beyond their coethnics[4] - under a minimal number of scope conditions.

Under what conditions are we likely to see "inclusive influencers" emerge?

2.1 Diversity at the Local Level and the *Potential* For Inclusive Networks

For influencers to adopt inclusive strategies, it first needs to be possible for them to include non-coethnics in their networks, which of course requires that ethnic diversity exists at the local level. When ethnic diversity exists at the local level, influencers must be able to transcend ethnic boundaries and mobilize members of other groups. While this is unlikely to be the case in places in which crossing boundaries between groups is costly (for instance, in post-conflict situations) or in which norms entirely prohibit contact between members of different groups, this should nonetheless be the case in a large number of places, in which such severe restrictions do not exist.

Even when they can influence non-coethnics, intermediaries and local leaders will still find it easier to influence coethnics. This neither implies that influencers systematically manage to attract all of their coethnics at the local level within their network nor that their network is en-

[3]See for instance http://www.hindustantimes.com/assembly-elections/assembly-elections-2017-how-bjp-won-and-the-sp-bsp-lost-the-dalit-vote-in-up/story-rPZmjerz0e5figRRuKUdTP.html

[4]Inclusive influencers may be equally good at mobilizing coethnics but surpass lower-quality influencers when it comes to non-coethnics. Or they may simply be better at mobilizing both coethnics and non-coethnics.

tirely composed of coethnics. It simply implies that they have a strategic advantage to include coethnics vis-à-vis non-coethnics in their networks. Generally speaking, being an influencer takes a lot of effort.[5] In this context, attempting to influence non-coethnics is even costlier in time and/or resources. This may be for a number of reasons. In light of residential segregation, non-coethnics may be more costly to reach, approach and convince. Besides, communication may be more difficult with non-coethnics. Most importantly, levels of trust may, at least at the outset, be lower. Influencing non-coethnics is accordingly costlier, since this is a harder task, though not an impossible one.

For influencers to be inclusive, there also need to be non-coethnics willing to be influenced by an individual who does not belong to their own group. But we argue that many voters should in fact be ready to join the local network a non-coethnic influencer, including in the most ethnicized of societies. There are two reasons for this. First, it can be assumed that there are voters within each ethnic group for whom the quality of an intermediary is disproportionately more important than her ethnicity. The second reason has to do with inequalities across groups; not all groups are equal in their preference for co-ethnics, leaving some groups relatively open to the influence of a non-coethnic. These inequalities across groups may be driven by several factors. Members of smaller groups may first wish to associate with leaders from bigger groups, either for strategic reasons or because they internally generate too small a pool of potential leaders to ensure that one always exist at the very local level. More generally speaking, the disadvantaged socio-economic status of one's group may limit the ability of that group to generate its own influencer, insofar as these actors typically need to be educated and to possess several forms of capital (wealth, social capital, symbolic capital, or connectedness).[6] In addition, class inequalities within groups may limit the intensity of the coethnic bias (Huber and Suryanarayan, 2016). Last but not least, in highly hierarchical systems, the coethnic bias of some groups may be higher among groups in the higher ranks of the hierarchy, assuming that a higher rank in the hierarchy implies a more prestigious position. Thus, in almost all ethnicized societies, there should thus be an abundance of voters willing to be part of multi-ethnic influence networks.

[5] As evidenced from the growing literature on political brokerage, local intermediaries allocate much of their time to discussing problems with citizens who approach them, and subsequently solving these problems by approaching various officials through their network. In electoral periods, they in turn allocate the lion's share of their time to mobilizing the electorate for a host of political events in the lead-up to Election Day, and to convincing voters.

[6] Our data on influencers in appendix C affirms this point.

2.2 Political Competition and the *Incentive* to Maintain Inclusive Networks

In addition to being able to mobilize, assist and influence non-coethnics, influencers also need to have an incentive to do so, especially since reaching non-coethnics is costlier than reaching coethnics. In this article, we argue that electoral competition is sufficient to generate incentives for local influencers to consider including non-coethnics in their network, despite the fact that doing so is relatively difficult. Suppose, for the sake of the argument, that there is no real electoral competition in a political system. In this case, influencers get a fixed rent from elites for their work, since there is no competition for their services. Even though there might be several influencers at the local level, there is no market for quality influencers, since there is only one likely buyer of their influence or clientele.

If political competition at the elite level exists, by contrast, the incentives of local influencers may change. When multiple credible winners of elections exist, local influencers can simultaneously entertain offers from several elites before elections.[7] Assuming they are not purely ideological or partisan actors, they develop significant abilities to bargain with elites and renegotiate the terms of their agreements with them. In this market-like situation, influencers now have an explicit incentive to increase the size of their coalition since elites prefer influencers with large followings in order to win elections (Szwarcberg, 2015). Under the pressure of electoral competition, the rents delivered to locally dominant influencers may thus increase.

In this context, in which a larger network may now be adequately rewarded by elites who wish to win close elections, we should expect influencers to be more motivated to increase the size of their network than under a system that lacks political competition. As incentives and rent increase, we should simultaneously expect to observe more influencers emerge and increased competition *at the local level* between influencers to gain the favors of political elites.[8] As noted above, competition at the local level between influencers need not however be homogenous across locations.[9] In some cases, one of these actors may become a "dominant influencer" at the local level – that is, his quality far surpasses that of other local influencers[10], leading him

[7] In a related paper, we show that most influencers are not partisan actors in the context of our study. This enables them to renegotiate with elites.

[8] Political competition and the ensuing emergence of a market for local influence on voters accordingly is a double-edged sword for local influencers. On the one hand, it creates the conditions for them to potentially increase their rents. On the other, it creates more competition at the local level, between influencers, which makes their ability to increase their rents a more strategic game.

[9] Note here the difference between our scope condition - that political competition exists in the system, among elites - and this competition between influencers *at the local level*.

[10] This may be because he employed a more successful strategy or was more talented to begin with. We are agnos-

to maintain a comparatively larger network. In other cases, several "contested influencers" of relatively similar abilities and network sizes may coexist.

3 Model: Competition Between Influencers in Ethnicized Societies

Under these rather minimal scope conditions – 1. diversity at the local level, 2. a strategic advantage in including co-ethnics in one's network, and 3. political competition among elites (and the ensuing development of a market for local intermediaries) – we argue that some of the most prominent influencers should in fact be "inclusive influencers". The central intuition of our formal model is that "dominant influencers" are much more likely than "contested influencers" to invest their finite time and resources towards including non-coethnics in these networks. "Dominant influencers" are, in other words, comparatively more likely to be "inclusive influencers".

This prediction emerges from the highly stylized model detailed in the following pages. In this model, we analyze how a potentially inclusive influencer, influencer j, allocates effort to coethnics and non-coethnics when faced with a contesting influencer, j'. The goal here is to demonstrate that when j possesses significantly higher ability than j', it is in his best interests to build an inclusive coalition. If, on the other hand, the competing influencer, j', possesses levels of ability similar to j, then j is pushed to allocate more effort towards his own coethnics. Implications for the relative ethnic inclusiveness of an influencer's coalition are derived here.

3.1 Preferences and Actions

Consider a set of groups $G_1, \ldots, G_K$ with sizes $N_1, \ldots, N_K$, so the total population is $N = \sum_k N_k$. Further consider a citizen $i \in \mathcal{V}$ and influencers $j, j' \in \mathcal{I}$. Let $g()$ be an indexing function that denotes the group membership of i or j. So if $i \in G_k$ and $j \in G_q$, then $g(i) = k$ and $g(j) = q$.

The *ability* of influencer j is given by $\Gamma_j \in \mathbb{R}_+$. The influencer chooses an action of how much to emphasize his ability, $\alpha_j \in \{0, A\}$ with $A > 0$, and how much to emphasize his coethnic appeal, $\lambda_j \in \{\lambda_L, \lambda_{Hj}\}$ with $\lambda_{Hj} > \lambda_L > 0$ and the strength of the high appeal specific to influencer j. Because influencers must necessarily tradeoff between emphasizing ability and ethnicity, their action spaces consists of two (pure strategy) ordered pairs:

$$(\alpha_j, \lambda_j) \in \{(0, \lambda_{Hj}), (A_j, \lambda_L)\}$$

tic in this manuscript as to what leads to this dominance.

Simply put, an influencer can put low emphasis on ability and high emphasis on ethnic identity or the reverse. Concretely, an influencer can structure his appeal to citizens based on his ability to help citizens through high-level bureaucratic and political connections or personal speaking ability or charisma that encourages all citizens to turn up for events and rallies – or he may choose to simply appeal to his ethnic group through emphasizing shared symbols and practices with this group. Obviously, the willingness to appeal based on high-level connections or personal charisma will be a function of the strength and scale of such abilities.

The ethnic appeal of the influencer has a positive impact on co-ethnics, but has no positive impact on citizens from other groups, with a stronger impact on co-ethnics if the influencer chooses to emphasize ethnic identity. At the same time, *all* citizens care about the overall ability of an influencer (compared to his competitor), so higher ability than a competitor is positively viewed across the board – but only if the influencer chooses to emphasize these differences in ability and play down ethnic identity. Taking these considerations into account, the utility for citizen i from influencer j when compared against j', $U_{ij,j'}$, is given by:

$$U_{ij,j'} = \begin{cases} \alpha_j b(\Gamma_j - \Gamma_{j'}) + \lambda_j + \frac{1}{2}\gamma_{g(i)} + \frac{1}{2}\varepsilon_i & \text{if } g(i) = g(j) \\ \alpha_j(\Gamma_j - \Gamma_{j'}) + \frac{1}{2}\gamma_{g(i)} + \frac{1}{2}\varepsilon_i & \text{if } g(i) \neq g(j) \end{cases} \tag{3.1}$$

$$\varepsilon_i \sim Unif(-c_{g(i)}, c_{g(i)})$$

$$c_{g(i)} \sim \chi^2(c)$$

$$\gamma_{g(i)} \sim N(\gamma, \sigma_\gamma^2)$$

$$b \in \mathbb{R}_+$$

Each of the stochastic terms in the above formulation, ε_i,$c_{g(i)}$, and $\gamma_{g(i)}$, are independently and identically distributed. Note here that ε_i is an idiosyncratic level of support (ideological/personal attachment) for voter i that is group-specific – allowing for different levels of ideological heterogeneity by group ($c_{g(i)}$) and different biases ($\gamma_{g(i)}$) towards/against influencer j as compared to j' (the fraction, $\frac{1}{2}$ is a normalization factor that will keep calculations cleaner later on), with $E(\gamma_{g(i)}) = \gamma$ and $E(c_{g(i)}) = c$. These differences may emanate from, for instance, differing levels of social connectedness within a caste group and differing levels of association (based on social status) with an influencer's caste. Finally, b is parameter which measures the importance of relative ability differences to the citizen (as b goes to zero it matters less and less – which maps to

the scope conditions for inclusive influencers described above).

