Beyond Coethnicity: Political Influencers in Ethnically Diverse Societies[*]

Neelanjan Sircar   Simon Chauchard

CPR, Delhi   Dartmouth College

August 31, 2017

Abstract

In India, as in many other ethnically diverse democracies, ethnicity plays an oversized role in voters' preferences. Over the past decade, empirical studies have shown that voters often have a strong preference for coethnics. Yet votes are only partially determined by these individual-level preferences. This is because voting decisions are also the product of coordination mechanisms and influence by local political intermediaries (hereafter referred to as "influencers"). In this article, we explore the extent to which influence networks at the local level are strictly organized around ethnic lines when ethnic differences are salient - in other words, whether influencers and voters are coethnics. We argue that local influencers in ethnicized societies are often likely to be "inclusive influencers". This is because diversity at the local level, if combined with political competition and the existence of a market for local intermediaries, provides strong incentives for dominant influencers to reach beyond their coethnics, in an effort to differentiate themselves from other intermediaries. To test this argument, we deploy a novel empirical strategy to identify influencers across a large sample of villages in rural Bihar (India). We then rely on an innovative cross-referencing exercise between influencers and voters to compare the networks of locally dominant influencers vs. other influencers. Results support our argument: locally dominant influencers generally maintain inclusive networks, contrary to other influencers. These findings contribute to the empirical literature on brokers and political influence. They also help explain why ethnic preferences do not always transform into ethnic votes: namely, because the 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

In most developing democracies, local-level actors mediate interactions between political elites and voters (Hicken, 2011; Stokes et al., 2013). These actors draw crowds to meetings, canvass and organize political events (Björkman, 2014; Szwarcberg, 2015; Muñoz, 2014). They also play a role between elections, as they serve as intermediaries between elected officials and citizens, helping citizens access the state and assisting elected officials with the implementation of their policies (Manor, 2000; Krishna, 2002; Van de Walle, 2007; Berenschot, 2012; Baldwin, 2013; Bussell, 2014; Auerbach, 2016; Kruks-Wisner, Forthcoming). This places these informal actors in a privileged position to influence voters ahead of elections. Their methods may vary: from legitimate discussion and canvassing to more problematic strategies such as gift-giving, or even coercion. But regardless of the method (or more likely, the mix of methods) they employ, these local-level actors will be the main actors directly influencing voters on the ground, simply because they are the closest to them.

The ability of different types of local intermediaries to influence the choices of voters during elections is our focus in this article. Given our substantive interest on this aspect of their work, we refer to these actors as "influencers" hereafter. Contrary to what much of the literature on clientelism has argued, influencers are not always long-term partisan actors (Holland and Palmer-Rubin, 2015; Larreguy, Montiel and Querubin, 2017). In a related paper (Chauchard and Sircar, 2017), we show that ethnic ties in some cases shape the contours of local-level influence networks much more strongly than partisan ones do. Building on this finding, we explore the extent to which local influence networks are *strictly* organized around ethnic lines in highly ethnicized societies, i.e., the extent to which influencers and those whom they influence belong to the same ethnic group.[1]

The identity of local influencers and the ethnic composition of their network are 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

---

[1]In this regard, we are in this manuscript more interested in *who* influencers are than in *how* they influence voters.

with voters. When the influence networks built around these local influencers are multi-ethnic in nature, 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 networks are strictly organized around ethnic lines, thus, allows us to document the extent to which ethnic preferences should be expected to transform into ethnic votes.

We explore this question 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 the composition of influence networks in this context is important for a number of reasons. Insofar as Bihar resembles much of rural India, it first allows us to better understand how candidates and party elites organize to mediate the vote of over 600 million rural Indian voters with extremely diverse ethnic identities. In particular, it allows us to document the profiles of the individuals upon which higher-level elites rely to reach, mobilize and convince voters. Whether or not the local networks maintained by these actors are mono-ethnic is substantively important because it tells us much about the strength of ethnic boundaries and about the ability of local political actors to reach out to and mobilize members of other ethnic groups. The identity of the individuals who mobilize and influence voters on the ground is also normatively important, as a strict division of local influence networks along ethnic lines may lead to more rigid and more divisive forms of politics.

As we develop our argument, we distinguish between two stylized types of local influencers that may exist in an ethnicized polity such as Bihar. The first type may be described as an "ethnic influencer" - that is, a local intermediary who solely mobilizes and influences her coethnics. A second type of local influencer may be referred to as an "inclusive influencer". Inclusive influencers do not restrain themselves to mobilizing and influencing coethnics. Instead, they help, and later attempt to influence and mobilize, non-coethnics.

The literature on ethnic politics has often assumed that local influencers should be ethnic influencers (Shepsle and Rabushka (2008); Horowitz (1985); Fearon and Laitin (1996)). In this manuscript, however, we contend that some influencers should be expected to maintain inclusive networks under a limited set of scope conditions. Namely, wherever diversity at the local level is combined with genuine political competition, our theoretical model predicts the most prominent influencers should be "inclusive influencers". Political competition leads political elites to seek out the best influencers on the ground, i.e., those individuals who can influence

the greatest number of citizens (arguably, because they solve the problems of the greatest number of citizens between elections). Where political competition exists, a market for these individuals thus develops. Competition *at the local level* between influencers need not however be homogenous across locations. In some cases, one of these actors becomes a "dominant influencer" at the local level – that is, his quality far surpasses that of other local influencers[2], leading him to maintain a comparatively larger network. In other cases, several "contested influencers" with similar abilities and network sizes coexist. The central intuition of our 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".

To test this hypothesis, we identify influencers across a large sample of villages in rural Bihar. To do this, we ask a random sample of voters to select the two individuals that would be the most likely to influence them ahead of elections. The relative popularity of these two individuals indicates whether there exists a dominant influencer at the local level or whether several influencers compete at the local level. We later interview these two influencers. 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. Concretely, we ask these selected influencers to report 1. how able they would be to mobilize each of these voters for a political event and 2. how ready they would be to help each selected voter. These are two tasks that intermediaries commonly engage in.[3] This data, combined with voters' own choice of influencers, allows us to describe and compare the networks of "dominant" vs. "contested" influencers. It also allows us to establish how strongly each type of influencer ("dominant" vs. "contested") is able to influence each type of voter (coethnic vs. non-coethnic).

