MILAN VAISHNAV

Ethnic Identifiability in India

Evidence from a Voter Survey

ABSTRACT

The ethnic politics literature assumes that one's ethnic identity is readily identifiable. This paper evaluates how well this assumption holds in practice, drawing on a voter survey in the Indian state of Bihar. While voters often misidentify the identity of their candidate, the degree of error is small but systematic.

KEYWORDS: ethnic politics, voter behavior, elections, caste, India

INTRODUCTION

In recent years, there has been a surge of interest among social scientists in the study of ethnic politics. For instance, scholars of conflict and civil war have examined how ethnic grievances shape conflict outcomes.[1] Researchers have explored the conditions under which ethnic heterogeneity affects the possibilities for individuals to engage in collective action and whether ethnic

MILAN VAISHNAV is Senior Fellow in the South Asia Program at the Carnegie Endowment for International Peace, Washington, DC, and the author of *When Crime Pays: Money and Muscle in Indian Politics* (Yale University Press, 2017). He is grateful to Sanjay Kumar, Rakesh Ranjan, Rakesh Kumar, and Mukesh Kumar Rai of the Lokniti Programme of the Centre for the Study of Developing Societies and to Sanjay Kumar of the A. N. Sinha Institute for their assistance. He is also appreciative of comments from an anonymous reviewer, as well as Johnathan Guy, Dominika Koter, Phil Oldenburg, Arvind Subramanian, Steven Wilkinson, and participants at the 2011 American Political Science Association annual meeting, the 2013 Midwest Political Science Association annual meeting, and the DC India Politics workshop. Rebecca Brown and Owen McCarthy provided excellent research assistance. The author acknowledges financial support for this project from the Center for the Advanced Study of India at the University of Pennsylvania, the National Science Foundation (Doctoral Dissertation Improvement Grant SES #1022234), and the Center for Global Development. Email: <mvaishnav@ceip.org>.

1. James D. Fearon and David D. Laitin, "Ethnicity, Insurgency and Civil War," *American Political Science Review* 97:1 (February 2003): 75–90; Andreas Wimmer, Lars-Erik Cederman, and Brian Min, "Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set," *American Sociological Review* 74:2 (April 2009): 316–37.

Asian Survey, Vol. 57, Number 4, pp. 738–763. ISSN 0004-4687, electronic ISSN 1533-838X. DOI: https://doi.org/10.1525/AS.2017.57.4.738.

diversity impacts the provision of public goods.[2] And, of course, there is growing interest in studying how ethnic identity can shape voting behavior.[3]

Although the ethnic politics literature is unquestionably diverse, almost all studies that explore the impact of ethnicity on a given outcome share a common assumption: that individuals can readily and correctly identify the ethnic identity of other individuals in their community without ambiguity.[4] This is the assumption of "ethnic identifiability," or the ability for person A to correctly identify person B as belonging to the same ethnic category that person B uses to identify him or herself. While this is a standard assumption in the ethnic politics literature, several recent micro-level analyses have demonstrated that it often does not hold in actuality.[5] In the words of one study, in real-world settings there can be meaningful discrepancies between the "subjective" and "objective" demographies of a given community.[6] Furthermore, the degree of imprecision might have an underlying structure that is nonrandom.

2. James D. Fearon and David D. Laitin, "Explaining Interethnic Cooperation," *American Political Science Review* 90:4 (December 1996): 715–35; Alberto Alesina, Reza Baqir, and William Easterly, "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics* 114:4 (November 1999): 1243–84; James Habyarimana, Macartan Humphreys, Daniel N. Posner, and Jeremy Weinstein, *Coethnicity: Diversity and the Dilemmas of Collective Action* (New York: Russell Sage Foundation, 2009).

3. Donald L. Horowitz, *Ethnic Groups in Conflict* (Berkeley: University of California Press, 1985); Kanchan Chandra, *Why Ethnic Parties Succeed: Patronage and Ethnic Head Counts in India* (New York: Cambridge University Press, 2004); Daniel N. Posner, *Institutions and Ethnic Politics in Africa* (New York: Cambridge University Press, 2005).

4. Kanchan Chandra, "Counting Heads: A Theory of Voter and Elite Behavior in Patronage Democracies," in Herbert Kitschelt and Steven Wilkinson, eds., *Patrons, Clients and Policies: Patterns of Democratic Accountability and Political Competition* (New York: Cambridge University Press, 2007): 96–97.

5. James Habyarimana, Macartan Humphreys, Daniel N. Posner, and Jeremy Weinstein, "Placing and Passing: Evidence from Uganda on Ethnic Identification and Ethnic Deception," unpublished manuscript, Department of Political Science, Columbia University, 2007; Daniel Corstange, "Ethnic Observables and Educated Guess, or, What Can We Learn From a Simple Dagger?" unpublished manuscript, Department of Government and Politics, University of Maryland, 2010; Adam S. Harris and Michael G. Findley, "Is Ethnicity Identifiable? Lessons from an Experiment in South Africa," *Journal of Conflict Resolution* 58:1 (February 2014): 4–33; Pavan Mamidi, "Signalling and Passing in the Caste System: Experimental Evidence from India," unpublished manuscript, Nuffield College, Oxford University, 2010. While the political science literature on this topic is relatively novel, there is a more established body of literature in social psychology that interrogates the assumption of "ethnic identifiability": Diana Rice and Brian Mullen, "Isaac, Ishmael, and Janus: Past and Future Lessons Regarding the Ethnic Categorization of Faces," *Applied Cognitive Psychology* 17:9 (November/December 2003): 1129–47.

6. Habyarimana et al., *Coethnicity*, 37.

This paper contributes to this literature by testing the common assumption of ethnic identifiability in a unique context that is highly pertinent to the field of political science: the act of voting. Specifically, this paper evaluates how well voters can identify the ethnic identity of their preferred candidate in an electoral context where ethnic differences are highly salient. Situating the issue of ethnic identifiability in the electoral domain is relevant for at least four reasons. First, there is a venerable literature that has argued that the ethnic identity of a politician has a considerable influence on an individual's voting decision.[7] Second, some studies have argued that the ethnic identity of one's elected representative can shape one's retrospective evaluation of incumbent performance.[8] In addition, several studies have shown that the ethnic identity of a politician can shape voters' prospective evaluations of candidates' future performance in office.[9] Finally, when voters and their representatives are co-ethnics, this shared status is thought to influence incumbent responsiveness.[10]

In contrast to recent studies on this topic, this paper is unique for at least two reasons. First, it explicitly tests the ethnic-identifiability assumption in an electoral context, using original data on voters and candidates. Second, this study's empirical contribution relies on real, as opposed to simulated, data. Previous studies in this area have typically adopted an experimental "lab in the field" approach. While such an approach has obvious advantages from the perspective of causal inference, it also has a downside in terms of limits on external validity.