3.2 Objective Function for Influencers

The objective of influencer j is to maximize the size of his coalition. A citizen i chooses the coalition of j over the coalition of j' if $U_{ij,j'} > U_{ij',j}$. Define $I_j^q = 1$ if $q = g(j)$ and 0 otherwise, i.e., an indicator function for co-ethnicity. Keeping the scales comparable (so a positive value of ε indicates a predilection towards influencer j), citizen prefers j to j' if:

$$U_{ij,j'} > U_{ij',j} \Rightarrow \tag{3.2}$$

$$\alpha_j b(\Gamma_j - \Gamma_{j'}) + I_j \lambda_j + \frac{1}{2}\gamma_{g(i)} + \frac{1}{2}\varepsilon_i > \alpha_{j'} b(\Gamma_{j'} - \Gamma_j) + I_{j'}\lambda_{j'} - \frac{1}{2}\gamma_{g(i)} - \frac{1}{2}\varepsilon_i$$

Some simple rearrangement yields:

$$U_{ij,j'} > U_{ij',j} \Rightarrow \tag{3.3}$$

$$(\alpha_j + \alpha_{j'}) b(\Gamma_j - \Gamma_{j'}) + I_j^{g(i)}\lambda_j - I_{j'}^{g(i)}\lambda_{j'} + \gamma_{g(i)} + \varepsilon_i > 0$$

Using the fact that ε is uniformly distributed, the proportion of group G_q voting for influencer j, p_j^q, is then given by:

$$p_j^q = \frac{c_q + (\alpha_j + \alpha_{j'}) b(\Gamma_j - \Gamma_{j'}) + I_j^q \lambda_j - I_{j'}^q \lambda_{j'} + \gamma_q}{2c_q}$$

under the simple assumption that $c_q > \left|(\alpha_j + \alpha_{j'})(\Gamma_j - \Gamma_{j'}) + I_j^q \lambda_j - I_{j'}^q \lambda_{j'} + \gamma_q\right|$, as shall be made throughout this exposition – implying that p_j^q is strictly between 0 and 1. The size of the coalition for influencer j then is given by:

$$\mathcal{C}_j = \sum_{q=1}^{K} p_j^q N_q \tag{3.4}$$

3.3 Strategy for Lowest-Ability Influencers

The next theorem demonstrates that if j has inferior ability than his local competitor, then he always emphasizes his ethnic identity.

Theorem 3.1 (Co-ethnic Strategy among Lower Ability Influencers). *If* $\Gamma_j < \Gamma_{j'}$, *then* $\alpha_j =$

$(0, \lambda_{Hj})$.

Proof: Because $\Gamma_j - \Gamma_{j'} < 0$, $U_{ij,j'}$ *is strictly decreasing in* α_j *(see equation (3.1)). This implies that* $\alpha_j = (0, \lambda_{Hj})$ *is a dominant strategy for influencer* j. $\square$

The above theorem suggests that in any competition between two influencers, the weaker influencer (i.e., the one with comparatively lower ability) will necessarily employ a strategy that emphasizes ethnic identity. The equilibrium is, thus, characterized by how the influencer of higher ability responds – a strategy that emphasizes either his relative ability or ethnic identity.

To develop some intuition about the higher ability influencer's strategy, imagine that $\Gamma_j = \Gamma_{j'}$. In this scenario, influencer j gains no advantage by emphasizing relative quality, so he will play a strategy emphasizing ethnic identity (maximizing the proportion of co-ethnics in the coalition). If, however, $\Gamma_j > \Gamma_{j'}$, then emphasizing relative ability can increase the proportion of individuals in the coalition across all groups, while emphasizing ethnic identity only helps among co-ethnics. Thus, as the relative ability influencer j grows sufficiently large as compared to j', influencer j switches from an ethnic strategy to an inclusive strategy that emphasizes relative ability.

3.4 Strategy for Highest-Ability Influencers

3.4.1 The Ethnic Strategy

Remember that $E(c_k) = c$ and $E(\gamma_k) = \gamma$ for all groups G_k, so there is no a priori difference in expectation of ideological heterogeneity or predilection towards an influencer (outside of those resulting from a co-ethnic tie) for a co-ethnic group as compared to a non-co-ethnic group. Furthermore, each of these terms is independent of all other terms in the expressions. Using theorem 3.1, the expected proportion of the co-ethnics in the highest ability influencer's coalition when he undertakes an ethnic strategy, $(\alpha_j, \lambda_j) = (0, \lambda_{Hj})$ is given by:

$$p_j^{g(j)}(\alpha_j, \lambda_j) = \begin{cases} \frac{c+\gamma+\lambda_{Hj}-\lambda_{Hj'}}{2c} & \text{if } g(j) = g(j') \\ \frac{c+\gamma+\lambda_{Hj}}{2c} & \text{if } g(j) \neq g(j') \end{cases} \tag{3.5}$$

Using theorem 3.1, the expected proportion of the non-co-ethnics in the highest ability influencer's coalition when he undertakes an ethnic strategy, $(\alpha_j, \lambda_j) = (0, \lambda_{Hj})$ for $q \neq g(j)$ is given

by:

$$p_j^q(\alpha_j, \lambda_j) = \begin{cases} \frac{c+\gamma}{2c} & \text{if } g(j) = g(j') \\ \frac{c+\gamma-\frac{N_{g(j')}}{N-N_{g(j)}}\lambda_{Hj'}}{2c} & \text{if } g(j) \neq g(j') \end{cases} \tag{3.6}$$

The only part of j's strategy that shows up in these expressions is λ_{Hj}, suggesting that only the proportion of co-ethnics in j's coalition responds to an ethnic strategy. More importantly, in this scenario, the popularity of j among non-co-ethnics is only a function of γ, and thus the proportion of co-ethnics supporting j *will not converge* to the proportion of non-co-ethnics supporting j. Furthermore, for a fixed value of γ, the overall popularity in the population is function of relative popularity among co-ethnics, so the most popular influencers will also exhibit the highest values of λ_{Hj}.

3.4.2 The Inclusive Strategy

Using theorem 3.1, the expected proportion of the co-ethnics in the highest ability influencer's coalition when he undertakes an inclusive strategy, $(\alpha_j, \lambda_j) = (A, \lambda_L)$ is given by:

$$p_j^{g(j)}(\alpha_j, \lambda_j) = \begin{cases} \frac{c+\gamma+A_j b(\Gamma_j-\Gamma_{j'})+\lambda_L-\lambda_{Hj'}}{2c} & \text{if } g(j) = g(j') \\ \frac{c+\gamma+A_j b(\Gamma_j-\Gamma_{j'})+\lambda_L}{2c} & \text{if } g(j) \neq g(j') \end{cases} \tag{3.7}$$

Using theorem 3.1, the expected proportion of the non-co-ethnics in the highest ability influencer's coalition when he undertakes an ethnic strategy, $(\alpha_j, \lambda_j) = (0, \lambda_{Hj})$ for $q \neq g(j)$ is given by:

$$p_j^q(\alpha_j, \lambda_j) = \begin{cases} \frac{c+\gamma+A_j b(\Gamma_j-\Gamma_{j'})}{2c} & \text{if } g(j) = g(j') \\ \frac{c+\gamma+A_j b(\Gamma_j-\Gamma_{j'})-\frac{N_{g(j')}}{N-N_{g(j)}}\lambda_{Hj'}}{2c} & \text{if } g(j) \neq g(j') \end{cases} \tag{3.8}$$

Importantly, both the coalitions of co-ethnics and non-co-ethnics are impacted by the term $(\Gamma_j - \Gamma_{j'})$. Furthermore, under the inclusive strategy, j moves α_j from 0 to A so that his relative ability advantage impacts both coalitions. This is unlike the ethnic strategy, which only impacts co-ethnics. Finally, given that λ_L is fixed, we note that the proportion of non-co-ethnics in the coalitions will *converge* to the proportion of co-ethnics in the coalition as the difference in relative ability grows. The speed of this convergence is determined by the responsiveness in the population to ability (b) and the influencer-specific skill in advertising this difference in ability (A_j).

3.5 Equilibrium Behavior

In order to conduct the analysis, we look at the difference in the coalition size between the highest quality influencer selecting the ethnic strategy versus the inclusive strategy (an identical calculation whether or not the competing influencer is or is not a coethnic) – namely, we calculate whether the following expression is positive or negative:

$$\Delta \mathcal{C}_j = (p_j^{g(j)}(0, \lambda_{Hj}) - p_j^{g(j)}(A_j, \lambda_L))N_{g(j)} + \sum_{q \neq g(j)} (p_j^q(0, \lambda_{Hj}) - p_j^q(A_j, \lambda_L))N_q \tag{3.9}$$

$$= \frac{(\lambda_{Hj} - \lambda_L) - A_j b(\Gamma_j - \Gamma_{j'})}{2c} N_{g(j)} - \frac{A_j b(\Gamma_j - \Gamma_{j'})}{2c} \sum_{q \neq g(j)} N_q$$

$$= \frac{(\lambda_{Hj} - \lambda_L)}{2c} N_{g(j)} - \frac{A_j b(\Gamma_j - \Gamma_{j'})}{2c} N$$

The ethnic strategy is preferred to the inclusive strategy if $\Delta \mathcal{C}_j > 0$. Some simple rearranging of terms demonstrates that the ethnic strategy is preferred if:

$$(\lambda_{Hj} - \lambda_L)\frac{N_{g(j)}}{N} > A_j b(\Gamma_j - \Gamma_{j'}) \tag{3.10}$$

This simple inequality in (3.10) provides the key factors in determining the ethnic or inclusive strategy. As the relative gap in ability grows, or the responsiveness of the population to relative ability and/or influencer-specific skill in advertising ability difference grows, the highest ability influencer is more likely to play an inclusive strategy. As the relative strength of the ethnic appeal grows, or the size of the influencer's ethnic group grows, one should expect that the highest ability influencer is more likely to play the ethnic strategy.

Finally, note that under both the ethnic and inclusive strategies, the higher quality influencer will do worse among co-ethnics if facing a co-ethnic competitor for influence. This is because the competitor will necessarily play an ethnic strategy that cuts into co-ethnic support. For this reason, large populations of one ethnic group are likely to have countervailing effects on the level of support among co-ethnics. On one hand, the relative size of the co-ethnic group makes the ethnic strategy more appealing; on the other hand, the competitor is far more likely to be a co-ethnic as well.