Results support our argument, as they show that dominant influencers generally maintain more inclusive networks. First, we demonstrate that intermediaries whom voters see as the most likely to influence them are precisely those who maintain multi-ethnic networks (i.e. they are "inclusive influencers"). Furthermore, we show that dominant influencers are more likely to help and mobilize non-coethnics than contested influencers.

These analyses contribute to at least three prominent literatures in comparative politics.

---

[2]This may be because he employed a more successful strategy or was more talented to begin with. Our model is agnostic as to what leads to this dominance.

[3]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.

They first contribute to the comparative literature on ethnic politics. Authors in this lterature have frequently assumed that local influencers cater primarily to their own ethnic groups in highly ethnicized environments. Our paper demonstrates that the most prominent influencers actually have incentives to move beyond their own ethnic group. In fact, the more ethnicized a society, the more important it is for the most skilled influencers to differentiate themselves by catering to 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.

These analyses also contribute to the literature on brokerage and clientelism. So far, this 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. In addition, the literature has focused on very specific categories of intermediaries, 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), as opposed to political influence more generally. In this paper, we develop a cross-referencing empirical exercise that identifies a broad class of influencers, characterizes their relative popularity, and generates detailed information interactions between influencers and individual voters to address this lacuna. This novel data allows us to investigate the relationship between the dominance of an influencer and her ability to construct an multi-ethnic coalition.

Finally, while political brokerage has been a major topic of study in Indian political behavior (Srinivas, 1955; Kothari, 1964; Wilkinson, 2007), there has so far been no attempt to characterize differences in the coalitions built by local-level intermediaries. Since dominant influencers cater across ethnic groups, we find that voters are frequently influenced by individuals who do not belong to their own 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.

2 Theory: Influence Networks in Ethnicized Societies

To what extent are influence networks built along ethnic lines? The literature on ethnic politics has so far remained relatively silent 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, relative ethnic homogeneity at the local level simply makes it implausible that voters could enter the network of a non-coethnic.

Where and when political brokerage 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.[4]

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[5] - under a minimal number of scope conditions.

---

[4]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

[5]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.

2.1 Scope Conditions

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

2.1.1 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 entirely 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.[6] 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

---

[6]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.

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).[7] 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.

2.1.2 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 however 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.[8] 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

---

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

[8]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.

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 an additional clientele 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.[9] As noted above, competition at the local level between influencers need not however be homogenous across locations.[10] 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[11], leading him to maintain a comparatively larger network. In other cases, several "contested influencers" of relatively similar abilities and network sizes may coexist.

2.2 A Model of Influence 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 (presented in detail in Appendix A) 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". While our formal model properly details these predictions, the following paragraphs provide basic intuitions on our argument.

---

[9]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.

[10]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*.

[11]This may be because he employed a more successful strategy or was more talented to begin with. We are agnostic in this manuscript as to what leads to this dominance.

2.2.1 General Setup

Suppose first that there are multiple influencers at the local level. Whether a specific voter's electoral choices are in turn influenced by a specific influencer depends on a combination of two factors: the ethnicity of the influencer and her quality irrespective of ethnicity. For ease of understanding, her quality may here be best understood as her ability to solve the voter's problems; an influencer that has successfully assisted a voter in the past may be seen as more likely to influence them in the present.[12] All else held constant, voters unsurprisingly prefer coethnic influencers and high-quality influencers (i.e., those who solve their problems).

Each influencer allocates her finite time and resources as she wishes between coethnics and non-coethnics. This allocation strategy defines the composition of their influence network. Allocating more of her finite time and resources towards solving the problems of her coethnics overall helps her include coethnics in her network, and eventually influence them, and vice-versa.

Because we are in an ethnically divided society, influencers are assumed to have a strategic advantage in influencing their own co-ethnics. This concretely yields the following assumptions. First, effort spent towards influencing coethnics has a greater marginal impact than effort spent towards non-coethnics. Second, no influencer can credibly maintain a network which includes a smaller proportion of the coethnic population than that of the non-coethnic population. It does not however straightforwardly follow that they should always focus on coethnic voters. This is because they can only probabilistically hope to influence them and because co-ethnic voters may alternatively choose to join another network. That is, given the presence of multiple influencers, they may find someone else to solve their problems, who will then be in a position to influence them. Furthermore, their relative quality may make it preferable for them to target non-coethnics.

In this context, the extent to which influencers are willing to add non-coethnics to their network (that is, to help and eventually influence them) depends on the degree of contestation on the influencer market *at the local level*. As noted above, the market for influencers is contested at the local level, since different influencers strive to help voters, and eventually gain influence on them.[13] We define quality here as an influencer's ability to help and in time influence others given her limited time and resources.[14] A high-quality influencer is, one can imagine,

---

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

[13]In this article, we simply take the quality of the competition that influencers face at the local level as exogenous.

[14]This quantity is independent of her ethnic identity.

more charming, charismatic or efficient than her competitor at the local level, and is as a result likely to develop a larger influence network, time and resources held constant. The degree of contestation at the local level in turn describes the difference in quality between the various influencers that exist at the local level.[15] As noted above, contestation at the local level between influencers is likely to be heterogenous. 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, leading him to maintain a comparatively larger network. In other cases, several "contested influencers" of similar ability may coexist. We refer to an influencer whose quality is much higher than her competition as a "dominant influencer", as opposed to a "contested influencer", whose quality is similar or only slightly different from that of her competition.