Empirically, this paper relies on data drawn from a post-election survey of more than 2,000 randomly selected voters in the North Indian state of Bihar that was conducted immediately following the conclusion of voting in state assembly elections in October-November 2010. Bihar, a large state that is home to more than 100 million Indians, is an ideal setting in which to explore issues of ethnic identifiability, as scholars have long characterized politics in

7. Chandra, "Counting Heads"; Karen Ferree, "Explaining South Africa's Racial Census," *Journal of Politics* 68:4 (November 2006): 803–15.

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

9. Ferree, "Explaining South Africa's Racial Census"; Leonard Wantchekon, "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin," *World Politics* 55:3 (April 2003): 399–422.

10. Gwyneth McClendon, "Race and Responsiveness: A Field Experiment with South African Politicians," *Journal of Experimental Political Science* 3:1 (April 2016): 60–74.

the state as principally revolving around identity-based competition.[11] Particularly over the last several decades, Bihar has witnessed intense political mobilization around ethnic identities, with parties and candidates making explicit ethnically based appeals to voters during elections. The survey this paper draws from asked voters to identify the ethnic identity of the candidate they voted for in the recently concluded election. These responses were then matched with a unique dataset on the actual ethnic identities of the major party candidates, which the research team constructed independently.

The empirical analysis is in two parts. The first part tests the identifiability assumption and explores some descriptive patterns regarding the nature of errors voters make when attempting to identify candidates' identities. The second part attempts to model the determinants of accurate identification, using survey data to test hypotheses generated by the comparative literature. To preview the paper's principal findings, I find that nearly one in three survey respondents do not accurately identify the ethnic identity of the candidate they voted for. In a fragmented, multi-ethnic environment, it is perhaps not surprising that we do not observe perfect identification. On the other hand, respondents are often not far off in their assessments: the majority of errors occur within, rather than across, aggregate ethnic categories.

Also, as prior studies have shown, there is considerable structure to the misclassification errors. Modeling the determinants of accurate identification more formally, I find that co-ethnicity improves accuracy, while partisanship modestly reduces it, and education (used as a proxy for an individual's access to information) has no discernible impact. One of the most interesting findings is that Muslims clearly stand apart from other ethnic groups: they are better at identifying, and at being identified, than individuals belonging to other ethnic categories. This is the true even though Muslims vote for fellow Muslims in only a minority of instances.

Before proceeding, a clarification about terminology is in order. Following Chandra, I use *ethnic identity* to refer to "identities based on ascriptive categories, including caste, language, tribe, and religion."[12] In Indian politics, and particularly in Bihar, caste and religion are the two most salient social identifiers. Thus, this paper focuses on caste and religious identities under

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

12. Chandra, *Why Ethnic Parties Succeed*, xvii.

this larger heading of ethnicity. Undoubtedly, caste and religion are distinct phenomena (in the Indian case scholars often talk about "communal" issues in reference to religious differences, as distinct from "caste" issues). Yet, with respect to electoral politics, these identity categories also exhibit similar political-electoral dynamics.

The remainder of the paper is organized as follows. In the next section, I describe the electoral context in Bihar and summarize the survey data used here. In the third section, I present some preliminary results on patterns of identifiability that seek to shed light on three fundamental questions: To what extent do voters misidentify candidates? Are certain groups systematically prone to such misidentification? And when voters misclassify candidates, how badly do they do so? The fourth section presents hypotheses, gleaned from the literature, which might explain individual voters' ability to correctly identify the ethnic identity of their preferred candidate. In the fifth section, I present the results of a statistical analysis of the determinants of accurate identification. The last section offers some parting thoughts.

SETTING AND SURVEY DATA

This section begins with a brief description of politics in Bihar, highlighting in particular the relevance of identity-specific attributes in the political-electoral discourse, before describing how I operationalize the concept of ethnic identity. Finally, I discuss the details of the post-poll survey and related, complementary data gathered for this study.

Identity Politics in Bihar

Ethnic identity, particularly caste identity, has been a defining feature of politics in contemporary India. Scholars have noted that the caste system has played a crucial role in organizing social relations and giving legitimacy to social hierarchies in what has been historically a largely agrarian economy.[13] For a variety of reasons, including the nature of colonial structures of land

13. Nicholas B. Dirks, *Castes of Mind* (Princeton, NJ: Princeton University Press, 2001); Susan Bayly, *Caste, Society, and Politics in India from the Eighteenth Century to the Modern Age* (New York: Cambridge University Press, 2001); Dipankar Gupta, *Caste in Question: Identity or Hierarchy?* (New Delhi: Sage, 2000).

tenure, demographics, and the influence of the Congress Party, identity politics has been particularly prevalent in the state of Bihar.[14]

Situated along North India's Indo-Gangetic plain, Bihar is India's third-most populous state, with over 100 million residents; in area it is similar to South Korea. A largely agrarian state, Bihar is one of India's poorest, exhibiting human development indicators on par with the worst-performing countries of sub-Saharan Africa. Long a bastion of the traditionally upper-caste-dominated Congress Party, much of Bihar's post-independence politics has centered on questions of caste dominance: that is, which group or groups controlled the primary levers of economic and political power. As a result, politics and elections have traditionally been viewed through the lens of identity, rather than programmatic, politics.[15]

Identity considerations became particularly acute in the 1980s and 1990s, when Bihar was ground zero for the movement for lower-caste empowerment. The so-called Other Backward Classes (OBCs) are numerically dominant in Bihar and elsewhere in North India, yet for much of the post-independence period they have been politically marginalized. As Congress dominance slowly eroded and OBCs established their own political movements, competition among ethnic groups for political power in the state was a source of intense conflict.[16] The torchbearer for this movement was a wily politician named Lalu Prasad Yadav, who became the state's chief minister in 1990 and would dominate state politics for 15 years.[17] Yadav skillfully combined rustic language, ethnic symbolism, and savvy retail politics to build a political coalition of disparate castes and communities that would contest, and eventually oust, the long-established hold of the upper castes. Although caste had been a factor in Bihar's politics for centuries, Yadav's reign was associated with a deepening of identity politics; while in power, Yadav explicitly prioritized social justice over all other policy

14. Francine R. Frankel, "Caste, Land and Dominance in Bihar," in Francis R. Frankel and M. S. A. Rao, eds., *Dominance and State Power in Modern India: Decline of a Social Order* (New Delhi: Oxford University Press, 1990); Christophe Jaffrelot, *India's Silent Revolution: The Rise of the Lower Castes in North India* (New York: Columbia University Press, 2003).

15. Jeffrey Witsoe, "A View from the States: Bihar," in Atul Kohli and Prerna Singh, eds., *Routledge Handbook of Indian Politics* (New York: Taylor & Francis, 2011): 298–308.