3.6 Predictions

Under what conditions should we expect to see the proportion of coethnic and non-coethnic voters included in dominant influencers' coalition converge - that is, under what conditions should we observe inclusive coalitions?

3.6.1 *Conditions for Divergence*

Summarizing the model, we find that if the most dominant influencers (i.e., those with the comparatively largest coalitions at the local level) are also those that are most able to muster the strongest ethnic appeals, then we should expect divergence between the proportion of co-ethnics in the most popular influencer's coalition and the proportion of non-co-ethnics in the coalition. This is because only the size of the co-ethnic population in the coalition increases as a function of the ethnic strategy through λ_{Hj}, while the size of the non-coethnic population in coalition remains unchanged.

3.6.2 *Conditions for Convergence*

By contrast our model suggests that if the most popular influencers (i.e., those with the largest coalitions) largely play an inclusive strategy, then we should expect convergence between the proportion of co-ethnics in the most popular influencer's coalition and the proportion of non-co-ethnics in the coalition for reasons described in detail above. In other words, *the probability that a non-coethnic is included in an influencer's network converges to the probability that a coethnic is included in the influencer's network as her dominance increases.* This is the hypothesis we test in the rest of this article.

It is important to note that this is not a claim about an increase in the number of non-coethnics and coethnics in the coalition, but rather that the marginal propensity to include non-coethnics in the coalition grows faster than the marginal propensity to include coethnics as the dominance of the influencer grows. This convergence is a direct result of an inclusive strategy, as more dominant influencers playing an stronger ethnic strategy would yield the opposite result (as discussed above).

4 Research Design

In order to test this hypothesis, we rely on an innovative research design that allows us to identify local influencers, and to interview them along with voters as part of a cross-referencing exercise.

4.1 The Context: Bihar

This project was conducted in the Indian state of Bihar immediately after the 2015 state elections. Bihar is known as a state where caste and religion are highly salient social cleavages, as well as a state in which political mobilization has taken place along caste lines (Witsoe, 2013; Vaishnav, 2017). Until recently, Bihar politics was largely characterized by competition between the Rashtriya Janata Dal [RJD] and the Janata Dal (United) [JD(U)]. The RJD, led by the charismatic Lalu Prasad Yadav, is often described as having a core base made of Yadavs and Muslims, the so-called "Y-M coalition," governed the state from 1997 to 2005.[11] The JD(U), which has ruled the state since, is largely associated with "other backward castes" (OBCs) outside of the Yadav population. In the 2015 election, once bitter foes RJD and JD(U) joined forces in a pre-electoral alliance to contest against an ascendant Bharatiya Janata Party (BJP), a party largely associated with upper caste Hindus in India.

In this highly ethnicized context, observers often assume that political intermediaries and influencers are arrayed according to caste and religion, consistent with the state's political history, and thus one should expect to find a preponderance of "ethnic influencers." At the same time, there is extraordinary ethnic heterogeneity at the village level; in fact, the 1931 Indian Census (the last year for which we have reliable data on caste composition in India) reveals that the single largest caste group (Yadavs) were a mere 11% of the population. In our own data, there are 99 distinct caste groups and the average size of the largest caste group in the village is just 25% of sampled voters. Furthermore, the incumbent chief minister, Nitish Kumar, the chief of JD(U), is purported to be a leader who can make broad-based development appeals cutting across ethnic lines (Joshi, Ranjan and Sircar, 2015). This suggests that, even in a place like Bihar where ethnicity is highly salient, there is an opportunity for particularly skilled local-level influencers to build coalitions beyond ethnicity and become "inclusive influencers."

[11] Mr. Yadav is widely considered as one of the key faces of the "post-Mandal era" in Indian politics, where certain backward castes (particularly Yadavs) gained greater political power and representation in the system (Jaffrelot and Kumar, 2009).

4.2 Generating the Sample

In this project, we seek to identify influencers at the most local level of politics in Bihar. In order to do so, we selected the polling booth area (PBA) as the lowest sampling unit for the study. State-level legislators, or Members of Legislative Assembly (MLAs), are selected from assembly constituencies (ACs). Each AC contains approximately 200 polling booths, and polling booths average approximately 1000 voters. The polling booth is the smallest political unit in India (much like a precinct in the United States). Large villages may contain multiple polling booths, while smaller villages are typically represented by a single one. As such, the PBA is a political unit often smaller than a village and thus characterized by extremely dense social relations.

We sampled polling booths from 3 districts of Bihar: Buxar, Nalanda, and Vaishali. These were chosen to ensure some minimal levels of cultural, political and socio-economic diversity in our sample.[12] In each of these three districts, we then randomly selected three blocks.[13] In each of these blocks, we randomly selected polling booths using a variant of systematic random sampling.[14] The outcome of this process was a random sample of 179 PBAs over 9 blocks and 3 districts of Bihar.

4.3 Phase 1: Identifying Influencers

In the first phase of our study, we identified influencers in each selected polling booth area (PBA).

4.3.1 Generating a 'Long List' of Influencers

To generate such a list, the research team asked the following three questions in at least five locations within each selected PBA, in the following order:[15]

[12] Nalanda is located about 50-100 km south of Patna (the state capital) and is a Magahi-speaking area. Buxar is located about 125-200 km west of Patna and is located in the Bhojpuri-speaking area of the state. Finally, Vaishali is located just across the Ganges River, north of Patna, and is located in the Maithili-speaking area of the state.

[13] In each district, we excluded a small number of blocks that would be prone to flooding (which would have made the work of the research team complicated during the rainy season) as well as several blocks that were not easily accessible by road, prior to random sampling. This was to ensure the security of survey teams as well as to guarantee that our implementing partner (SUNAI) would be able to implement the complex protocol detailed below in a timely fashion.

[14] We broke the list of polling booths in each block into 40 intervals with approximately the same number of polling booths, randomly selected whether we would take odd or even numbered intervals (i.e., first, third,..., or second, fourth,...), and then we randomly selected a polling booth in each interval. Each interval contained consecutive "polling booth numbers" which means that they are likely spatially clustered. This protocol thus minimizes the likelihood that we select neighboring polling booths.

[15] These areas were by design dispersed within the booth, based on information about the caste/religious composition of the booth.

1. *Who among residents of this area is most influential?*

2. *When it comes to social issues, whose opinions do people listen to the most around here?*

3. *When people seek to solve small problems outside the family in this village without approaching the panchayat or political party, who do they go to?*

In this first step, we deliberately asked questions that were *not* specific to politics, so as to generate as large a sample as possible.[16] Although the research team was tasked with collecting at least ten different names in each PBA, they collected over fifteen on average, suggesting that enumerators typically had no trouble identifying names. Nonetheless, if fewer than ten names were collected at the end of this process, the research team was mandated to visit additional locations until the list included ten names. Obtaining at least ten names in such a small area ensures that we obtain a relatively exhaustive list of intermediaries, including elected officials, local party leaders and other potentially influential individuals at the local level, such as "social workers."[17] As they obtained names in response to these questions, the research team asked for a few additional details about each of the individuals named (their phone number, whether they hold a position in a political party, whether they are elected in any political or non-political local body, as well as their profession, age, and community/social group). The research team used their responses to create a 'long list' of influencers which we subsequently used in our voter survey.

4.3.2 Selecting The Two Leading Influencers

Using official voters' lists, we randomly sampled twelve male voters in each PBA.[18] We first asked these voters to report information about their caste, education, occupation, living conditions and political preferences, which we use in our statistical analyses below.

Most importantly, we asked these respondents to choose the two most influential individuals from the 'long list' of influencers described above.[19] Specifically, we asked each respondent

[16]The 3rd question specifically excludes partisan and elected individuals - so as to ensure that villagers do not feel compelled to focus on these actors. The first two questions however potentially include these individuals. As a result, villagers named both elected or partisan individuals as well as more informal actors, usually known as "social workers".

[17]As it turns out, and as reported in Chauchard and Sircar (2017), remarkably few of the individuals identified through this process were either elected or partisan. Most were informal "social workers".

[18]While we were interested in differential response by gender, we chose not to select female respondents to minimize risks, insofar as we feared that some responses provided by female respondents could put them at risk. In practice, we used a Kish table to randomly select one male voter from the set of male voters in the household.

[19]In addition to picking two "real" influencers from our list, we also select a 3rd "benchmark influencer" in each

to name two individuals on the long list of influencers in response to the following question(s): *Which two of these people would you (you personally) be most likely to listen to and follow? Which of these would most influence you?* To illustrate this somewhat abstract concept of influence and anchor it in the context of electoral politics, interviewers followed up with the following statement: "if you had to choose between two candidates in elections, which of these individuals would most likely affect your decision?". Since it is unlikely that our relatively exhaustive "long list" of influencers missed the most influential individuals at the PBA-level, and since a random sample of voters answered this question, we are confident that this process allows us to select the most popular influencers at the PBA-level.

To select our first influencer, we simply selected the most popular individual on the list (that is, the influencer that had been nominated as influential by the largest number of our 12 respondents, in each PBA). We went down the list in order of popularity if he or she was unavailable or if did not provide consent.[20] We will refer to the first chosen influencer as the $T1$ influencer. The second influencer (the $T2$ influencer) was the most popular influencer among those remaining on the list.[21,22]

We build several important variables based on the number of villagers who recognized our $T1$ and $T2$ as influential. The difference in popularity between $T1$ and $T2$ provides us with a convenient way to measure the relative dominance of $T1$ influencers: the larger the difference in votes received between $T1$ and $T2$, the most dominant $T1$ can be considered to be (by extension, the smaller the difference between $T1$ and $T2$, the more contested $T1$ can be considered to be). This provides us with the key independent variable we use to test our hypothesis. Second, a binary variable denoting whether or not each voter selected the influencers ($T1$ and/or $T2$) in our sample helps us disentangle the set of voter characteristics (and dyadic relationships between influencer and voter) that predict selection of $T1$ and $T2$ in phase 1. Finally, a simple characterization of which influencer is designated as $T1$ and which is designated $T2$ by each voter helps

booth. This third influencer ($T3$) simply was the head of household of a randomly drawn household within the PBA. We do not make use of data about $T3$ in this paper.

[20] If there was a tie in popularity, the research team randomly picked one of the (equally) popular individuals.

[21] As above, if there is a tie in popularity, the team randomly picked among equally popular individuals.