2.2.2 First Case: Dominant Influencers

Several configurations are at this point possible. Suppose first that an influencer is a "dominant influencer". Because of this dominance, that individual can overall include more people in her network than her competitors for each minute of effort she spends on assisting voters. She is of course eager to do so, since she is interested in increasing the size of her network in order to gain additional rents and become a sought-after local influencer in the eyes of political elites. When this is the case, we predict that that "dominant influencer" should pick an inclusive strategy and allocate a greater share of her time and resources towards non-coethnics in order to increase the size of her network. Note that this choice of an inclusive strategy is not merely a consequence of numerical constraints on the size of her group, since it can be assumed that she always has additional coethnics to include.[16]

That a dominant influencer opts for an inclusive strategy makes sense if her competitor is a non-coethnic; since her quality is much higher than that of her competitor, she should expect to be able to attract many of her competitor's own coethnics, despite their bias. At the same time, this inclusive allocation of resources would not constitute a great risk; she is very unlikely not lose any of her own coethnics to this competitor, since she is both more talented AND benefits from a coethnic bias among voters from her own group. Choosing an inclusive strategy also makes sense if her competitor is a coethnic. Including non-coethnics is in that case under-

---

[15]In this section, we focus on a case in which there are two influencers, though this is generalized to other cases in our model in appendix A.

[16]Our data below confirms the credibility of this assumption, as even extremely dominant influencers in our sample are not predicted to manage to include more than 80 percent of their coethnics in their network.

standable, since she has a much greater ability to help (and hence influence) them than her co-ethnic competitor, who is of far inferior quality. Of course, redirecting resources towards non-coethnics to some extent reduces her ability to serve more coethnics. This however remains a preferable strategy in order to include the greatest number of individuals in her network. This is simply because the marginal impact of her being a far superior influencer is particularly large among non-coethnics. The intuition we derive from our model is here interesting, and worth emphasizing: since a low-quality competitor from her group performs almost as well as she does among her coethnics because of the latters' coethnic bias, her talents are better allocated elsewhere. In other words, when voters have a strong coethnic bias, the best way for a high-quality influencer to differentiate oneself from her competition may be to further direct her efforts towards non-coethnics.

2.2.3   Second Case: Contested Influencers

Let us now suppose now that our influencer is a contested influencer at the local level. Once again, this means that her quality is equal to or only slightly higher than that of her local competition.

In this case, her marginal ability to influence non-coethnics generally goes down, since she is not much better than her competition. As a result, she should be expected to return towards a less inclusive strategy. Interestingly, and as above, this does not depend on the identity of her competitor. Suppose first that her competitor is a non-coethnic. In this case, she is at a clear disadvantage among non-coethnics given her ethnicity, and her relative quality is not sufficient to allow her to overcome this disadvantage. As a result, she is probably better off allocating her time and resources assisting, and later trying to influence, her coethnics. She thus chooses an ethnic strategy.

Suppose now that her competitor is a coethnic. In this case as well, she is better off allocating most of her resources towards her coethnics. To understand why, it is important to remember that influencers can only credibly maintain networks which includes a larger proportion of the coethnic population than that of the non-coethnic population. In light of this assumption, an influencer should redirect their effort towards coethnics in order to maintain their coethnic advantage, relative to their competition. Since their competitor is equally talented and benefits from a similar coethnic advantage, they have no other option.

2.3 Main Hypothesis

In light of our review of these various cases, we expect "dominant influencers" to be more inclusive than "contested" ones.[17] Given the continuous (rather than binary) nature of our model, this practically implies that we test the following hypothesis:

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.

3   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.

3.1   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.[18] 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

---

[17]This is true regardless of the identity of their competition at the local level.

[18]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).

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."

3.2 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.[19] In each of these three districts, we then randomly selected three blocks.[20] In each of these blocks, we randomly selected polling booths using a variant of systematic random sampling.[21] The outcome of this process was a random sample of 179 PBAs over 9 blocks and 3 districts of Bihar.

---

[19]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.

[20]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.

[21]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.

3.3 Phase 1: Identifying Influencers

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

3.3.1 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:[22]

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?*

We deliberately posed questions that were *not* specific to politics, so as to generate as large a sample as possible.[23] 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 generating 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."[24] 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.

---

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

[23]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".

[24]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".

3.3.2 Selecting The Two Leading Influencers

Using official voters' lists, we randomly sampled twelve male voters in each PBA.[25] 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.[26] Specifically, we asked each respondent 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.[27] We will refer to the first chosen influencer as the *T*1 influencer. The second influencer (the *T*2 influencer) was the most popular influencer among those remaining on the list.[28,29]

We build several important variables based on the number of villagers who recognized our *T*1 and *T*2 as influential. The difference in popularity between *T*1 and *T*2 provides us with a

---

[25]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.

[26]In addition to picking two "real" influencers from our list, we also select a 3rd "benchmark influencer" in each booth. This third influencer (*T*3) simply was the head of household of a randomly drawn household within the PBA. We do not make use of data about *T*3 in this paper.

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

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

[29]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.

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

3.4 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,[30], 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).

In appendix B, 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.[31]

4 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

---

[30]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.

[31]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.

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.

4.1 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 C provides a detailed comparison between our two selected influencers (T1 and T2) and our randomly sampled male voters, which can be assumed 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 $T2$.[32]

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 bi-

---

[32]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.

ases 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.

4.2 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 $T2$ accordingly 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 D, we show that dominance of the $T1$ influencer is correlated to whether she is elected and has passed class 12.[33] 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)).[34] 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.