16. Jaffrelot, *India's Silent Revolution*.

17. Lalu Prasad Yadav was forced to step down after 10 years, in the wake of a massive corruption scandal. While nominally resigning his post, he managed to install his wife, Rabri Devi, as chief minister.

priorities.[18] In 2005, the Lalu Prasad Yadav era gave way to a period of reform under Chief Minister Nitish Kumar. While Kumar campaigned and later governed on a platform of good governance and inclusive development, he too has skillfully manipulated social divisions to build and maintain political support.[19]

One fact worth emphasizing is the widespread identification of politicians and voters by their caste or religious identities in Bihar, as is the case in many settings across India. In Bihar in particular, the scholarly literature and journalistic reporting on this point are quite clear. For instance, an article published before the 2010 elections described the Congress Party's slate of nominated candidates in explicitly ethnic terms:

> Hoping for a return of minorities towards Congress . . . the party has given tickets [nominations] to 47 Muslims in 221 seats for which candidates have been declared so far. Upper Castes . . . have also got major share of Congress tickets. They have already got tickets for 79 seats, maximum 32 of which have gone to Rajputs followed by 25 to Bhumihars and 18 to Brahmins. The party has announced OBCs as candidates in 56 seats so far of which Yadavs have got 20. A senior Congress leader said the strategy is to dent the M-Y (Muslim-Yadav) combination of Lalu Prasad. The Scheduled Castes have got 37 seats of which Paswan community has got 12.[20]

The focus on the caste identity of the candidate is important to highlight because it updates the traditional view of ethnic politics, which typically emphasizes the "ethnic labels" of parties. While a party's association with a particular community is certainly important, more recent work has found that voters also consider the ethnicity of the specific candidate on the ballot when formulating their voting decisions.[21] My own interviews conducted during the 2010 state assembly election in Bihar corroborate the widespread use of ethnic identities as a kind of shorthand in talking about candidates for election. Qualitative fieldwork in more than 10 constituencies revealed that voters, party workers, and journalists almost always began their characterization

18. Sankarshan Thakur, *Subaltern Saheb: Bihar and the Making of Laloo Yadav* (New Delhi: Picador, 2006).

19. Witsoe, "A View from the States: Bihar."

20. "Cong's Balancing Act in Bihar Ticket Distribution," Press Trust of India, October 13, 2010.

21. Simon Chauchard, "Unpacking Ethnic Preferences: Theory and Micro-Level Evidence from North India," *Comparative Political Studies* 49:2 (February 2016): 1–32.

of the local electoral contest by providing the names of the leading candidates, their party affiliations, and their ethnic identities.[22]

Operationalizing Ethnic Identity

In broad terms, this paper conceives of an individual's ethnic identity in Bihar as belonging to one of six overarching categories: Upper Caste (UC); Other Backwards Classes (OBC); Scheduled Caste (SC); Scheduled Tribe (ST); Muslim; or Other religious affiliation. As mentioned in the introduction, this conceptualization considers caste and religious identities in a single classificatory scheme. This maps onto the electoral-political discourse in Bihar and is in line with much of the literature on ethnic politics in India. A graphical representation of the six categories can be found in Figure A1 (in the Appendix). The first four aggregate categories constitute the modern-day caste system in India.[23] UCs occupy the highest rank in the hierarchy, followed by OBCs, and then SCs and STs. These four groups are further divided into individual castes, known as *jatis*—Bhumihar, Yadav, Dhobi, etc. It is these caste, or jati, identities that are commonly thought to be the salient distinctions in social and political life in Bihar (and in many other parts of India).

As identity politics in India has evolved over the decades, these jati-level identities have become more prominent in everyday political discourse. As Chauchard notes, political parties regularly "appeal to these more specific subunits instead of appealing to larger caste group categories," which means, in practical terms, that parties target some jatis within the larger caste categories while ignoring or downplaying others.[24] Because the broader caste categories are so large and diverse, politicians have found merit in carving up these groups into smaller subsets. In Bihar, for instance, many politicians opposed to Lalu Prasad Yadav complained that his rule was synonymous with *yadav raj*, or rule of the Yadavs (an OBC jati), at the expense of other OBCs on whose behalf he claimed to be advocating.[25]

22. The ubiquity of discussion about the caste composition of candidates and the electorate is even more striking in light of the fact that the last caste census took place in 1931, under the British Raj. The government carried out an update in 2011, but its results were never released.

23. Although, strictly speaking, STs fall outside of the traditional Hindu social hierarchy, and the position of SCs was thought to be so inferior that they were simply left out of the hierarchy entirely.

24. Chauchard, "Unpacking Ethnic Preferences," 9.

25. Jeffrey Witsoe, "Corruption as Power: Caste and the Political Imagination of the Postcolonial State," *American Ethnologist* 38:1 (February 2011): 73–85.

Indeed, as mentioned, discussions of politics in Bihar are usually framed in terms of individual jatis when it comes to members of the Hindu UC, OBC, and SC communities.[26] And this is by no means specific to Bihar. Recent scholarship has empirically demonstrated that party-system ethnification in India, or the degree to which parties have a unique ethnic basis of support, is best captured by jati-level distinctions, as opposed to umbrella caste categories or religious differences.[27]

The last two ethnic categories are Muslim and Other; the latter refers to individuals from non-Hindu, non-Muslim communities (Sikhs, Buddhists, etc.). While caste hierarchies do exist among non-Hindu communities in India and elsewhere in South Asia, the popular discourse in Bihar typically stays at the level of the broad umbrella categories of "Muslim" or "Sikh."[28]

Survey Data

The survey data for this paper were collected in collaboration with the Lokniti program of the Centre for the Study of Developing Societies. The data used in this paper were collected as part of a special module Lokniti carried out in parallel with its standard post-election survey of voters.

India is a federal parliamentary democracy comprising 29 states and seven union territories. Under the Indian constitution, states are endowed with primary responsibility over a wide range of day-to-day governance issues, ranging from electricity to land policy. All states have directly elected assemblies whose terms last five years. Assemblies are divided further into assembly constituencies governed by identical first-past-the-post, single-member-district electoral rules.

To construct the sample, Lokniti first randomly selected 40 assembly constituencies (out of a total of 243 in the entire state). Assembly constituencies were selected using probability proportionate to size.[29] Within each of

26. In states that have large ST populations (Bihar does not), individual tribes can have distinct identities.

27. John D. Huber and Pavithra Suryanarayan, "Ethnic Inequality and the Ethnification of Political Parties: Evidence from India," *World Politics* 68:1 (January 2016): 149–88. Using their preferred measure of party system ethnification, Huber and Suryanarayan find that using jati-level data produces the highest ethnification score 59% of the time.

28. In Muslim-dominated constituencies, of which there are several in Bihar, individual Muslim "castes" (jatis) are likely to be more salient.