[22] In our design, an additional requirement existed for T2: they had to be currently unelected. This allowed us to ensure that we did not only select influencers who currently are (or just were) elected. Note that in a vast majority of cases analyzed here, we however do not select a single elected individual, as the two most popular influencers, as defined here, often are unelected. Because of this, the difference between our T1 and T2 influencers is best thought as a mere difference in popularity (with T1 being named as the most "influential", and T2 being named as the second most "influential"). Thus, we always pick two different individuals, and at least one individual who is not elected.

us understand how the networks of our most popular influencers ($T1$) differ from that of less popular ones ($T2$).

4.4 Phase 2: Interviewing Influencers and Cross-Referencing

In the second phase of the study, we conduct interviews with our selected $T1$ and $T2$.

This allows us to collect basic demographic information on these two influencers and hence to "cross-reference" the profile of these influencers with voter information from phase 1. This also allows us to generate two variables indicative of the degree of influence of $T1$ and $T2$ over different voter profiles. In order to further understand who $T1$ and $T2$ can influence, we measure their *ability to mobilize* and their *willingness to help* each of the voters. These variables are collected at the dyadic level; that is, the influencer gives a separate response for each voter. The behaviors are operationalized as described below:

- **Ability to Mobilize.** *If a party leader that you respect asked you to bring people from this area, how easy would it be for you to convince this person to turn up?* This variable is coded as a four-point scale, with 1 denoting "very difficult" and 4 denoting "very easy". This item allows us to assess the extent to which different types of influencers feel able to mobilize different types of voters for political, campaign-related events. We view the ability to mobilize as a natural manifestation of of the influencer's underlying ability to influence, especially when it comes to voting behavior.

- **Willingness to Help.** *If these 12 people come to ask you for help at same time, then who would you help first and who would you help last? Rank these individuals from 1 to 12, from the first person you would help to the last person you would help.* Because previous assistance is likely to be correlated to the ability to influence an individual,[23], this variable constitutes a good predictor of influence. More generally speaking, local intermediaries are widely thought to have influence on voters during elections because they assist them, solve their problems and help them access the state in between elections (Auyero, 2001; Stokes et al., 2013; Berenschot, 2012). Because the willingness to help is often a decision subject to constraints on time and money, we opted for a measure that would force the influencer to report the *relative* willingness to help voters interviewed as part of phase 1 (see Schneider and Sircar (2015) for a detailed discussion of this type of empirical strategy).

[23] As noted above, influencers are in our data 38 percent more likely to be see as influential when they have helped respondents at some point in the past.

In appendix A, we demonstrate that these two variables are strongly and independently related to the probability of being chosen by a voter in phase 1. Voters tend, in other words, to select influencers who claim to be able to mobilize them and who declare a willingness to help them. This suggests that these measures are related to the concept of influence, by showing that voters themselves recognize the highest quality influencers as those who report being able to mobilize them for a political event and a particular willingness to help them.

Since each of these variables is measured at the dyadic level, we can "cross-reference" the reported behavior with characteristics of the voter and the dyadic relationship between voter and influencer. For instance, by measuring the caste of both the influencer and the voter, we are able to determine whether each voter is a coethnic of the influencer. We can then associate coethnicity of the voter with the influencer's ability to mobilize the voter. This cross-referencing design, building on Schneider and Sircar (2015), presents a number of advantages. It first allows us to understand the networks and the preferences of real-world voters and influencers, as opposed to hypothetical characters presented as part of vignettes. Second, it helps circumvent social desirability concerns, insofar as it allows us to make inferences as to whether ethnicity influences the behaviors of influencers without having to explicitly mention the ethnicity of voters.[24]

5 Descriptive Data: Who Are Our Selected Influencers?

This research design provides us with a number of complementary strategies to test the hypothesis developed in section 2. Before we present these tests - over the next two sections-, it is useful to describe the sample of individuals we select through this process. In light of our main independent variable (whether or not influencers are "dominant influencers"), we also describe how the two influencers we select within each PBA relate one to another.

5.1 Characteristics of Influencers

In order to understand who our selected influencers are, and how T1 influencers differ from T2 influencers, it is useful to compare their characteristics to that of the general population we sample. To do this, the table in appendix B provides a detailed comparison between our two selected influencers (T1 and T2) and our randomly sampled male voters, which can be assumed

[24] An influencer may be reticent to directly reveal that he is more willing to help a coethnic if ethnicity is explicitly mentioned, whereas in this study the researcher can simply cross-reference an influencer's willingness to help with a measure of coethnicity obtained from the voter and influencer surveys.

to be representative of the population.

A cursory look at the data demonstrates that the selected $T1$ and $T2$ influencers are overall "higher status" individuals, compared to the population, and that this is especially true of $T1$. As compared to the general population, a significantly higher percentage of influencers belong to upper castes (and markedly fewer belong to the Scheduled Castes); they are much more educated (i.e. many more have passed class 12); they also tend to live in larger houses (as measured by the number of rooms) and are much more likely to live in permanent (*pucca*) structures. While this is significantly more than in the population, only 38% of $T1$ and 33% of $T2$ influencers however consider themselves a member of any party.

These data thus suggest that a higher status correlates with the ability of these individuals to have political influence at the local level. At the same time, the relatively low levels of party membership among this group (explored at length in a companion paper, Chauchard and Sircar (2017)) suggest that influencers derive their popularity from their underlying "quality" rather than from their partisan affiliations. A further substantiation that the difference between the $T1$ and $T2$ influencer has to do with the popularity of these characters comes from the fact that $T1$ influencers report an average of 29.48 visitors per week, whereas $T2$ influencers report an average of 19.45 visitors per week, supporting the idea that $T1$ overall influences a larger coalition than $T2$.[25]

Finally, it is worth emphasizing that the data are consistent with our underlying scope condition of a society in which ethnicity is highly salient; as such, there exist strong co-ethnic biases in the networks of influencers. In particular, influencers in our sample are nominated by a non-co-ethnic citizen 32.7% of the time, whereas they are nominated 44.7% of the time by co-ethnic citizens. The remainder of this manuscript is devoted to demonstrating that there is variation in the ethnic inclusiveness of these networks, depending on the extent of dominance of the influencer.

5.2 Dominance and Contestation Among Influencers

As noted in section 3, we view the number of voters nominating our $T1$ or $T2$ as influential as a proxy for her underlying quality. The relative difference in votes between $T1$ and $T2$ accord-

[25]We simply asked influencers to estimate the number of villagers that visited them every week to seek their help. These relatively large numbers, although they need to be taken with a grain of salt since they are self reports, confirm that the individuals we select are intermediaries frequently described as "social workers" or *dalals* in the scholarship on state-citizens relations in India.

ingly measures the relative difference in quality between the top two influencers, or the degree of contestation between the top two influencers. That is, when $T1$ has significantly more votes than $T2$, then we view $T1$ as a "dominant influencer". When the difference in votes between $T1$ and $T2$ is small, then we view the $T1$ influencer as a "contested influencer" in a competitive local market for influencers. In appendix C, we show that dominance of the $T1$ influencer is correlated to whether she is elected and has passed class 12.[26] This points to dominance being associated with underlying measures of social status and quality.

Figure 1 displays the number of villagers (out of a total of 12 in each PBA) who named $T1$ and $T2$ as most likely to influence them (figure 1(a)), as well as the difference in such votes between the $T1$ and $T2$ influencers in a polling booth (figure 1(b)).[27] Three points should be clear from looking at the data. First, a $T1$ influencer, nominated by an average of 5.56 villagers in each PBA, is about twice as likely to get selected by a respondent as compared to a $T2$ influencer, with an average 2.82 nominations; thus, $T1$ influencers are significantly more popular that $T2$ influencers. Second, despite being the most popular influencers, less than a majority of villagers in each PBA name $T1$ as most influential on average, suggesting a lot of heterogeneity in these estimates within PBAs. Third, while average difference in nominations between $T1$ and $T2$ is a little under 3, the difference in nominations between $T1$ and $T2$ varies between 0 and 9 at the polling booth level, suggesting great variation in levels of "dominance" of $T1$ in our sample.

6 Tests: Do Voters Name Coethnics As Influencers?

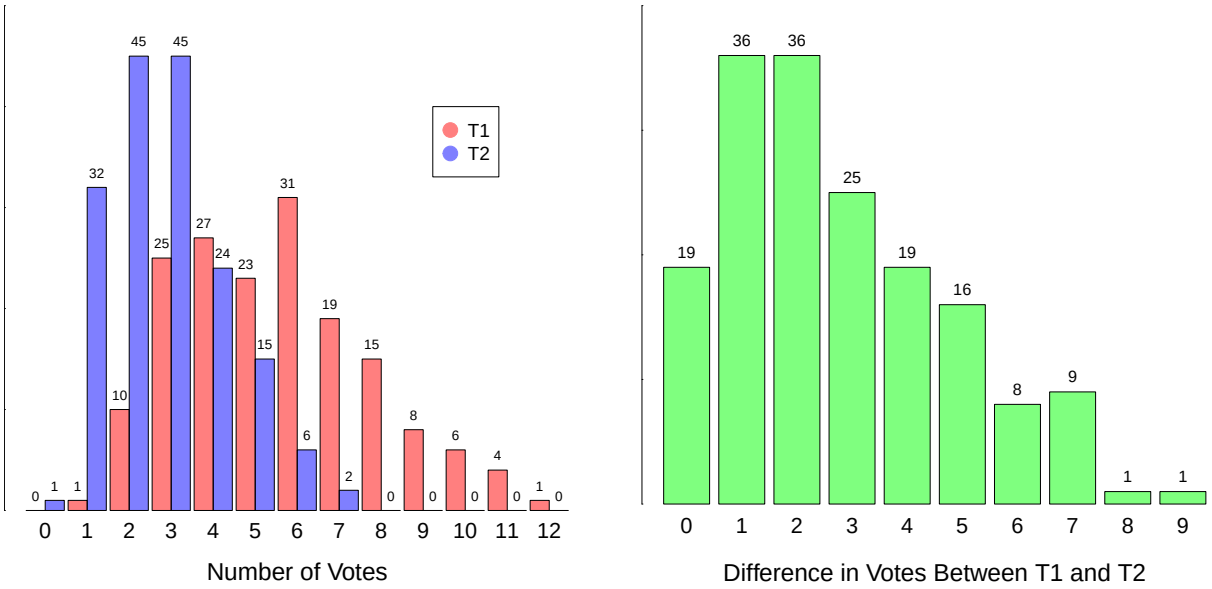
In order to test our hypothesis, we first evaluate the extent to which voters are prone to name coethnics as influencers, and the extent to which this varies across "dominant" vs. "contested" influencers. In other words, we test the hypothesis that as an influencer becomes more dominant, his coalition of supporters becomes more inclusive, i.e., non-coethnics make up a higher percentage of his supporters.