---

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

[34]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 $T2$
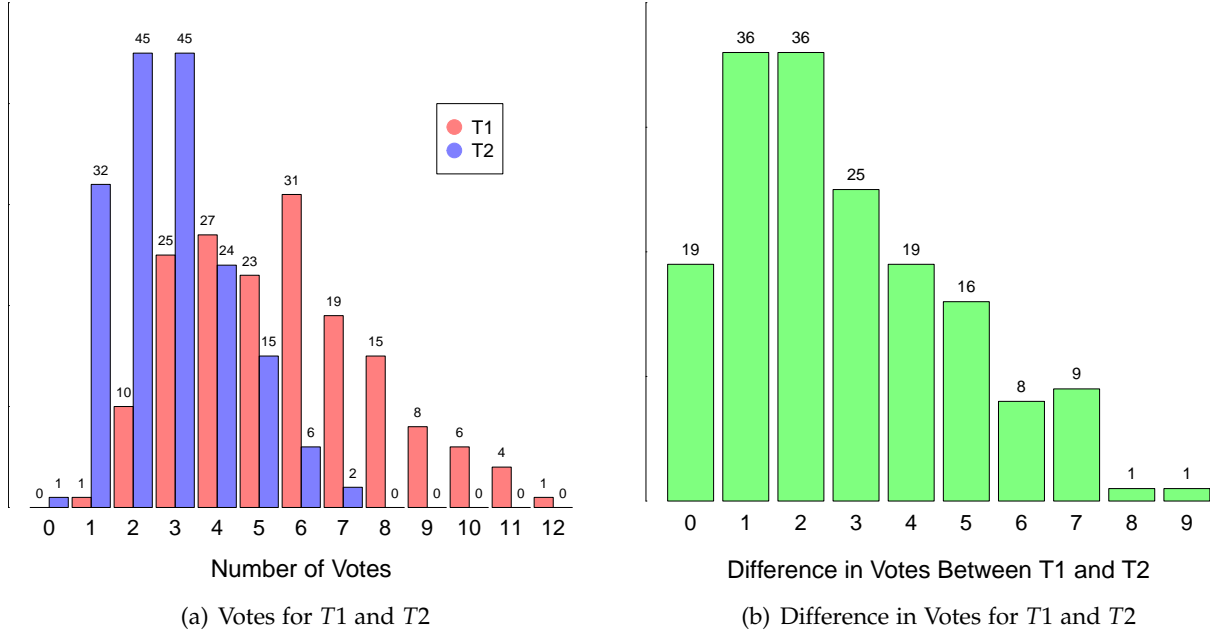
5   Tests: Do Voters Name Coethnics As Influencers?

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 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. 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 and C_{ij} be a binary variable taking the value of 1 when respondent i is a co-ethnic of influencer j. The core regression model is:

$$P(choose_{ijk} = 1) = logit^{-1}(\beta_0 + \beta_1\Delta_k + \beta_2 C_{ij} + \beta_3 C_{ij} * \Delta_k + \mathbf{Z}_i\gamma + \alpha_i + \alpha_j + \alpha_k) \qquad (5.1)$$

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

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.[35]

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 $\Delta \leq 0$). 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.

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.[36]

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 coali-

---

[35]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.

[36]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 $\hat{\pi}$ be a vector of values drawn from the posterior distribution of a parameter of interest. Then, we define $\underline{\pi} = 2 * P(\hat{\pi} < 0)$. 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.

tion 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.

Figure 2: The Relationship Between Dominance and Inclusive Support
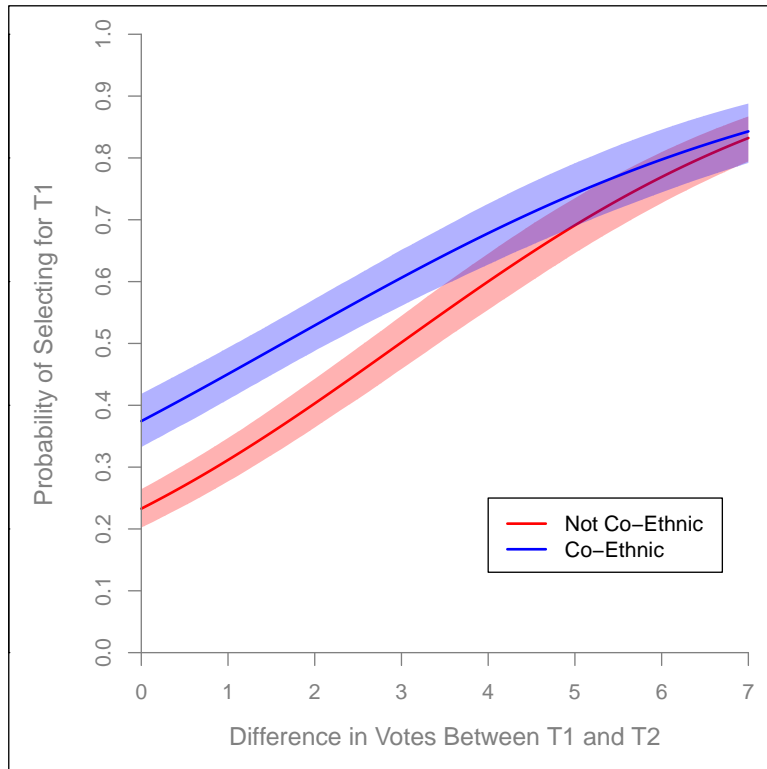
6 Tests: Who Do Selected Influencers Mobilize and Assist?

An alternative and complementary strategy to test our hypothesis is to evaluate whether dominant influencers themselves declare maintaining more inclusive influence networks.

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 B). 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 ability 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 ability to influence said voter. 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 her ability to maintain a more inclusive influence network.

6.1 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 (6.1). 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;

$$y_{ijk} = \beta_0 + \beta_1 \Delta_k + \beta_2 C_{ij} + \beta_3 C_{ij} * \Delta_k + \mathbf{Z}_i \gamma + \alpha_i + \alpha_j + \alpha_k + \varepsilon_{ijk} \tag{6.1}$$