29. This method ensures that locations with larger population are sampled; the underlying assumption is that locations with larger population will be more representative of the diversity within the universe.

these 40 constituencies, four polling stations were then chosen at random as sites to carry out the survey. Polling stations were again sampled with probability proportionate to size. After sampling polling stations, 20 respondents were randomly selected from the electoral rolls provided by the chief electoral officer of Bihar. Respondents were sampled using systematic random sampling, which is based on a fixed interval ratio between two respondents. For each polling station a sampled respondents list was prepared, which is a comprehensive list of the selected respondents with their complete name, address, age, and gender. Lokniti does not allow substitution in its surveys. Out of a target sample size of 3,200 respondents, Lokniti successfully interviewed 2,333 respondents (a response rate of 73%). Survey enumerators, trained by Lokniti and residents of the state, conducted face-to-face interviews at the respondent's place of residence in the respondent's native tongue.

The 2010 elections were held in six phases, with a unique set of constituencies going to the polls in each phase. After the conclusion of voting in each phase (usually within a few days), Lokniti carried out its survey with the target sample. To obtain information on respondents' vote choice, Lokniti begins its surveys by simulating an election. For respondents who affirmed that they were able to vote in the election, enumerators then produced a dummy ballot box and asked voters to cast their vote by marking the symbol of the party they voted for on a dummy paper ballot and depositing it in the box. The dummy ballots are tagged with a unique identifier, which enables easy merging with the rest of the respondent's information.

The key survey question from this study's perspective is the open-ended query, "What is the caste [*jati*/*biradari*] or tribe name of the candidate you voted for?" To construct a measure of identifiability, with the help of two research assistants I collected data on the caste/tribe of all major-party candidates contesting the elections. In the Bihar elections, these parties were Janata Dal (United), JD(U); Bharatiya Janata Party (BJP); Rashtriya Janata Dal (RJD); Lok Jan Shakti (LJP); Indian National Congress; Bahujan Samaj Party (BSP); National Congress Party; Communist Party of India; and Communist Party of India (Marxist-Leninist Liberation). To obtain the ethnic identity of major-party candidates, the research team visited party offices in Bihar's capital, Patna, and in several other districts, soliciting information from party officials on how party-affiliated candidates chose to self-identify.

For each major party candidate, we tried to obtain confirmation from at least two different sources within the party.[30]

I measure a voter's degree of accuracy in identifying a candidate's ethnic identity using a binary variable, *Accuracy*, which is 1 if the voter's perception of the candidate's ethnic identity matches the candidate's actual ethnic identity, and 0 otherwise. To be clear, an accurate match is when a voter can correctly identify the jati of his or her preferred candidate in the election. This is straightforward for UC, OBC, and SC communities (there are no ST candidates in the sample).[31] For non-Hindus—that is, Muslims and Others—I treat these broad categories as equivalent to jatis since they are the salient categories in politics. Although 2,333 respondents were surveyed, the final dataset contains 2,043 observations. There are two reasons for this. First, I was forced to drop respondents who voted for Independent candidates because we did not collect information on their ethnic identity.[32] Second, 181 respondents either did not vote in the election or did not recall who they voted for.

PATTERNS OF IDENTIFIABILITY

This section explores variation in the ability of voters to correctly identify the identity of the candidate they just voted for in the state assembly elections. This section is oriented around three questions. To what extent do voters misidentify the identity of the candidate they voted for? When voters do misidentify, what is the degree of their misidentification? Finally, are certain groups better or worse at identification?

To What Extent Do Voters Misidentify Candidates?

In the remainder of the paper, I refer to voters' identification of their preferred candidate's identity as "perceived identity," and candidates' true

30. As an additional validity check, a third research assistant independently collected data on the identity of politicians using his own sources.

31. For instance, a voter who identifies a Rajput candidate as a Bhumihar would receive an *Accuracy* score of 0. Although both Rajputs and Bhumihars belong to the UCs, the specific jati is the relevant distinction. In contrast, a voter must only identify a Muslim candidate as a Muslim to receive a score of 1.

32. Furthermore, because the mock voting procedure provides information on which party a respondent voted for, rather than the candidate's name, it is impossible to link a respondent's vote choice to a specific candidate (there are typically multiple independent candidates contesting elections in a constituency).

identity as "actual identity." The survey data suggest that nearly one-third (29.2%) of voters in Bihar misidentified the ethnic identity (jati) of the candidate they voted for in the election. Given the political science literature's emphasis on the salience of ethnic identity in politics and elections (not to mention the subset of this literature which examines the Indian case generally and Bihar specifically), the percentage of misidentifiers is not insignificant. Interestingly, respondents in Bihar exhibit identification rates toward the more accurate end of the comparative spectrum, based on findings of recent studies.

For instance, Habyarimina et al. do not report an overall summary statistic for their experimental study of 300 Ugandans from a Kampala neighborhood, but they do provide group-wise data on misclassifications.[33] Some 41% of Banyankole are correctly identified as such, and nearly 70% of Baganda are accurately classified as Baganda. Yet, for some groups the success rate is extremely low: only 6% of Banyoro are correctly classified. Mamidi studies the ability of individuals to successfully identify other individuals' caste categories, drawing on a sample of Indian college students.[34] He reports 74% success in detecting when someone else signals truthfully about their caste category, but only a 34% success rate for detecting when individuals attempt to "pass" as belonging to another caste group. Harris and Findley report much lower rates of correct classification in their study from the Eastern Cape of South Africa.[35] While they too do not provide an overall summary statistic, no group is correctly identified more than 50% of the time. Indeed, Xhosas are correctly identified 45% of the time, and the group with the next-best identification rate is the Zulu, who are correctly classified 28% of the time. Corstange reports that roughly 50% of Yemenis surveyed could accurately identify a fellow citizen as a "tribesman" in the absence of any additional cues.[36]

In India, there are clear differences in the rate of misidentification across jatis. Because the number of candidates from any one group is relatively small, to discern broad patterns it helps to collapse jatis into their four aggregate categories, as shown in Table 1.[37] The misidentification rates of

33. Habyarimana et al., "Placing and Passing"; Habyarimana et al., *Coethnicity*.

34. Mamidi, "Signalling and Passing in the Caste System."

35. Harris and Findley, "Is Ethnicity Identifiable?"

36. Corstange, "Ethnic Observables and Educated Guess."

37. In the sample, there are no candidates who belong to the Scheduled Tribe or Other categories.

TABLE 1. Misidentification Rates, Aggregated by Candidate Ethnic Group

Candidate ethnic group	*Correctly identified*	*Misidentified*	*% misidentified*	*n*
Upper Castes	342	139	28.9	481
Other Backward Castes	603	346	36.5	949
Scheduled Castes	189	85	31.0	274
Muslims	313	26	7.7	339
Total	1447	596	29.2	2043

SOURCE: All tables and figures by author.
NOTE: Identification is considered accurate if voters identify the correct *jati* of the candidate.

the three categories of Hindu candidates (UC, OBC, and SC) are roughly similar in magnitude: they range from 28.9% for UCs to 36.5% for OBCs. The difference in the misidentification rates between UCs and OBCs is statistically significant ($p < .01$), but the difference between UCs and SCs is not. In addition, there are stark differences when it comes to Muslim candidates, as voters seem to have a much easier time identifying them: only 7.7% of Muslim candidates are perceived as non-Muslims.