In order to operationalize this problem, we modeled the probability an individual respondent named the influencers we selected as $T1$ and $T2$ as "likely to influence [them]". Two predictors are crucial for this analysis. First, our cross-referencing design enables us to reliably

[26] We also note that the magnitude on being upper caste is quite high, although not significant.

[27] We note that in a small number of villages, we received vote totals that were unreliable. This data, and subsequent analysis, is restricted to the 170 polling booths in which the data on selection of influencers was reliable.

Figure 1: Distribution of Votes for $T1$ and $T2$



(a) Votes for $T1$ and $T2$

(b) Difference in Votes for $T1$ and $T2$

code whether our $T1$ or $T2$ influencer is a co-ethnic or non-coethnic of the individual respondent. Second, the difference in the aggregate number of votes for $T1$ and $T2$ measures how dominant $T1$ is in the polling booth area. In empirical terms, this allows us to measure whether the probability of a non-coethnic nominating $T1$ converges to the probability of a coethnic nominating $T1$, *as* $T1$ *becomes increasingly dominant*.

We model this as such. Let $choose_{ijk}$ denote a binary variable that takes the value of 1 when respondent i votes for influencer $j \in \{T1, T2\}$ in polling booth k. Let Δ_k denote the difference in votes between $T1$ and $T2$ in polling booth k. Further define C_{ij} to be a binary variable taking the value of 1 when respondent i is a co-ethnic of influencer j and Q_{ij} as a similarly defined binary variable denoting co-partisanship between respondent i and influencer j. The core regression model is:

$$P(choose_{ijk} = 1) = logit^{-1}(\beta_0 + \beta_1\Delta_k + \beta_2 C_{ij} + \beta_3 C_{ij} * \Delta_k + \mathbf{Z}_i\gamma + \alpha_i + \alpha_j + \alpha_k) \tag{6.1}$$

$$\alpha_i \sim \mathcal{N}(0, \sigma_i^2); \quad \alpha_j \sim \mathcal{N}(0, \sigma_j^2); \quad \alpha_k \sim \mathcal{N}(0, \sigma_k^2)$$

where β are the main parameters of interest in the model, $\mathbf{Z}_i$ is a matrix of control variables

with associated parameter vector γ, and the α terms correspond to random effects at the voter, influencer, and polling booth level in a hierarchical model. The control variables in these regression are whether influencer j is a party member and whether i is a co-partisan of influencer j (based on identical vote preferences in the last two elections). The random effects help address sources of variation at each of the voter, influencer, and polling booth level, as well as a complex "clustering" in the data at these levels. It is thus a conservative model of the data, and in our opinion the most adequate one.[28][29]

Table 1 reports the results of the regression, and figure 2 displays the results and its implications for the probability of choosing $T1$ as a function of dominance. Figure 2 displays the 90 percent simulated intervals from the regression, plotting the probability of voting for $T1$ among co-ethnics and non-co-ethnics as function of the vote difference between $T1$ and $T2$ (the level of dominance of $T1$). These simulations further assume that the influencer is a party member and a co-partisan of the influencer, thus controlling for two prominent explanatory variables for our dependent variable.[30]

As can be seen from these analyses, we find that greater "dominance" (that is, greater difference in the number of villagers naming $T1$ relative to $T2$ as influential) predicts that $T1$'s coalition will include a higher percentage of non-coethnics. This provides strong evidence for our hypothesis that more dominant influencers build more inclusive networks.

Two points are particularly important to understand the shape of $T1$ influencers' networks. First the probability of support among non-coethnics eventually *converges* to that of coethnics, suggesting that dominant influencers tend to have more inclusive coalitions (and very dominant influencers almost perfectly inclusive coalitions). Second, while the rate of support grows among co-ethnics as $T1$ becomes more dominant (suggesting that their network is overall larger), it grows faster among non-co-ethnics, explaining the convergence phenomenon. This is perfectly consistent with our hypothesis: as an influencer becomes dominant, a higher proportion of her attention is geared towards non-coethnics.

[28]The (Bayesian) regression is fit using Markov Chain Monte Carlo (MCMC) using the `JAGS` software (called within the `R` framework) and is based on 3 chains and 3750 simulations of the posterior distribution. This protocol applies to all regression models run in this paper.

[29]Three points are worth noting with respect to this model. First, the main parameter of interest in β_3, as $\beta_3 < 0$ denotes a situation in which the $T1$ influencer becomes more inclusive as he becomes more dominant. Second, this model simultaneously models choice for $T1$ (for which $\Delta \geq 0$) and $T2$ (for which $\Delta = 0$). Third, β_1 is mechanically guaranteed to be positive, but including this term guarantees that we have an appropriate baseline from which to determine whether an influencers coalition is becoming more or less inclusive.

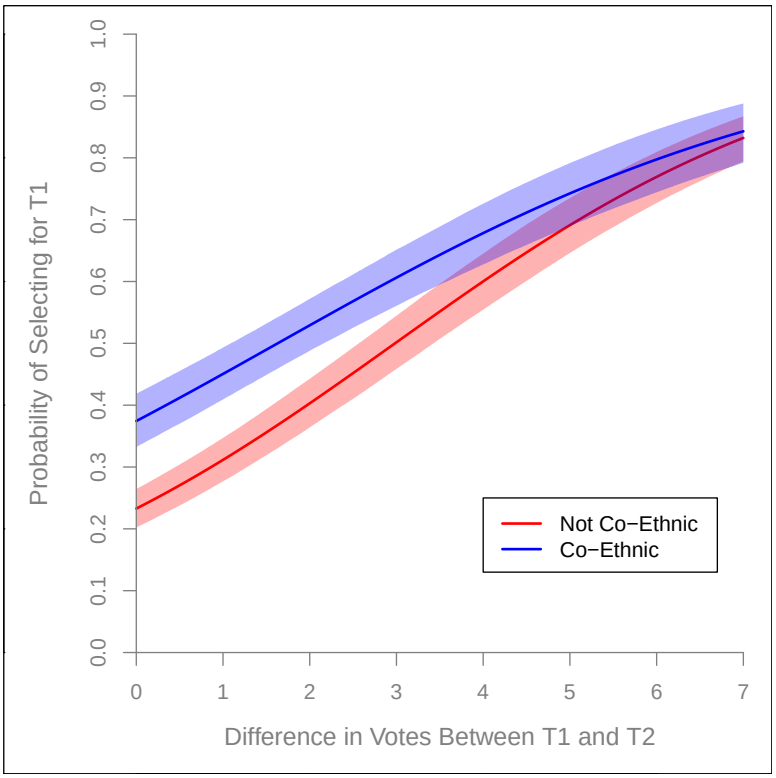
[30]Though the visualization is robust to other choices.

Table 1: Regression Results: The Relationship Between Dominance and Inclusive Support

	Dependent variable: Influencer Chosen (Logit)
Intercept	-1.479^{***}
	(0.078)
Δ_k	0.397^{***}
	(0.025)
Party Member	0.129
	(0.087)
C_{ij}	0.638^{***}
	(0.105)
$\Delta_k \times C_{ij}$	-0.075^{*}
	(0.041)
Q_{ij}	0.170^{**}
	(0.084)
pD	267.8
DIC	4833.1
Note:	$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Results report estimates from a 3750 posterior simulations from a regression model estimated in a Bayesian framework through Markov Chain Monte Carlo (MCMC) with 3 chains and diffuse priors on all parameters, using the program `JAGS`. Standard deviations of the posteriors on the respective parameters are given in parentheses. Statistical significance in the model is given with respect to the posterior distribution. In particular, let $\widehat{\pi}$ be a vector of values drawn from the posterior distribution of a parameter of interest. Then, we define $\underline{\pi} = 2 * P(\widehat{\pi} < 0)$. The deviance information criterion (DIC) is a measure of fit that is defined as the sum of one-half of the estimated variance of deviance (pD) and the expected value of the deviance. The lower value of DIC is taken to be a better fit, with pD entering as a penalty for overfitting the data.

Figure 2: The Relationship Between Dominance and Inclusive Support



Note that there is nothing tautological to this result, insofar as more dominant influencers (i.e., those with overall larger coalitions) could grow larger coalitions that feature similar ratios of coethnics and non-co-ethnics. As see in Figure 2, they may for instance continue adding co-ethnics as their dominance increases, since even very dominant influencers do not "max out" on coethnics at the local level. Yet this is not what the figure suggests: as their dominance increases, T1 influencers progressively include a higher proportion of non-coethnics, up to the point where they do not discriminate between coethnic and non-co-ethnic (on the right-hand side of the figure). This suggests a change of strategy.

7 Tests: Who Do Selected Influencers Mobilize and Assist?

An alternative and complementary strategy to test our hypothesis is to evaluate whether dominant influencers themselves declare maintaining more inclusive influence networks. Importantly, we cannot directly observe all of the appeals influencers make to their citizens, but we can understand the breadth of the appeals from reported behaviors by looking at convergence/divergence

in proportions of coethnics and nonethnics towards whom influencers display these behaviors.

To do this, we interview the selected $T1$ and $T2$ influencers in order to understand two of their behaviors vis-a-vis different types of voters. As noted in section 3, these behaviors are typical of the work of local intermediaries in rural India and are arguably related to their ability to influence voters (as we show in Appendix A). We first measure the extent to which the influencer declares being *able to mobilize* each sampled voter, which we think of as a manifestation or an illustration of their capacity to influence said voter. Second, we measure the influencer's relative *willingness to help* each sampled voter, which we think of as predictive of their capacity to influence said voter. To the extent that these capacities are general enough to be displayed across ethnic groups, they constitute the sort of ability (Γ) we discuss in the model above. On the other hand, if these capacities are restricted to coethnic populations, then they are likely driven by the ethnic appeal of the influencer (λ). Regressing the differences of votes between $T1$ and $T2$ on these variables allows us to explore whether the degree of local dominance of an influencer correlates with the maintenance of a more inclusive influence network.

7.1 Ability to Mobilize

As discussed in section 3, our self-reported measure of ability to mobilize varies from 1 to 4 (4 being highest). There is some prima facie evidence that relative differences in the quality of influencers are reflected in differences in the capacity to mobilize. Our $T1$ sample has an average capacity to mobilize score of 2.95 with the $T2$ sample reporting an average of 2.84. This may seem like a small difference but it is actually quite significant given that influencers tend to report very high capacity mobilize the respondents, as the t-statistic associated with this difference is 3.21 ($p = 0.001$).