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

25

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 $\hat{\pi}$ be a vector of values drawn from the posterior distribution of a parameter of interest. Then, we define $\underline{\pi} = 2 * P(\hat{\pi} < 0)$. 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
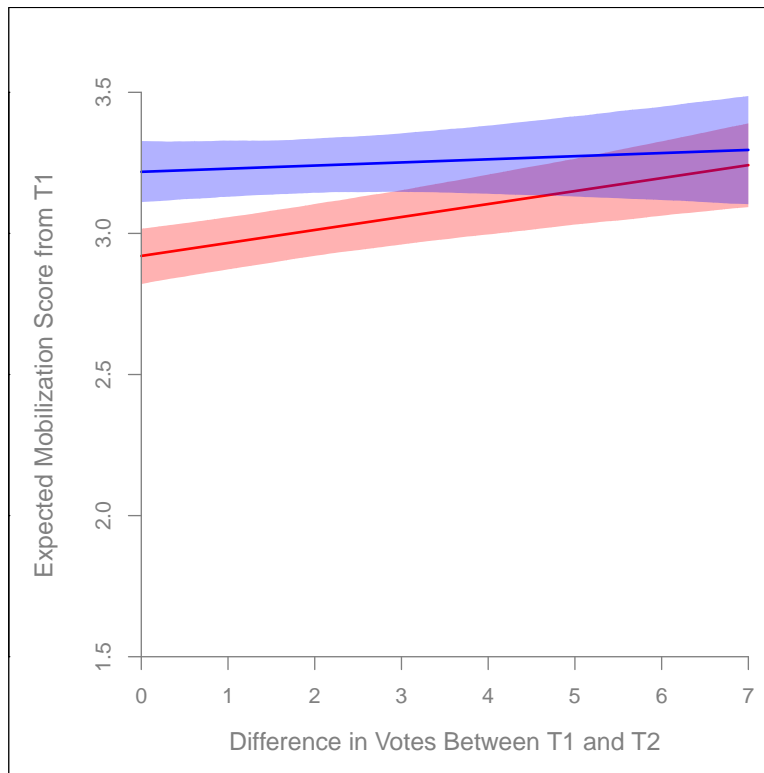
Two points are evident from this figure. First, the capacity to mobilize non-co-ethnics converges to the capacity converge co-ethnics for $T1$ as a function of dominance. Second, the capacity to mobilize co-ethnics does not really change as a function of dominance; the convergence results from a greater 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.

6.2   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 influencers may help any cit-

izen, 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.[37]

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 $\tilde{x}_i = x_{ijk} - \bar{x}_j$ for each $j \in \mathcal{I}$, where $\bar{x}_J$ 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. Further define Q_{ij} as a binary variable denoting co-partisanship between respondent i and influencer j. We may now write the regression equation:

$$rank_{ijk} = \zeta + \tilde{\mathbf{X}}_i \beta + \alpha_i + \varepsilon_{ijk} \tag{6.2}$$

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

---

[37]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.

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.

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.

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 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.

Figure 4: The Relationship Between Dominance and Willingness to Help
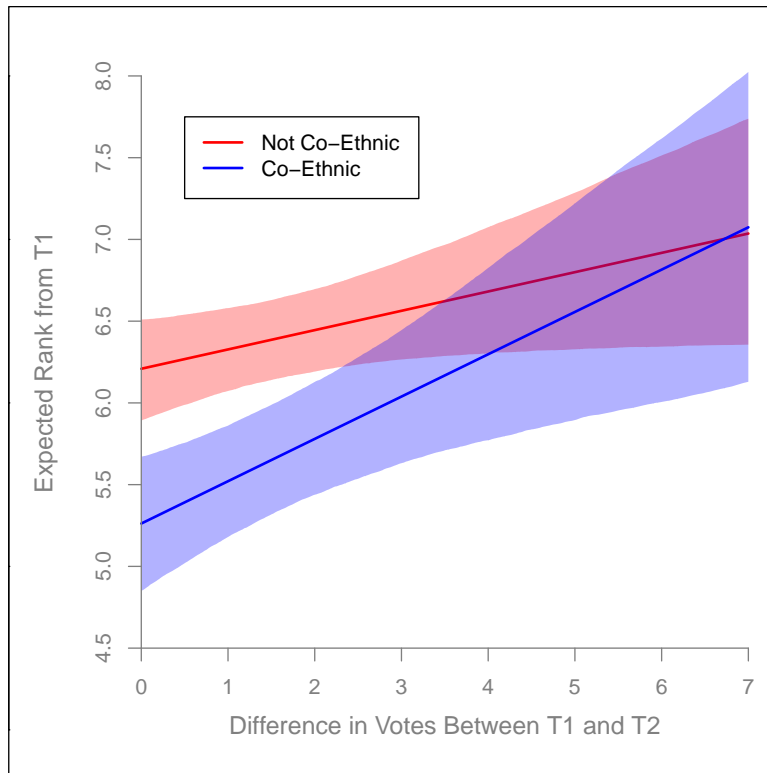
7   Robustness

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.

7.1 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.[38]

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 B, 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.

---

[38]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.

7.2 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 ability to maintain 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 exercise effort - 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 E, 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 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' quality. 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.

8 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.[39] 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 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

---

[39]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.

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 Michael. 2016. "Clients and Communities: The Political Economy of Party Network Organization and Development in India's Urban Slums." *World Politics* 68(1):111–148.

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.

Bussell, Jennifer. 2014. "Representation Between the Votes: Informal Citizen-State Relations in India." Unpublished.

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.

Hicken, Allen. 2011. "Clientelism." *Annual Review of Political Science* 14:289–310.

Holland, Alisha C. and Brian Palmer-Rubin. 2015. "Beyond the Machine: Clientelist Brokers and Interest Organizations in Latin America." *Comparative Political Studies* 48(9):1186–1223.

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.

Krishna, Anirudh. 2002. *Active Social Capital*. Columbia University Press.

Kruks-Wisner, Gabrielle. Forthcoming. *Active Citizenship: Claim-Making & the Pursuit of Social Welfare in Rural Indi*. Cambridge University Press.

Larreguy, Horacio, Cesar Montiel and Pablo Querubin. 2017. "Political Brokers: Partisans or Agents? Evidence from the Mexican Teacher's Union." *American Journal of Political Science* Forthcoming.

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.

Muñoz, Paula. 2014. "An Informational Theory of Campaign Clientelism: The Case of Peru." *Comparative Politics* 47(1):79–98.

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.

Van de Walle, Nicholas. 2007. Meet the new boss, same as the old boss? The evolution of political clientelism in Africa. In *Patrons, Clients, and Policies*, ed. Herbert Kitschelt and Steven Wilkinson. Cambridge 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 A Model of Competition Between Influencers

This is a highly stylized model in which we analyze how a potentially inclusive influencer, say influencer j, allocates effort to coethnics and non-coethnics when faced with contesting influencers who are targeting their coethnics. The goal here is to demonstrate that when j possesses significantly higher ability than the influencers contesting against him, it is in his best interests to build an inclusive coalition. If, on the other hand, if the contesting influencers possess levels of ability (either in the coethnic group of j or in one of the non-coethnic groups of j), then j is pushed to allocate more effort towards his own coethnics.