What Is the Degree of Misidentification?

Next, I explore the degree of misidentification. In other words, when voters misidentify the ethnic identity of their preferred candidate, how far off are their perceptions? To investigate, I again collapse jatis into four aggregate categories (UC, OBC, SC, and Muslim). The data in Table 2 show that when voters get it wrong, in most cases they do not get it very wrong. That is, while nearly one-third of respondents misidentify the jati of their preferred candidate, for almost 90% of all respondents the "perceived identity" falls within the correct overarching ethnic category. For instance, a voter might mistakenly identify a candidate as an upper-caste Bhumihar, when in fact the candidate belongs to the upper-caste Brahmin community. The voter misidentified the jati of candidate but was (correctly) able to locate him/her in the UC category. In other words, when voters make classification errors, most often they do so within (rather than across) broader ethnic categories.

Table 3 reports the degree of misidentification (in percentage terms), categorized by the four broad ethnic categories. For instance, the lower-right cell indicates that 84.9% of the time, voters identified a UC candidate

TABLE 2. What is the Degree of Misidentification?

Ethnic group	*Frequency*	*Percent*
Misidentified	218	10.7
Correctly identified	1,825	89.3
Total	2,043	100

SOURCE: By author.
NOTE: For the purposes of this table, identification is considered accurate if voters locate candidates within the correct umbrella ethnic group (Upper Caste, OBC, SC, Muslim).

TABLE 3. Degree of Misidentification by Ethnic Category, of Those Misclassified

		Actual ethnic group of candidates			
		Muslim	*OBC*	*SC*	*Upper Caste*
Perceived ethnic group of candidates	Muslim	0	6.7	0	.7
	OBC	73.1	53.5	10.6	13.7
	SC	0	2.3	88.2	0.7
	Upper Caste	26.9	37.6	1.2	84.9
	Total	100	100	100	100

SOURCE: By author.
NOTE: For the purposes of this table, identification is considered accurate if voters locate candidates within the correct umbrella ethnic group (Upper Caste, OBC, SC, Muslim). Numbers represent the percentage of candidates correctly identified as belonging to their overarching caste category.

as UC (although they got the candidate's specific jati wrong). Similarly, nearly 88.2% of voters identified SC candidates as hailing from an SC jati, while 10.6% of SC candidates were mistaken for OBCs. When voters incorrectly identified OBC jatis, in nearly half of the cases voters did not even place the candidate in the general OBC category: nearly 38% of voters of this type mistook OBC candidates as belong to a UC jati.

But, all in all, as Table 3 suggests, the degree of misclassification was not great. Voters may not always get it exactly right, but roughly 90% of the time they are in the ballpark. This echoes a point made by Habyarimana et al.: in their Uganda study, errors of classification tended to occur within (and not across) a set of broader regional categories.[38] But even if the degree of misclassification is not large, it is noteworthy, given how much of the political

38. Habyarimana et al., *Coethnicity.*

TABLE 4. Misidentification Rates by Voter Ethnic Group

Voter ethnic group	*Misidentified*	*Correctly identified*	*Total*
Upper Caste	119	216	335
	35.5%	*64.5%*	*100%*
OBC	327	696	1,023
	32%	*68%*	*100%*
SC	107	253	360
	29.7%	*70.3%*	*100%*
Muslim	43	282	325
	13.2%	*86.8%*	*100%*
Total	596	1,447	2,043
	29.2%	*70.8%*	*100%*

SOURCE: By author.
NOTE: For the purposes of this table, identification is considered accurate if voters identify the correct *jati* of the candidate.

mobilization in Bihar takes place on jati-specific lines, as opposed to broader caste-based markers. Due to the degree of caste polarization in Bihar—and in much of North India—parties have increasingly engaged in a process referred to as "social engineering" to entice specific jatis to join their core party supporters and broaden their appeal.[39]

Are Certain Groups Better (or Worse) at Identification?

A final issue in this exploratory section is whether there are systematic differences in the ability of certain types of voters to correctly identify the identity of their preferred candidate. The previous subsection argued that when it comes to the identity of candidates, there are small differences in misidentification rates across candidate types—with the notable exception of Muslim candidates. But what about misidentification rates across voter types? Table 4 displays the accuracy rates by voters' overarching ethnic category. Although SC voters are more likely to accurately identify the identity of the candidate they voted for than either OBCs or UCs (with 35.5% of the latter misidentifying their preferred candidate), the differences are not statistically significant. Yet,

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

here again, Muslims appear to be outliers: Muslims are far better at accurately identifying the identity of their preferred candidate. Nearly 87% of Muslim voters correctly assess the identity of their preferred candidate.

DETERMINANTS OF IDENTIFIABILITY

This section explores more systematically the determinants of voters' ability to accurately identify the identity of their preferred candidate. I begin by briefly describing some of the more compelling hypotheses from the literature that are thought to predict accurate identification. Next I describe the model and data used to test these hypotheses, and then conclude this section with the statistical results.

Hypotheses

One of the most consistent findings to emerge from the literature on ethnic identifiability to date is that individuals are significantly more likely to accurately identify fellow co-ethnics.[40] Turning to the experimental evidence, Habyarimana et al. find that in an experimental setting in Uganda, co-ethnicity increases the likelihood of correct identification by eight percentage points.[41] For instance, for the average member of the largest ethnic group in their sample—the Baganda—the probability of correctly identifying a fellow Baganda is 70%, while the probability of correctly identifying someone from the non-Baganda community is just 20%. Based on an experimental sample from South Africa, Harris and Findley also find that co-ethnicity is a robust predictor of successful ethnic identification.[42]

A second hypothesis relates to access to information. Voters often differ in how much they know about the candidates standing for election. Some voters might not be aware of the personal attributes of the candidate for whom they eventually cast their vote.[43] On the other hand, some scholars have argued that ethnic identity takes on even greater importance for low-information

40. This, for instance, underpins the supposed advantage of co-ethnics in "policing" their own, which can help explain why we often observe cooperation between ethnic groups in divided societies. Fearon and Laitin, "Explaining Interethnic Cooperation," 715–35.

41. Habyarimana et al., "Placing and Passing"; Habyarimana et al., *Coethnicity*.

42. Harris and Findley, "Is Ethnicity Identifiable?"

43. Christopher H. Achen and Larry M. Bartels, *Democracy for Realists: Why Elections Do Not Produce Responsive Government* (Princeton, NJ: Princeton University Press, 2016).

voters because it is a cognitive shortcut that provides voters with useful information about politicians' future behavior without requiring them to learn about specific policy positions.[44] Both hypotheses could be advanced in the Indian context, where ethnicity motivates a great deal of politics but literacy, media penetration, and political awareness exhibit significant variation.

A third hypothesis suggests that voters who have strong partisan identities will place relatively less emphasis on identity considerations because they are primarily concerned with partisan, rather than candidate-level, considerations.[45] As a result, strong partisan identifiers will be less likely to invest in learning about candidates' backgrounds. Under such conditions, the identity of the candidate takes a backseat to the identity of the party.