In order to empirically characterize the relationship between ability to mobilize and dominance of the $T1$ influencer, we make inferences primarily based upon relative differences between $T1$ and $T2$ and not on the absolute levels of the measure. We do this because self-reported measures such as this one may include "response bias," e.g., influencers may have an incentive to inflate their capacity to mobilize and their reported closeness to respondents.

In order to evaluate the extent to which the level of dominance of the $T1$ influencer affects her ability to mobilize coethnic and non-coethnic voters, we run a linear model variant of (7.1). In particular, we define y_{ijk} as the reported ability to mobilize vis-s-vis respondent i by influencer j in polling booth k. The resulting equation is;

$$y_{ijk} = \beta_0 + \beta_1 \Delta_k + \beta_2 C_{ij} + \beta_3 C_{ij} * \Delta_k + \mathbf{Z}_i \gamma + \alpha_i + \alpha_j + \alpha_k + \varepsilon_{ijk} \tag{7.1}$$

$$\alpha_i \sim \mathcal{N}(0, \sigma_i^2); \quad \alpha_j \sim \mathcal{N}(0, \sigma_j^2); \quad \alpha_k \sim \mathcal{N}(0, \sigma_k^2); \quad \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$$

where β are the main parameters of interest in the model, $\mathbf{Z}_i$ is a vector of control variables with associated parameter vector γ, and the α terms correspond to random effects at the voter, influencer, and polling booth level in a hierarchical model. The control variables in these regression are once again whether influencer j is a party member and whether i is a co-partisan of influencer j (based on identical vote preferences in the last two elections).

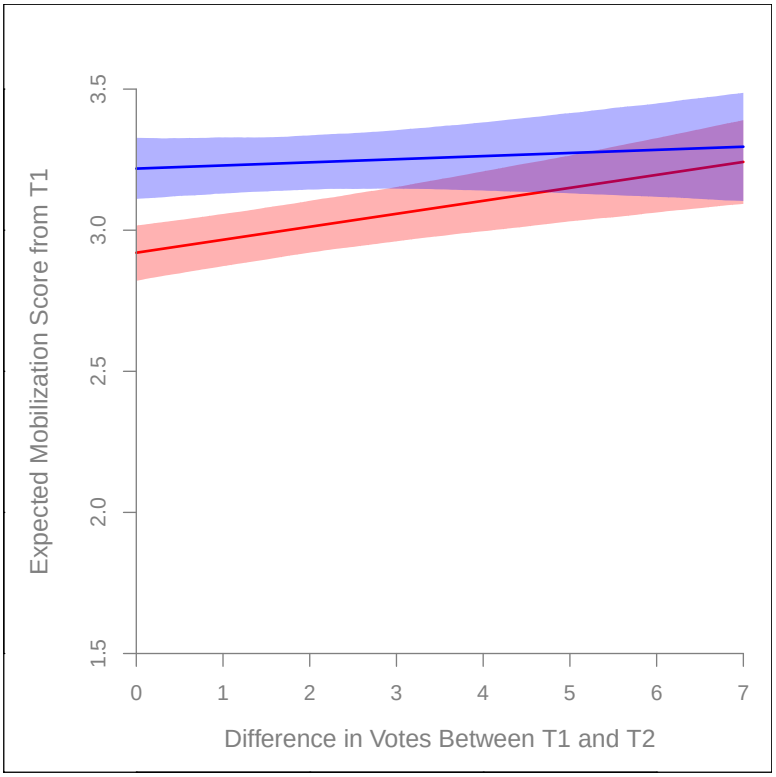
Table 2: Regression Results: The Relationship Between Dominance and Ability to Mobilize

	Dependent variable: Mobilization
Intercept	2.684^{***}
	(0.040)
Δ_k	0.046^{***}
	(0.013)
Party Member	0.144^{***}
	(0.034)
C_{ij}	0.298^{***}
	(0.053)
$\Delta_k \times C_{ij}$	-0.035^{*}
	(0.021)
Q_{ij}	0.092^{**}
	(0.042)
pD	3741.8
DIC	15410.3
Note:	$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Results report estimates from a 3750 posterior simulations from a regression model estimated in a Bayesian framework through Markov Chain Monte Carlo (MCMC) with 3 chains and diffuse priors on all parameters, using the program `JAGS`. Standard deviations of the posteriors on the respective parameters are given in parentheses. Statistical significance in the model is given with respect to the posterior distribution. In particular, let $\widehat{\pi}$ be a vector of values drawn from the posterior distribution of a parameter of interest. Then, we define $\underline{\pi} = 2 * P(\widehat{\pi} < 0)$. The deviance information criterion (DIC) is a measure of fit that is defined as the sum of one-half of the estimated variance of deviance (pD) and the expected value of the deviance. The lower value of DIC is taken to be a better fit, with pD entering as a penalty for overfitting the data.

While table 2 provides regression results, figure 3 presents the relationship between the degree of dominance of $T1$ and her capacity to mobilize. Once again, the simulations are based off of 90 percent predictive intervals, assuming that the influencer is a party member and that the respondent is a co-ethnic, thereby controlling for two critical correlates of one's ability to mobilze.

Figure 3: The Relationship Between Dominance and Ability to Mobilize



Two points are evident from this figure. First, the capacity to mobilize non-co-ethnics converges to the capacity converge co-ethnics for $T1$ as a function of dominance. Second, the capacity to mobilize co-ethnics does not really change as a function of dominance; the convergence results from a greater capacity to mobilize non-co-ethnics when $T1$ is dominant. We infer from these results that the relative popularity of an influencer is a function of her greater capacity to mobilize and influence non-coethnics. Furthermore, these results suggest that co-ethnics are relatively easy to mobilize, but greater skill is required to mobilize non-coethnics, and this is exactly how more dominant influencers distinguish themselves.

7.2 Willingness to Help

We also seek to understand how the degree of dominance of $T1$ predicts that influencer's willingness to assist non-coethnics, and as a result maintain inclusive influence networks.

As discussed in detail in Schneider and Sircar (2015), estimating an individual's willingness to help or target benefits is not straightforward. In principle, an influencer may help any citizen, but they prefer to help certain people *first*. In order to model this process, our procedure asked influencers to *rank*, from 1 to 12, the voters they would prefer to help in order of priority.

Before we move to the statistical model, two things need to be kept in mind as we interpret this rank data. First, because the ranks must be arrayed from 1 to 12, the mean rank must be 6.5. Second, it is important to remember that a *lower* rank is evidence of a greater bias, so if co-ethnics have a lower average rank, there is a bias towards the co-ethnic population. Because the mean rank if fixed, the regression model must address this "loss of a degree of freedom." This insight is incorporated into the regression protocol. Because the average ranking of respondents is fixed at 6.5, every predictor in a regression is centered at its mean value at the level of the influencer (e.g., adjusting for the the mean level of co-ethnicity at the level of the influencer). As a result, the constant term in an ordinary least squares type regression will yield a mean of 6.5 on the intercept term.

Furthermore, adjusting the means at the influencer and polling booth level through random effects in a hierarchical model would make little sense here since the mean will be fixed (a voter-level random effect is still meaningful). On the other hand, influencer-level and polling booth-level variables may enter into the analysis when interacted with voter level or dyadic-level variables. Thus, in these models we explicitly interact the extent to which $T1$ is dominant with dyadic variables of interest, namely co-ethnicity and co-partisanship.[31]

To formalize this intuition, for some variable x, the notation x_{ijk} once again denotes the value of x for respondent i, influencer j, and polling booth k. Let $\tilde{x}_i = x_{ijk} - \overline{x}_j$ for each $j \in \mathcal{I}$, where $\overline{x}_J$ denotes $E(x_{ijk}|j = J)$. Then given a vector of predictors $\mathbf{X}_{ijk}$ (without the constant), let $\tilde{\mathbf{X}}_i$ denote the vector where each of the elements is transformed according to the transformation described above. Let $rank_{ijk}$ denote the rank given to respondent i by influencer j in polling booth k, with C_{ij} as a binary variable denoting co-ethnicity between respondent i and influencer j and Δ_k as the difference in votes between $T1$ and $T2$ in polling booth k, as before. As before,

[31] We may also formulate this model as an ordered logit or probit, given the scale. For ease of fitting and interpretation, we have kept the model as linear model with normally distributed error.

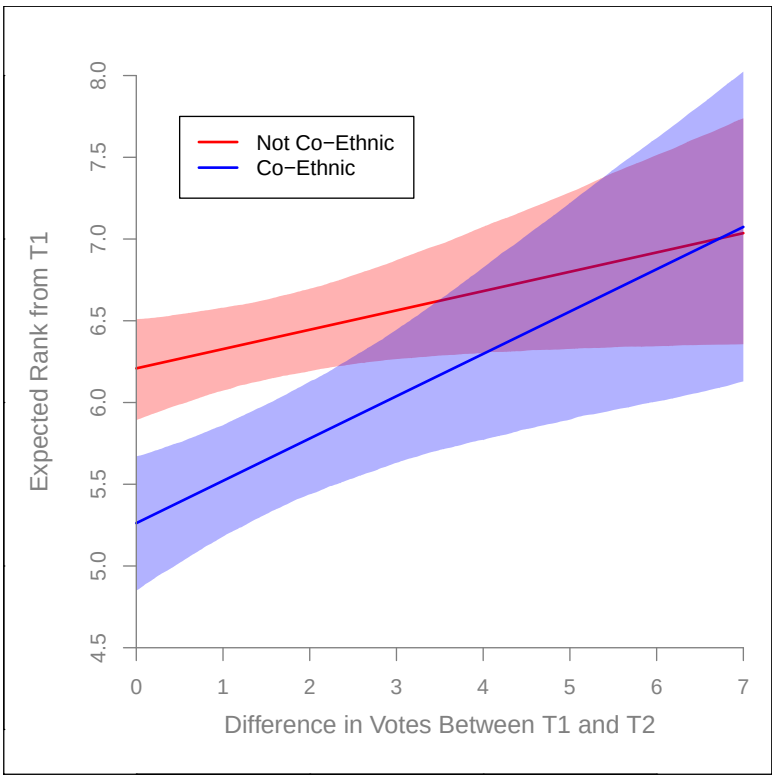
define Q_{ij} as a binary variable denoting co-partisanship between respondent i and influencer j and C_{ij} as a binary variable denoting coethnicity between respondent i and influencer j. We may now write the regression equation:

$$rank_{ijk} = \zeta + \tilde{\mathbf{X}}_i\beta + \alpha_i + \varepsilon_{ijk} \tag{7.2}$$

$$\alpha_i \sim \mathcal{N}(0, \sigma_i^2); \quad \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$$

where ζ is a constant term, β are fixed parameters estimated in the regression, and α_i is a random effect for voter i estimated through a hierarchical model. The above discussion should make clear that $\zeta = 6.5$. In our model $\mathbf{X}_{ijk} = (C_{ij}, \Delta_k * C_{ij}, Q_{ij}, \Delta_k * Q_{ij})$. Figure 4 displays the simulated 90 percent intervals of the mean ranks for co-ethnics and non-co-ethnics by $T1$ as a function of the level of dominance by $T1$ under the assumption that the respondent is a co-ethnic.