Consider a set of groups $G_1, \ldots, G_K$ with sizes $N_1, \ldots, N_K$, so the total population is $N = \sum_k N_k$.

Consider a voter $i \in \mathcal{V}$ and influencer $j \in \mathcal{I}$. 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 action of the influencer j towards group k is defined as $\alpha_{j,k} \in [0,1]$, the proportion of his/her ability that is spent towards group k, where $\sum_k \alpha_{j,k} = 1$. Let λ denote the constant marginal impact on a coethnic voter's utility from increased effort towards her group (for a non-coethnic voter, the marginal impact is normalized to 1), so $\lambda > 1$. The utility for voter i from influencer j, U_{ij}, is given by:

$$
U_{ij} = \begin{cases} \lambda \alpha_{j,g(i)} \Gamma_j + \varepsilon_{ij} & \text{if } g(i) = g(j) \\ \alpha_{j,g(i)} \Gamma_j + \varepsilon_{ij} & \text{if } g(i) \neq g(j) \end{cases}
$$

$$
\varepsilon_{ij} \sim \mathcal{N}\left(0, \frac{1}{2}\right)
$$

Note here that ε_{ij} is an idiosyncratic level of support (ideological/personal attachment) for voter i to influencer j uncorrelated to support for any other influencer. For ease of notation, define:

A.1 Aggregation and Assumptions

Consider two influencers $j, q_k \in \mathcal{I}$, where j is a potentially inclusive influencer and q_k is the highest ability influencer in group k (outside of j). Formally,

$$
q_k = \underset{r \in \{q:g(q)=k\}\setminus\{j\}}{\arg\max} \Gamma_r
$$

A voter, $i \in \mathcal{V}$, prefers j to q_k if $U_{ij} > U_{iq_k}$. Then, j is preferred if:

$$\lambda \alpha_{j,g(i)} \Gamma_j - \lambda \alpha_{q_{g(i)},g(i)} \Gamma_{q_{g(i)}} + \varepsilon_{ij} - \varepsilon_{iq_{g(i)}} > 0 \quad \text{if } g(i) = g(j) \tag{A.1}$$

$$\alpha_{j,g(i)} \Gamma_j - \lambda \alpha_{q_{g(i)},g(i)} \Gamma_{q_{g(i)}} + \varepsilon_{ij} - \varepsilon_{iq_{g(i)}} > 0 \quad \text{if } g(i) \neq g(j)$$

Then expected number of voters for influencer j, $\mathbb{E}(V_j)$, under a particular profile of actions by j is given by (using the fact that ε is normally distributed and idiosyncratic):

$$\mathbb{E}(V_j) = N_{g(j)} * \Phi \left(\lambda \alpha_{j,g(j)} \Gamma_j - \lambda \alpha_{q_{g(j)},g(j)} \Gamma_{q_{g(i)}} \right) + \sum_{k \neq g(j)} N_k * \Phi \left(\alpha_{j,k} \Gamma_j - \lambda \alpha_{q_k,k} \Gamma_{q_k} \right) \tag{A.2}$$

where Φ is defined as the cumulative distribution function of a standard normal distribution.

The term $\mathbb{E}(V_j)$ is the objective function each influencer j seeks to maximize subject to choices of α_{jk} under the straightforward constraint the sum of these terms equal to 1. The first order condition for the constrained maximization function (Lagrangian) $\mathcal{L}$ with respect to α_{jk} is given by:

$$\frac{\partial \mathcal{L}}{\partial \alpha_{j,k}} = \lambda \Gamma_j N_k * \phi \left(\lambda \alpha_{j,k} \Gamma_j - \lambda \alpha_{q_k,k} \Gamma_{q_k} \right) - c \quad \text{if } g(j) = k \tag{A.3}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_{j,k}} = \Gamma_j N_k * \phi \left(\alpha_{j,k} \Gamma_j - \lambda \alpha_{q_k,k} \Gamma_{q_k} \right) - c \quad \text{if } g(j) \neq k$$

where c is the Lagrange multiplier ϕ is the probability density function of a standard normal distribution.[40]

Two points are noted without proof. Existence of an equilibrium is guaranteed by a standard (Brouwer) fixed point argument, and the constraint obviously binds since the objective function is increasing in α_{jk} for each k, holding other α terms constant. A simple "algorithmic" re-parameterization will be useful here.

A.1.1 Assumptions

We note that this model is consistent with a broad class of outcomes. Here, we are interested in a simplified model in which the influencers operate in a universe where ethnicity is highly salient. In a general exposition, many of these assumptions can be discarded or weakened. We

---

[40]We note here that we omit Lagrange multipliers for the bounds on α, although we will check for corner solutions below.

choose these assumptions, as they fit our scope conditions well and allow for an intuitive expo-
sition of equilibrium behavior.

The first assumption is that in an ethnically divided society, our class of influencers must
have co-ethnic bias (or at least cannot credibly commit to not have it). Empirically, this means
the proportion of co-ethnics on an influencer's side must be great than the share in any other
group.

Assumption A.1 (Co-Ethnic Bias). *An influencer has a higher probability of receiving a vote from a
co-ethnic than from a non-coethnic always. For all influencers $j \in \mathcal{I}$:*

$$\Phi\left(\lambda \alpha_{j,g(j)} \Gamma_j - \gamma_{q_{g(j)},g(j)} \Gamma_{q_{g(j)}}\right) > \Phi\left(\alpha_{j,k} \Gamma_j - \gamma_{q_k,k} \Gamma_{q_k}\right) \quad \textit{for all } k \neq g(j)$$

The second assumption is that influencers can only credibly commit to putting spending
at least as much of their abilities on coethnics as compared to non-coethnics. Empircally, this
means that for influencer j $\alpha_{j,g(j)} > \alpha_{j,k}$ for all other $k \neq g(j)$.