Model and Data

To formally test the hypotheses about what determines the variation in identifiability, I estimate a series of linear probability models of the form:

$$Accuracy_{ip} = \beta_0 + \beta_1 Coethnicity_{ip} + \beta_2 Education_{ip} + \beta_3 Copartisan_{ip} + \beta_4 X_{ip} + \alpha_p + \varepsilon_{ip}$$

The dependent variable, *Accuracy*, is a binary measure of accurate identification, equal to 1 if a voter's perception of a candidate's identity matches his or her actual identity (and 0 otherwise). *Co-ethnicity* is a binary variable equal to 1 if the voter's caste (jati) identity matches the actual caste identity of the candidate. In the survey, we do not have data on respondents' political awareness or media habits. However, following other researchers, we use an individual's level of education as a proxy for their media exposure and overall political awareness.[46] *Education* is a categorical variable with four categories: nonliterate, primary-educated, high school-educated, and college-educated. From this, I create separate indicator variables for each category. To capture a voter's

44. Chandra, *Why Ethnic Parties Succeed*; Posner, *Institutions and Ethnic Politics*; Karen E. Ferree, *Framing the Race in South Africa: The Political Origins of Racial Census Elections* (New York: Cambridge University Press, 2010).

45. Michael Bratton, Ravi Bhavnani, and Tse-Hsin Chen, "Voting Intentions in Africa: Ethnic, Economic or Partisan?" *Commonwealth & Comparative Politics* 50:1 (February 2012): 27–52.

46. R. Michael Alvarez, *Information and Elections* (Ann Arbor: University of Michigan Press, 1998).

partisan orientation, I construct a binary variable, *Copartisan*, which is 1 if the voter voted for candidates of the same party in the 2010 state elections and the prior 2009 national elections.[47] X is a vector of voter-level controls for factors such as the voter's general ethnic category, age, gender, and a normalized index of household assets. The unit of analysis is the individual voter, i, who is nested within a given polling booth, p. All models include fixed effects for polling booths as well as standard errors that are clustered at the level of the polling booth. Including a fixed effect term for polling booths helps account for any unobserved variation across polling booths that is relatively time-invariant, thereby mitigating the effect that confounding contextual variables might otherwise have on identification rates. For instance, factors such as the information environment or the intensity of political competition are unlikely to vary at the level of the polling booth.[48]

RESULTS

Column 1 of Table 5 reports the regression results of a bivariate regression of *Accuracy* on *Co-ethnicity*. In this bivariate specification, co-ethnicity increases the probability of accurate identification by 16% ($p < .01$). In a multivariate regression with controls, as shown in Column 2, the effect of co-ethnicity remains unchanged. These results seem to support the earlier findings of Habyarimana et al. and Harris and Findley, who also detect a co-ethnic advantage in identifiability. While most of the controls are insignificant, the ethnic group indicator variables demonstrate an interesting pattern. Compared to Muslims (the reference category), Hindus across caste groups (SCs, OBCs, and UCs) are all significantly *less* likely to accurately identify the jati of their preferred electoral candidate.

Columns 3 and 4 of Table 5 examine the hypothesis concerning the impact of education on ethnic identification. The models contain indicator variables for three categories of education status: non-literate; primary-educated; and high school-educated (college-educated is the reference category). Both models report no impact of education on accuracy. While the magnitudes of the coefficients increase as education level decreases, the point estimates are

47. This is not an ideal measure of partisanship; one would prefer a more direct measure or at least one based on a longer track record of voting. Unfortunately, this is the best one can do with the present survey data.

48. Summary statistics can be found in Table A1, in the Appendix.

TABLE 5. Co-ethnicity and Identification

DV:	*−1* *accuracy*	*−2* *accuracy*	*−3* *accuracy*	*−4* *accuracy*	*−5* *accuracy*	*−6* *accuracy*	*−7* *accuracy*	*−8* *accuracy*
coethnic	0.16***	0.16***					0.15***	0.15***
	(0.04)	(0.04)					(0.04)	(0.04)
nonliterate			0.03	0.03			0.04	0.04
			(0.03)	(0.04)			(0.04)	(0.04)
primary			0.02	0.02			0.02	0.03
			(0.04)	(0.03)			(0.03)	(0.03)
metric			0.01	0.02			0.02	0.02
			(0.03)	(0.03)			(0.03)	(0.03)
copartisan					−0.06	−0.05	−0.07*	−0.07*
					(0.04)	(0.04)	(0.04)	(0.04)
voter_SC		−0.10*		−0.12**		−0.11*	−0.10*	−0.10*
		(0.06)		(0.06)		(0.05)	(0.06)	(0.06)
voter_OBC		−0.11**		−0.12**		−0.11**	−0.11**	−0.11**
		(0.05)		(0.05)		(0.05)	(0.05)	(0.05)
voter_UC		−0.18***		−0.18***		−0.19***	−0.17***	−0.17***
		(0.06)		(0.06)		(0.06)	(0.06)	(0.06)
age_upto25		−0.03		−0.02		−0.03	−0.02	−0.01
		(0.03)		(0.03)		(0.03)	(0.03)	(0.03)
age_26_35		−0.03		−0.02		−0.03	−0.02	−0.02
		(0.02)		(0.03)		(0.02)	(0.03)	(0.03)
age_36_45		−0.03		−0.02		−0.02	−0.03	−0.03
		(0.02)		(0.02)		(0.02)	(0.02)	(0.02)
age_46_55		−0.02		−0.01		−0.02	−0.01	−0.01
		(0.02)		(0.02)		(0.02)	(0.02)	(0.02)
male		−0.00		0.00		−0.01	0.01	0.00
		(0.02)		(0.02)		(0.02)	(0.02)	(0.02)
asset index norm		0.01		0.02		0.02	0.02	0.02
		(0.01)		(0.02)		(0.02)	(0.01)	(0.01)
rural								0.09
								(0.10)
resident 10 years								0.06
								(0.04)

(continued)

TABLE 5. *(continued)*

DV:	*−1 accuracy*	*−2 accuracy*	*−3 accuracy*	*−4 accuracy*	*−5 accuracy*	*−6 accuracy*	*−7 accuracy*	*−8 accuracy*
Constant	0.69***	0.81***	0.69***	0.82***	0.74***	0.87***	0.82***	0.69***
	(0.01)	(0.05)	(0.02)	(0.06)	(0.02)	(0.05)	(0.06)	(0.11)
Observations	2,043	2,042	2,013	2,012	2,043	2,042	2,012	2,012
R-squared	0.02	0.04	0.00	0.02	0.00	0.02	0.04	0.04
Polling booths	159	159	159	159	159	159	159	159

SOURCE: By author.