Figure 4: The Relationship Between Dominance and Willingness to Help



While table 3 provides the results of this regression, figure 4 presents the relationship between the degree of dominance of $T1$ and her willingness to help a voter. Much like in the case

Table 3: Regression Results: The Relationship Between Dominance and Willingness to Help

	Dependent variable: Help Rank
C_{ij}	-0.947^{***}
	(0.200)
$\Delta_k \times C_{ij}$	0.141^{**}
	(0.079)
Q_{ij}	-0.291
	(0.185)
$\Delta_k \times Q_{ij}$	0.118^{*}
	(0.072)
pD	4247.0
DIC	23291.3
Note:	$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Results report estimates from a 3750 posterior simulations from a regression model estimated in a Bayesian framework through Markov Chain Monte Carlo (MCMC) with 3 chains and diffuse priors on all parameters, using the program `JAGS`. Standard deviations of the posteriors on the respective parameters are given in parentheses. Statistical significance in the model is given with respect to the posterior distribution. In particular, let $\widehat{\pi}$ be a vector of values drawn from the posterior distribution of a parameter of interest. Then, we define $\underline{\pi} = 2 * P(\widehat{\pi} < 0)$. The deviance information criterion (DIC) is a measure of fit that is defined as the sum of one-half of the estimated variance of deviance (pD) and the expected value of the deviance. The lower value of DIC is taken to be a better fit, with pD entering as a penalty for overfitting the data.

of ability to mobilize, these analyses provides evidence that as the influencer becomes more dominant, he displays less bias towards his coethnics. As seen in the figure, there is convergence in the influencer's behavior towards co-ethnics and non-coethnics as $T1$ becomes more dominant. While the fact that both the co-ethnic and non-coethnic curve are upward-sloping may seem odd, it is important to remember that it is assumed that the voter is a co-ethnic. Thus, the upward slope indicates that the impact of co-partisanship also diminishes as the influencer become dominant. This points to a more general phenomenon: when the dominant influencer prefers to build a large coalition, he may forsake targeting biases predicated upon ethnicity.

8 Robustness

These results demonstrate that behaviors related to influence, namely ability to mobilize and willingness to help, correlate with the degree to which influencers are dominant. This adds to the evidence reviewed in the previous section, and strengthens our finding that when $T1$ influencers are locally dominant, they have incentives to maintain or build networks that are relatively inclusive vis-à-vis ethnicity.

8.1 Did We Manage to Select Influential Individuals?

For this conclusion to hold, we first need to demonstrate that we have selected truly influential individuals at the local level. A potential concern here might be that we were only able to interview marginally influential individuals.

The number of votes received by each of our selected influencers however confirm that the T1 influencers we managed to interview were on average seen as very influential by voters in our sample: on average, close to 6 (out of a total of 12) voters in each PBA chose these individuals as being the most likely to influence them, on a list that contained an average of over 15 names.

A second and interrelated concern may be that the individuals on our initial "long list" of influencers may be influential in general, though not necessarily *during elections* or in politics. To alleviate this concern, it is necessary to return to the process through which we selected our influencers. It is first the case that a number of influencers relevant to political discussions made it to that list, insofar as voters were asked to list individuals in response to a question about political and social discussions (*when it comes to social issues, whose opinions do people listen to the*

most around here?). Second, and more importantly, voters were specifically asked which of these individuals were likely to influence them in reference to an electoral decision.[32]

A third potential concern relates to the dependent variables we rely on in the second part of the paper (self-report of ability of mobilize and willingness to help) and whether they do effectively relate to influence, as we have argued they do. The concern here is that the response of influencers may not be a good indicator of their ability to be influential in the real world, and especially so in politics. This may be because influencers overestimate their ability or because these variables are not correlated to influence. In appendix A, we alleviate this concern by showing that *voters* see as influential individuals who later declare being able to mobilize them and willing to help them. While this is admittedly not a causal case, it at least shows that these variables are related one to another in a strong and significant way, and accordingly that voters see as influential individuals who declare being able to influence them. This provides a form of validation for these variables.

8.2 Are We Measuring Influence or Social Closeness?

To further enhance the robustness of these results, it is also useful to show that the effect of dominance on the maintenance of inclusive networks is not simply due to differences in social connectedness, but are instead really due to real differences in ability between "dominant" and "contested" influencers. It could indeed be that influence is merely a function of social connectedness, and that dominant influencers are naturally better connected to non-coethnics, and hence that they do not need to emphasize their general abilities - as we have assumed they do - in order to maintain inclusive networks. If that was true, we would expect to see intermediaries influence individuals with whom they are strongly connected, whether or not they are coethnics.

To show that this is *not* the case, we rely on an additional survey item. In addition to the two influence-related questions listed above, our instrument also asked influencers to report the extent to which they are generally *close* to each respondent. This measure of social closeness varied from 0 to 3 (3 being highest). As seen in the analyses presented in appendix D, models similar to the ones run in section 6.1 above show that the degree of local dominance of $T1$ does not predict how close the influencer is to co-ethnics vis-à-vis non-coethnics. This is contrary to

[32]Besides, note that our T1 and T2 influencers performed much better than randomly sampled citizens on average when it came to political mobilization for a meeting (as shown in(Chauchard and Sircar, 2017)). This suggests that these actors dabble in politics in real life, beyond our interview.

its predictive power on ability to mobilize and willingness to help.

This confirms that the two self-reports analyzed in section 6 do not simply proxy for social closeness, but rather measure influencers' underlying ability. These results in addition suggest that in an ethnically stratified society, influencing co-ethnics requires less skill, since all influencers, including the comparatively more mediocre ones, are equally connected to their coethnics; as a result, the most dominant influencers distinguish themselves by their ability to influence non-coethnics to whom they do not feel especially close.

9 Discussion

This intuition is relevant to politics in many ethnically-diverse contexts, far beyond Bihar and North India. As noted above, our model is based on a rather minimal set of scope conditions and assumptions, which apply to many ethnically diverse democracies. Diversity at the local level is common, including in places in which groups tend to be territorially concentrated. While the spatial concentration of ethnic groups in most of Africa implies that diversity at the local level may be harder to find that there than in India, this is not necessarily true of urban areas (Habyarimana et al., 2007). Contrary to what primordialist authors have long assumed, there are also in most ethnically-diverse contexts voters ready to join multi-ethnic coalitions (Madrid, 2008; Arriola, 2013). In India, our evidence suggests that this may derive from the fact that groups at the bottom of the caste hierarchy have a significantly lower preference for their coethnics than groups at the top.[33] But there may be other reasons that push voters to join inclusive networks elsewhere, the absence of leaders within the group (Barany, 2002) or strategic considerations related to the size of their group (Posner, 2005; Horowitz and Long, 2016). The existence of political competition - our second scope condition - is of course equally common, as competitive elections are now the rule in most emerging democracies. The subsequent development of a local-level market for intermediaries, while less discussed in the comparative literature on political networks, is well-documented across India (Manor, 2000; Berenschot, 2012; Björkman, 2014); this is presumably true elsewhere. Finally, our assumption that voters have a coethnic bias has been abundantly documented in studies run in Asia, Africa and America. While these results should be replicated and extended, we accordingly believe these analyses to

[33] Our data shows that an an upper caste person is 17% more likely to nominate $T1$ or $T2$ if one of these influencers is a coethnic, but someone who is not upper caste is only 4% more likely to nominate them if they are coethnic.

be relevant across many contexts.

These findings thus contribute to at least three literatures in comparative politics. They first contribute to the comparative literature on ethnic politics. Counter-intuitively enough, we show that voters are often mobilized by non-coethnics, including in strongly ethnicized environments such as Bihar. The argument we advance to explain this fact emphasizes the role of electoral incentives on the formation of multi-ethnic coalitions. When politics is competitive, we show that the most prominent local intermediaries differentiate themselves by targeting non-coethnics. This helps explain why ethnic preferences do not always transform into ethnic votes, namely because local-level influence networks are often multi-ethnic in nature. This also, more generally, shows that the development of inclusive coalitions does not always have to do with institutions (Posner, 2005). In our framework, it is changes in levels of political competition, because they create new incentives for local actors to build larger networks, which lead to more inclusive forms of politics.

Our analyses also add to the empirical literature on brokerage and clientelism. While much has been written about brokers over the past few years, most of it has been concerned with specific subtypes of brokers such as party agents (Calvo and Murillo, 2004), locally elected officials (Dunning and Nilekani, 2013; Schneider, 2014) or slum association leaders (Auerbach and Thachil, 2017). Our research design avoids these restrictions, as it allows us to more generally identify influential individuals at the local level, whether or not they fit in a pre-defined category. In line with recent work on brokerage in urban India (Auerbach and Thachil, 2017), we explore the strategy of these intermediaries and the contours of the coalitions they assemble - specifically, the extent to which coethnicity is the main principle according to which these coalitions are organized. By matching real intermediaries with real voters in a cross-referencing exercise, we document bonds between real actors. This allows us to identify which of these real intermediaries are locally dominant, and which are influential among more than their coethnics. We also make theoretical contributions to this literature. Much of the literature has not engaged with how political competition at the systemic level impacts the incentives of individual intermediaries, or the extent to which competition between these intermediaries shapes observable outcomes. By taking into account political competition, and the subsequent market for influencers that develops, we provide a novel theoretical argument as to why local influencers should have incentives to build multi-ethnic coalitions.

Finally, these analyses contribute to the country-specific literature on elections and political

behavior in India (Srinivas, 1955; Kothari, 1964; Chandra, 2004; Wilkinson, 2007). To the best of our knowledge, and in spite of the fact that characters such as our local influencers are often assumed to play an outsized role in Indian elections, our study constitutes the first quantitative attempt to identify prominent influencers across a large sample of villages in rural India and to document their strategies. We show that some influencers at the most local level of politics tend to lead multi-ethnic coalitions. As a result, voters are frequently mobilized and influenced by individuals who do not belong to their own caste group. This suggests that steady increases in levels of political competition in the country over the past thirty years - much of it through the rise of caste-based politics - have in some cases had an unlikely consequence: the creation of inclusive networks at the very local level.

References

Arriola, Leonardo R. 2013. "Capital and Opposition in Africa: Coalition Building in Multiethnic Societies." *World Politics* 65(2):233–272.