Assumption A.2 (Co-Ethnic Targeting). *For all influencers $j \in \mathcal{I}$:*

$$g(j) = \arg\max_{k \in \{1,\dots,K\}} \alpha_{j,k}$$

A third assumption is that targeting co-ethnic voters must have a sufficiently large pay-
off, i.e., λ is sufficiently large. That is, when influencers expend the increase effort towards co-
ethnics and a non-coethnic group by the same amount holding the capacity of other influencers
constant, the influencer should expect more voters from the co-ethnic group than the non-co-
ethnic group (the assumption follows from holding ϕ constant across groups in the first order
condition):

Assumption A.3 (Co-Ethnic Impact). *An influencer has higher marginal impact on co-ethnics than on
a non-co-ethnic, holding all else constant. For all influencers $j \in \mathcal{I}$:*

$$\lambda > \frac{N_{g(j)}}{\max_k N_k}$$

41

A.1.2   Re-Parameterization

Using the results from above, an "algorithmic" re-parameterization will make the results more intuitive. Given two groups k and k', we are interested in the *difference* in the first order conditions. We further define $\gamma^*_{j,k,k'} = \frac{\alpha^*_{j,k}}{\alpha^*_{j,k} + \alpha^*_{j,k'}}$, where $\alpha^*_{j,k}$ is the equilibrium value for $\alpha_{j,k}$. For an influencer j, starting with group $k = g(j)$,[41] for each group k, choose a group k' that minimizes the value of $\gamma^*_{j,k,k'}$ (i.e., the value when the difference in first order conditions for k and k' is considered). Then in the next step define k' as k and repeat. If no group can be found such that $\gamma_{j,k,k'} < 1$, then stop the algorithm. Formally, we may state the following algorithm:

1. Start with $g(j) = k$.

2. For group k, find a group k' that minimizes $\gamma^*_{j,k,k'} \in [0,1]$ in the equation:

$$\lambda \Gamma_j N_k * \phi \left(\lambda \gamma^*_{j,k,k'} \Gamma_j - \lambda \alpha^*_{q_k,k} \Gamma_{q_k} \right) - \Gamma_j N_{k'} * \phi \left((1 - \gamma^*_{j,k,k'}) \Gamma_j - \lambda \alpha^*_{q_{k'},k'} \Gamma_{q_{k'}} \right) = 0 \quad \text{if } g(j) = k$$

$$\Gamma_j N_k * \phi \left(\gamma^*_{j,k,k'} \Gamma_j - \lambda \alpha^*_{q_k,k} \Gamma_{q_k} \right) - \Gamma_j N_{k'} * \phi \left((1 - \gamma^*_{j,k,k'}) \Gamma_j - \lambda \alpha^*_{q_{k'},k'} \Gamma_{q_{k'}} \right) = 0 \quad \text{if } g(j) \neq k$$

3. If no such k' exists, stop the process

4. Re-define k' as k and start from step 2.

To see that this re-parameterization works, notice that there are at most $K - 1$ equations, the same number as α parameters. To see that this re-parameterization is equivalent to solution, for some $\alpha^*_{j,g(j)}$ the first iteration of the algorithm implies:

$$\alpha^*_{j,k'} = \frac{1 - \gamma^*_{j,g(j),k'}}{\gamma^*_{j,g(j),k'}} \alpha^*_{j,g(j)}$$

For each k thereafter in the iterations,

$$\alpha^*_{j,k'} = \frac{1 - \gamma^*_{j,k,k'}}{\gamma^*_{j,k,k'}} \alpha^*_{j,k}$$

---

[41]We make start with $g(j)$ due to assumption A.2.

The task then is to find $\alpha^*_{j,g(j)}$ such that the sum of the α^* terms is 1. This recursive form also makes clear that when the algorithm stops, any group, r, of non-coethnics that has not served as k' in the algorithm has $\alpha^*_{j,r} = 0$. A further implication is that $\gamma^*_{j,k,k'} > 0.5$ always.

A.2 Equilibrium Analysis

Our first result in characterizing equilibrium shows that when an influencer j faces a superior influencer in equilibrium in targeting his co-ethnic group $g(j)$, then j is a pure ethnic influencer (that is, $\alpha^*_{j,g(j)} = 1$).

Theorem A.4. *For any $j \in \mathcal{I}$, if:*

$$\lambda \alpha^*_{j,g(j)} \Gamma_j < \lambda \alpha^*_{q_{g(j)},g(j)} \Gamma_{q_{g(j)}}$$

then $\alpha^*_{j,g(j)} = 1$.

Proof:

The condition implies that $\Phi(\lambda \alpha^*_{j,g(j)} \Gamma_j - \lambda \alpha^*_{q_{g(j)},g(j)} \Gamma_{q_{g(j)}}) < 0.5$. Furthermore, assumption A.1 implies that:

$$\Phi(\lambda \alpha^*_{j,g(j)} \Gamma_j - \lambda \alpha^*_{q_{g(j)},g(j)} \Gamma_{q_{g(j)}}) > \Phi(\alpha^*_{j,k} \Gamma_j - \lambda \alpha^*_{q_k,k} \Gamma_{q_k}) \quad \text{for } k \neq g(j)$$

This implies that:

$$\phi(\lambda \alpha^*_{j,g(j)} \Gamma_j - \lambda \alpha^*_{q_{g(j)},g(j)} \Gamma_{q_{g(j)}}) > \phi(\alpha^*_{j,k} \Gamma_j - \lambda \alpha^*_{q_k,k} \Gamma_{q_k}) \quad \text{for } k \neq g(j)$$

Furthermore, assumption A.3 implies $\lambda N_{g(j)} > N_k$ for all $k \neq g(j)$. This in turn implies that:

$$\lambda \Gamma_j N_{g(j)} * \phi\left(\lambda \alpha^*_{j,g(j)} \Gamma_j - \lambda \alpha^*_{q_{g(j)},g(j)} \Gamma_{q_{g(j)}}\right) > \Gamma_j N_k * \phi\left(\alpha^*_{j,k} \Gamma_j - \lambda \alpha^*_{q_k,k} \Gamma_{q_k}\right) \quad \text{for all } k \neq g(j)$$

This final equation implies that no k' can be found in the first iteration of the algorithm described above. The result follows immediately. $\square$

The implication is that, when it comes to coethnics, an influencer must aim win more than 50% of the group. In order to accomplish this, the influencer will purely target coethnics until

this aim is met. Thus, if the abilities of two coethnic influencers are relatively close, the $T1$ influencer is pushed to target a higher proportion of coethnics (given that his challenger in the contest is purely targeting coethnics).