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

NOTE: All models include fixed effects for polling booths. Standard errors clustered at the polling booth level are in parentheses. Age group 55 and over is the reference category for age. Muslim is the reference category for voter ethnic identity.

themselves insignificant. Both models, therefore, provide inconclusive evidence with regard to the debate over access to information and ethnic identity. However, as with the previous models, Muslims continue to outperform Hindus of all castes when it comes to identifiability.

The next set of regression results, testing the copartisan hypothesis, is found in Columns 5 and 6. In both the bivariate and the multivariate cases, there is a negative—but statistically insignificant—impact of copartisanship on identifiability. Again, this result should be treated with caution, given that the copartisan variable is simply a measure of whether a voter voted for the same party in two consecutive elections. The null result perhaps should not be entirely surprising, given the scholarly literature on anti-incumbency in India and, thus, the existence of a notoriously fickle electorate.[49] Indeed, only 61 of the voters in the sample voted for the same party in the 2009 and 2010 elections, which were held just 18 months apart.

In Columns 7 and 8 of Table 5, I estimate a full model in which I include all explanatory and control variables on the right-hand side. The results in Column 7 indicate that co-ethnicity continues to be a robust predictor of accurate identification, even after controlling for alternative hypotheses. Hindu respondents continue to under-perform Muslims when it comes to

49. Yogesh Uppal, "The Disadvantaged Incumbents: Estimating Incumbency Effects in Indian State Legislatures," *Public Choice* 138:1 (January 2009): 9–27.

accurate identification. For the first time, the negative coefficient on the copartisanship variable is significant, although only at the 90% level. As in the previous models, education has no clear relationship with accurate identification.

In Column 8, I include additional control variables for a respondent's rural/urban location and whether he or she has been a resident in the same location for the past 10 years. Neither appears to have a significant impact on accurate identification. The inclusion of these additional controls has no impact on the results. Across all models, age, gender, and assets hold no obvious predictive value.

As a final test, I explore whether ethnic identifiability is higher for ethnic parties that have very clear caste constituencies.[50] One might hypothesize that voters would be more likely to accurately identify candidates fielded by ethnic parties, since these parties have very well-known "party brands," typically based on the identity of their charismatic party leader. This does not appear to be the case; in fact, the accuracy rate is 10 percentage points lower for candidates from ethnic parties. One reason for this could be that while party brands are well-known, many ethnic parties field a diversity of candidates to broaden their caste appeal. Indeed, our data on candidates' identities across parties confirm this.

The Muslim Puzzle

The evidence thus far suggests that Muslims are better at identifying, and at being identified, than individuals belonging to other ethnic categories. Muslim candidates standing for election are more easily identifiable by voters of all stripes. Unsurprisingly, Muslim voters accurately identify Muslim candidates at high rates (97.7%), but non-Muslim voters also do quite well: 94.6% for SCs; 92.1% for UCs; and 86.3% for OBCs.

One simple explanation for the identifiability of Muslim candidates could be the distinctiveness of Muslim (as compared to Hindu) names. If Muslim names are distinctive, it would not be surprising that the social identity of Muslim candidates sticks out in voters' minds. There are 131 candidates in the sample, with 58 distinct last names. Muslim candidates have 19 different last names, and these names are exclusively "Muslim"; no non-Muslim candidate

50. I code four parties as "ethnic" parties: BSP, JD(U), LJP, and RJD. Each has a well-defined core constituency (BSP, Chamars; JD(U), non-Yadav OBCs; LJP, Paswans; and RJD, Yadavs and Muslims).

bears any of these last names. However, when one looks at the 39 last names belonging to non-Muslims, in several instances last names are shared by multiple caste groups. Three last names are shared by candidates of two Hindu caste groups; for instance, Singh is used by 12 UC candidates as well as five OBC candidates, while another two last names are shared by three Hindu caste groups (Devi and Kumar are used by UC, OBC, and SC candidates).

While name distinctiveness might explain why Muslim candidates are more easily identified, it cannot explain why Muslim voters are also able to accurately identify the ethnic identity of the preferred candidate more accurately than other ethnic groups. The sample contains 325 Muslim voters, and their accuracy rate is 86.8%; in contrast, non-Muslim voters accurately identify candidates just 67.8% of the time. This finding becomes more intriguing when one considers that Muslims vote for fellow Muslims only 40.9% of the time; more often than not, Muslims are voting for candidates outside their ethnic community—yet they are still able to successful identify their ethnic identity. In fact, the accuracy of Muslims when they are asked to identify fellow co-ethnics is even higher: 94.3% (compared to 85.8% for non-Muslims asked to identify co-ethnics).

With the available data, it is difficult to identify the precise mechanism underlying this finding. However, there are at least two possibilities. It is possible that a broader sense of social distinctiveness—a feeling of other-ness—experienced by Muslims in India's majority-Hindu society could sharpen their awareness of the social identities of politicians more generally. While this could be the case, this distinctiveness cannot be explained purely in socioeconomic terms.[51] When one compares the socioeconomic characteristics of Muslim and non-Muslim voters in the dataset on several dimensions—educational attainment, mother's and father's literacy, index of household assets, monthly income, and housing quality—Muslims fare worse on four of the six measures. However, when one compares Muslim respondents to SC respondents, another traditionally marginalized social group, Muslims actually do better on most socioeconomic indicators.

A second hypothesis is that Muslim voters are less likely to make identification errors because the distribution of their guesses is much narrower.

51. The scholarly literature is replete with studies on the social and economic exclusion of Muslims in India's majority-Hindu society. Abusalleh Shariff, *Inclusive Development Paradigm in India: A Post-Sachar Perspective* (Washington, DC: U.S.-India Policy Institute, 2012).

Historically, candidates in Bihar are selected from a small number of large and traditionally dominant groups. Muslims, perhaps by virtue of their being "outsiders," could be knowledgeable about these large, dominant groups but less well-versed in many of the smaller jatis. Hindus, conversely, are possibly more steeped in the intricacies of individual jatis. The irony is that this greater level of familiarity provides a wider choice set to Hindu voters, leading them to make a wider variety of (often wrong) guesses, whereas Muslims stick to the most obvious jati labels. The latter strategy results in better identification because of the relative dominance of a small number of jatis. Indeed, the data in Table 6 suggest that the distribution of guesses made by Hindus is much wider than that of Muslims.[52]

CONCLUSION

The objective of this paper is to test the bedrock assumption that an individual's ethnic identity is readily identifiable by others. It does so through an assessment of how well this assumption holds up in the context of voting. Using an original dataset of voters and candidates from North India, this paper shows that while the identifiability assumption holds in the majority of instances, it is not foolproof. Voters frequently misidentify the ethnic identity of their preferred political candidate. While the evidence presented here suggests that errors tend to occur within large umbrella caste categories, political mobilization in Bihar largely takes place along jati lines, not within these overarching categories. Indeed, these fine distinctions have tremendous political relevance in contemporary Bihar. One of Nitish Kumar's most lasting political legacies, originating from his first term in office, was to drive a wedge between the Yadavs (the core supporters of his rival Lalu Prasad Yadav and his RJD party) and other backward communities. Importantly, this paper shows that errors in misclassification are not random and that their structure can be modeled.