Auerbach, Adam and Tariq Thachil. 2017. "Capability, Connectivity, and Co-Ethnicity: The Origins of Political Brokerage in India's Urban Slums." Manuscript.

Auyero, Javier. 2001. *Poor People's Politics*. Duke University Press.

Baldwin, Kate. 2013. "Why Vote with the Chief? Political Connections and Public Goods Provision in Zambia." *American Journal of Political Science* 57(4):794–809.

Barany, Zoltan. 2002. "Ethnic Mobilization Without Prerequisites: The East European Gypsies." *World Politics* 54(3):277–307.

Berenschot, Ward. 2012. *Riot Politics: Hindu-Muslim Violence and the Indian State*. New York: Columbia University Press.

Björkman, Lisa. 2014. "You can't buy a vote: Meanings of money in a Mumbai election." *American Ethnologist* 41(4):617–634.

Calvo, Ernest and Maria Victoria Murillo. 2004. "Who Delivers? Partisan Clients in the Argentine Electoral Market." *American Journal of Political Science* 48(4):864–871.

Carlson, Elizabeth. 2015. "Ethnic Voting and Accountability in Africa: A Choice Experiment in Uganda." *World Politics* 67(2):353–385.

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

Chandra, Kanchan. 2006. "What Is Ethnic Identity and Does It Matter?" *Annual Review of Political Science* 9(1):397–424.
URL: *http://www.annualreviews.org/doi/abs/10.1146/annurev.polisci.9.062404.170715*

Chauchard, Simon. 2016. "Unpacking Ethnic Preference: Theory and Microlevel Evidence from North India." *Comparative Political Studies* 49(2):253–284.

Chauchard, Simon and Neelanjan Sircar. 2017. "How Partisan is Political Brokerage in Rural India?" Manuscript.

Dunning, Thad and Janhavi Nilekani. 2013. "Ethnic Quotas and Political Mobilization: Caste, Parties, and Distribution in Indian Village Councils." *American Political Science Review* 107(1):35–56.

Fearon, James D. and David D. Laitin. 1996. "Explaining Interethnic Cooperation." *American Political Science Review* 90(4):715–735.

Ferree, Karen. 2011. *Framing the Race in South Africa: The Political Origins of Racial-Census Elections*. New York: Cambridge University Press.

Gay, Claudine. 2002. "Spirals of Trust? The Effect of Descriptive Representation on the Relationship Between Citizens and Their Government." *American Journal of Political Science* 46(4):717–732.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner and Jeremy Weinstein. 2007. "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101(4):709–725.

Horowitz, Donald L. 1985. *Ethnic Groups in Conflict*. University of California Press.

Horowitz, Jeremy and James Long. 2016. "Strategic voting, information, and ethnicity in emerging democracies: Evidence from Kenya." *Electoral Studies* 44:351–361.

Huber, John and Pavithra Suryanarayan. 2016. "Ethnic Inequality and the Ethnification of Political Parties: Evidence from India." *World Politics* 68(1):149–188.

Jaffrelot, Christophe and Sanjay Kumar, eds. 2009. *Rise of the Plebeians?: The Changing Face of the Indian Legislative Assemblies*. New Delhi: Routledge.

Joshi, Bhanu, Ashish Ranjan and Neelanjan Sircar. 2015. "What Do Bihar's Voters Want? Notes from the Field." Centre for Policy Research Working Paper.

Kothari, Rajni. 1964. "The Congress 'System' in India." *Asian Survey* 12(4):1161–1173.

Madrid, Raúl L. 2008. "The Rise of Ethnopopulism in Latin America." *World Politics* 60(3):475–508.

Manor, James. 2000. "Small-Time Political Fixers in India's States: 'Towel over Armpit'." *Asian Survey* 40(5):816–835.

Pepinsky, Thomas B., R. William Liddle and Saiful Mujani. 2012. "Testing Islam's Political Advantage: Evidence from Indonesia." *American Journal of Political Science* 56(3):584–600.

Posner, Daniel N. 2005. *Institutions and ethnic politics in Africa*. New York City: Cambridge University Press.

Schneider, Mark. 2014. "Does Clientelism Work? A Test of Guessability in India." Unpublished.

Schneider, Mark and Neelanjan Sircar. 2015. "Whose Side Are You On? Identifying Distributive Preferences of Local Politicians in India." CASI Working Paper, 15-01.

Shepsle, Kenneth and Alvin Rabushka. 2008. *Politics in Plural Societies: A Theory of Democratic Instability*. Pearson Education.

Srinivas, M. N. 1955. The Social Structure of a Mysore Village. In *Village India: studies in the little community*, ed. McKim Marriott. University of Chicago Press pp. 1–35.

Stokes, Susan C., Thad Dunning, Marcelo Nazareno and Valeria Brusco. 2013. *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*. New York: Cambridge University Press.

Szwarcberg, Mariela. 2015. *Mobilizing Poor Voters: Machine Politics, Clientelism, and Social Networks in Argentina*. New York: Cambridge University Press.

Vaishnav, Milan. 2017. *When Crime Pays: Money and Muscle in Indian Politics*. Yale University Press.

Wilkinson, Steven I. 2007. Explaining Changing Patterns of Party-Voter Linkages in India. In *Patrons, Clients, and Policies: Patterns of Democratic Accountability and Political Competition*, ed. Herbert Kitschelt and Steven I. Wilkinson. New York: Cambridge University Press.

Witsoe, Jeffrey. 2013. *Democracy against Development: Lower-Caste Politics and Political Modernity in Postcolonial India*. University of Chicago Press.

A The Relationship Between Influencer Measures and Selection of Influencer

	Dependent variable: Influencer Chosen (Logit)
Intercept	-1.249^{***}
	(0.031)
Closeness	0.309^{***}
	(0.034)
Ability to Mobilize	0.297^{***}
	(0.034)
Willingness to Help	-0.098^{***}
	(0.032)
Note:	$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Note that the predictor variables have been standardized by subtracting the mean and dividing by the standard deviation of the predictor to guarantee that they are on the same scale. The "closeness variable" is defined in the robustness section.

B Comparing Voter and Influencer Attributes

	T1 "most popular"	T2 "2cd most popular"	Random Voter
Party Member (%)	38	33	21
Upper Caste (%)	25	20	14
Yadav (%)	25	27	22
Other OBCs (%)	37	41	40
SC (%)	8	8	17
Under Class 5 (%)	7	12	32
Class 5 Pass (%)	6	6	14
Class 8 Pass (%)	7	12	15
Class 10 Pass (%)	26	23	18
Class 12 Pass (%)	53	48	21
Pucca House (%)	84	75	58
Number of Rooms (Avg.)	5.13	5.16	3.35
Age (Avg.)	50.42	51.47	45.35
Persons visiting/week	29	19	N/A
N	179	179	2148

C Relationship Between Influencer Characteristics and Dominance

	Coefficient	p-value
Elected	0.790	0.015
Party Member	0.260	0.424
Upper Caste	0.494	0.173
Social Worker	-0.463	0.146
Pucca House	0.319	0.444
Age	-0.002	0.824
Class 12 Pass	0.537	0.086

The above table reports a series of bivariate linear regressions with the difference in the number nominations between $T1$ and $T2$ as the dependent variable. Each row corresponds to a binary predictor (as listed in the first column) used in the bivariate regression, with associated regression coefficient and p-value.

D The Absence of a Relationship Between Dominance and Social Closeness

Figure 5: Simulated Effects from Regression

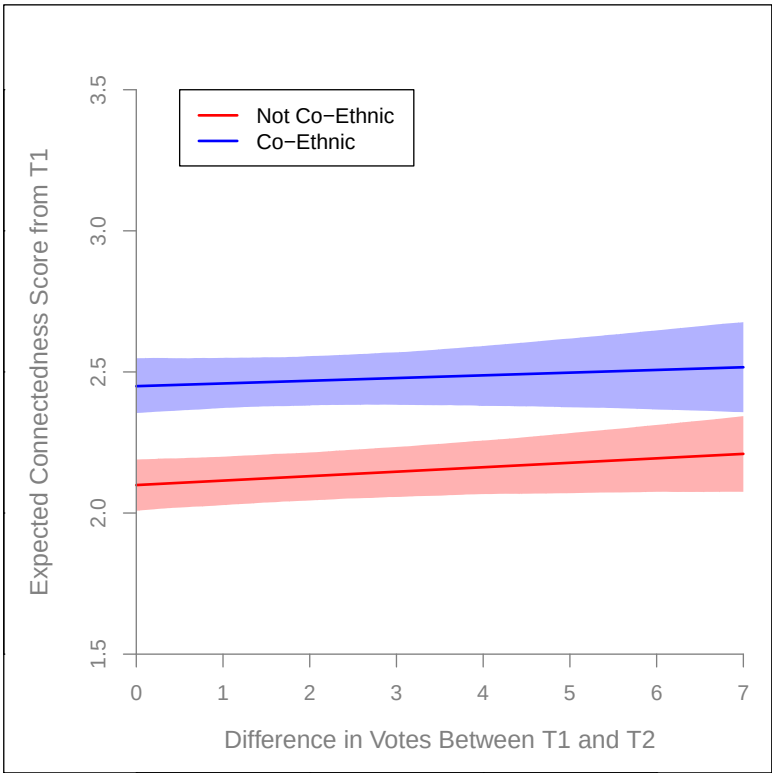


Table 4: Regression Results on Closeness

	Dependent variable: Closeness
Intercept	2.030^{***}
	(0.042)
Δ_k	0.016
	(0.012)
Party Member	0.072
	(0.050)
C_{ij}	0.350^{***}
	(0.037)
$\Delta_k \times C_{ij}$	-0.006
	(0.014)
Q_{ij}	-0.004
	(0.028)
pD	4385.5
DIC	12345.9
Note:	$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Results report estimates from a 3750 posterior simulations from a regression model estimated in a Bayesian framework through Markov Chain Monte Carlo (MCMC) with 3 chains and diffuse priors on all parameters, using the program `JAGS`. Standard deviations of the posteriors on the respective parameters are given in parentheses. Statistical significance in the model is given with respect to the posterior distribution. In particular, let $\widehat{\pi}$ be a vector of values drawn from the posterior distribution of a parameter of interest. Then, we define $\underline{\pi} = 2 * P(\widehat{\pi} < 0)$. The deviance information criterion (DIC) is a measure of fit that is defined as the sum of one-half of the estimated variance of deviance (pD) and the expected value of the deviance. The lower value of DIC is taken to be a better fit, with pD entering as a penalty for overfitting the data.