We now consider a case in which in equilibrium, influencer j provides some positive targeting to group $k \neq g(j)$. From the previous theorem, this also implies that j is the most influential among his own coethnics. We show that an exogenous increase in the ability of j's competitor in k creates *more* coethnic targeting by j.

Theorem A.5. *For all $j \in \mathcal{I}$, if there exists $k \neq g(j)$ such that:*

$$\alpha^*_{j,k} \Gamma_j < \lambda \Gamma_{q_k}, \quad \alpha^*_{j,k} > 0$$

then:

$$\frac{\partial \alpha^*_{j,g(j)}}{\partial \Gamma_{q_k}} > 0$$

Proof:

The condition implies that $\alpha^*_{j,k} > 0$ and thus $\alpha^*_{j,g(j)} < 1$. This implies that j is the strongest influencer in $g(j)$ from theorem A.5.

It follows that:

$$\lambda \Gamma_j N_{g(j)} * \phi \left(\lambda \alpha^*_{j,g(j)} \Gamma_j - \lambda \alpha^*_{q_{g(j)},g(j)} \Gamma_{q_{g(j)}} \right) - \Gamma_j N_k * \phi \left(\alpha^*_{j,k} \Gamma_j - \lambda \alpha^*_{q_k,k} \Gamma_{q_k} \right) = 0$$

Theorem A.5 also implies $\alpha^*_{q_{g(j)},g(j)} = 1$. We begin with the case where $\alpha^*_{q_k,k} = 1$ (which occurs when q_k has sufficiently low ability that he requires the corner solution solved in theorem A.5). The above equation can then be written as:

$$\lambda \Gamma_j N_{g(j)} * \phi \left(\lambda \beta^*_{j,g(j),k} \Gamma_j - \lambda \Gamma_{q_{g(j)}} \right) - \Gamma_j N_k * \phi \left(\left(1 - \beta^*_{j,g(j),k} \right) \Gamma_j - \lambda \Gamma_{q_k} \right) = 0 \qquad \text{(A.4)}$$

where β^* can be constructed from the recursive rule described above.

The implicit derivative can be messy to compute since β^* is a function of each Γ term, but a simple heuristic will suffice for the proof. Notice that $\alpha^*_{j,k} \Gamma_j < \lambda \Gamma_{q_k}$ implies that, holding $1 - \beta^*_{j,g(j),k} = b$ implies that $\phi \left(b \Gamma_j - \lambda \Gamma_{q_k} \right)$ is decreasing in Γ_{q_k}. Then the expression on the left hand

side of A.4 is increasing in Γ_{q_k}. It follows that the zero condition in equation A.4 can only be satisfied by increasing $\beta^*_{j,g(j),k}$. Since $\alpha^*_{j,g(j),k}$ is increasing in $\beta^*_{j,g(j),k}$, the theorem holds for $\alpha^*_{q_k,k} = 1$. To see that this holds in general, note that $\alpha^*_{q_k,k}\Gamma_{q_k}$ is increasing in Γ_{q_k} $\quad\square$

This theorem tells us that as an influencer faces an influencer with increasing ability in a non-coethnic group, he increases his targeting to coethnics.

B 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.

C   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

D 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.

E The Absence of a Relationship Between Dominance and Social Closeness
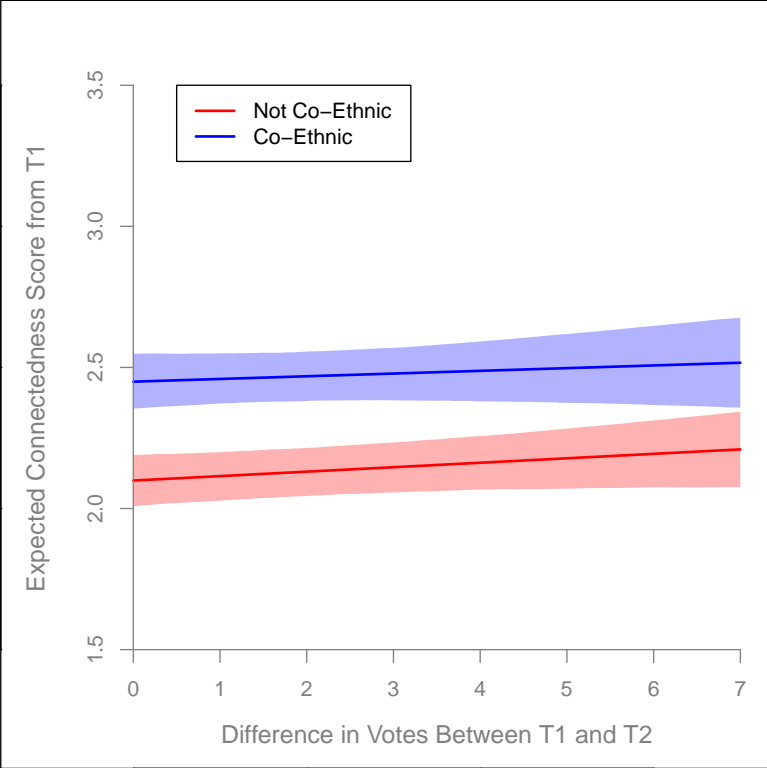
Figure 5: Simulated Effects from Regression

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 $\hat{\pi}$ be a vector of values drawn from the posterior distribution of a parameter of interest. Then, we define $\underline{\pi} = 2 * P(\hat{\pi} < 0)$. 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.