This paper contributes to a budding literature that questions the universal validity of the ethnic-identifiability assumption. Unlike previous studies, however, this paper situates this identifiability assumption in the realm of

52. A final hypothesis is that Muslims are better identifiers because they have more interaction and contact with political candidates, given that Muslims form a crucial bloc of swing voters (accounting for roughly 17% of the population) in Bihar. Parties strategize intensely on how best to woo "the Muslim vote," and Muslim voters might be more sensitive to the identities of those seeking to woo their vote. Although I cannot evaluate this hypothesis with this data, it deserves further exploration.

TABLE 6. Distribution of Responses, Muslim vs. Non-Muslim Voters

a) Muslim voters	*Frequency*	*Percent*	*b) Non-Muslim voters*	*Frequency*	*Percent*
Muslim	137	42.2	Yadav	414	24.1
Yadav	56	17.2	Bhumihar	260	15.1
Koeri	48	14.8	Rajput	204	11.9
Rajput	28	8.6	Muslim	200	11.6
Pasi	19	5.9	Koeri	107	6.2
Bhumihar	11	3.4	Pasi	89	5.2
Brahmin	6	1.9	Kurmi	83	4.8
Lowest SC	5	1.5	Chamar	82	4.8
Kewat	4	1.2	Brahmin	81	4.7
Other Upper Caste	3	0.9	Lowest SC	41	2.4
Kurmi	3	0.9	Dhobi	33	1.9
Other Craftsmen OBC	2	0.6	Teli	28	1.6
Dhobi	2	0.6	Sunar	26	1.5
Chamar	1	0.3	Other Peasant OBC	19	1.1
			Other Craftsmen OBC	19	1.1
			Kewat	12	0.7
			Trader OBC	8	0.5
			Other Upper Caste	4	0.2
			Gaderia	2	0.1
			Kayastha	1	0.1
			Kunbi	1	0.1
			Badhai	1	0.1
			Nai	1	0.1
			Vaishya	1	0.1
			Nomadic/Service SC	1	0.1

SOURCE: By author.

electoral politics and relies on observational data from a real election campaign, in contrast to simulated data from an experiment. Given the emphasis in political science on the importance of ethnicity in shaping voting behavior, it is remarkable that so few studies have interrogated the identifiability assumption in this context.

This paper's findings should not be interpreted as denying the reality that ethnic identity continues to be a powerful factor in shaping voter behavior in elections in India and elsewhere. After all, there is a large body of social

science research that backs this up. However, it may be time to revise our models of *how* ordinary voters interact with ethnicity. If the ubiquitous "ethnic signal" is noisier than we have been led to believe, are failures of identifiability consequential for outcomes of interest to political scientists? That is, if voters believe they are voting for a politician from the OBC community, but are in fact voting for a UC candidate, what effect does this have on politics? Do such errors influence voters' evaluations of politicians? Do they affect voters' expectations of what politicians should or should not deliver while in office? This study also raises intriguing hypotheses regarding the role the perceived degree of "dissimilarity" plays in ethnic identification. For example, an OBC Hindu may see themselves as "closer" to SC Hindus than to Muslims or UC Hindus. Might the perceived degree of dissimilarity between the ethnicity of a given candidate and one's own ethnic identity influence identification? For instance, the degree of dissimilarity could make a candidate's ethnic identity more or less salient, thus affecting the probability of correct identification. The answers to these questions have implications for how we understand what makes voters in multi-ethnic societies tick.

APPENDIX

FIGURE A1. Typology of Ethnic Groups in Bihar

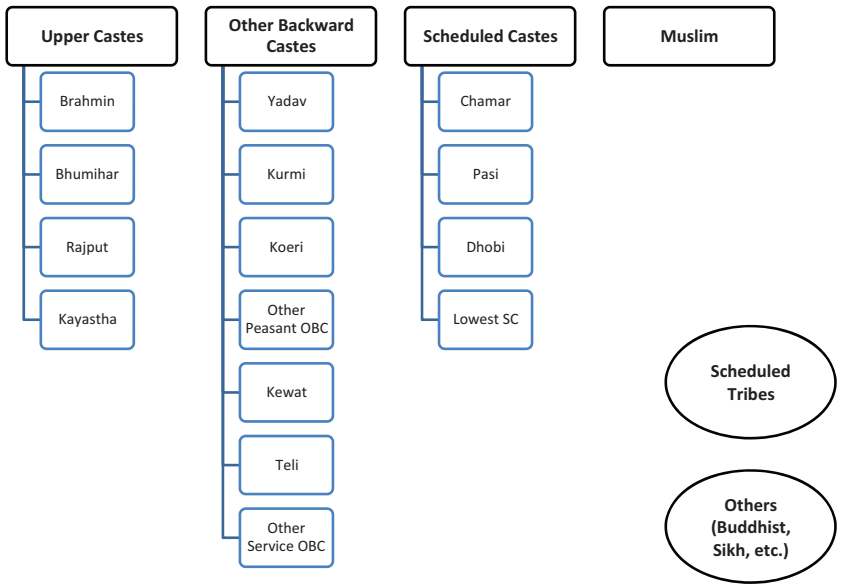

TABLE A1. Summary Statistics

Variable	*Obs*	*Mean*	*Std. dev.*	*Min*	*Max*
accuracy	2,043	0.71	0.45	0	1
group_accuracy	2,043	0.89	0.31	0	1
coethnic	2,043	0.14	0.35	0	1
group_coethnic	2,043	0.39	0.49	0	1
age_upto25	2,042	0.16	0.37	0	1
age_26_35	2,042	0.30	0.46	0	1
age_36_45	2,042	0.23	0.42	0	1
age_46_55	2,042	0.15	0.36	0	1
age_56_above	2,042	0.16	0.36	0	1
f_lit	1,953	0.41	0.49	0	1
m_lit	1,890	0.20	0.40	0	1
nonliterate	2,013	0.37	0.48	0	1
primary	2,013	0.18	0.39	0	1
metric	2,013	0.27	0.44	0	1
college	2,013	0.18	0.38	0	1
rural	2,043	0.85	0.35	0	1
voter_muslim	2,043	0.16	0.37	0	1
voter_OBC	2,043	0.50	0.50	0	1
voter_ST	2,043	0.00	0.00	0	0
voter_SC	2,043	0.18	0.38	0	1
voter_UC	2,043	0.16	0.37	0	1
voter_Others	2,043	0.00	0.00	0	0
copartisan	2,043	0.61	0.49	0	1
resident10_years	2,043	0.94	0.23	0	1
qualityhouse	2,018	0.30	0.46	0	1
hindu	2,043	0.83	0.37	0	1
muslim	2,043	0.16	0.36	0	1
religion_other	2,043	0.01	0.09	0	1
male	2,043	0.54	0.50	0	1
assetindexnorm	2,043	0.00	1.00	−1.05	4.11