The Expression of Religious Identity in India: Evidence from 505 Million Voter Records

Feyaad Allie[*]
Viktor Enssle[†]
Tanushree Goyal[‡]
Saad Gulzar[§]
Gufran Pathan[¶]

Research Note
May 2026

Abstract

We provide a theoretically grounded descriptive analysis of religious identity expression in India by examining personal names in over 505 million voter records. Standard theories of identity predict asymmetry: under rising discrimination, minorities should moderate visible markers while majorities express identity freely. In contrast, we find that both Hindus and Muslims show similarly high and increasing levels of religious distinctiveness in names. To characterize this rise, we examine the divine and social components of religious identity. We find no increase in doctrinally rooted names, inconsistent with growing religious devotion as an explanation. Instead, both groups increasingly draw from a smaller set of common names, and identity expression is stronger in segregated areas for the locally dominant group, highlighting the social function of religious identity. Together, these patterns reveal a precarious calculus in which minority moderation no longer holds, and cultural divides intensify alongside, and ahead of, majoritarian electoral politics

Keywords: Religious identity, inter-group relations, India, Hindu-Muslim relations, Social identity

[*]Department of Government, Harvard University
[†]Department of Politics and International Relations and University College, University of Oxford
[‡]Department of Politics, Princeton University
[§]Department of Politics and Keough School of Global Affairs, University of Notre Dame
[¶]CEGA, University of California, Berkeley

1 Introduction

Religion is among the world's most powerful identities. Nearly three-quarters of the global population identifies with a religious tradition,[1] and in societies split between religious majorities and minorities, it often shapes political and economic life. This is especially consequential in India, home to the world's largest Muslim minority (14 percent), where the Hindu-Muslim divide is a central axis of social and political organization. Hindu identity is increasingly tied to national belonging,[2] while Muslims face growing violence, discrimination, and socio-economic marginalization.[3] In this context, visible religious identity carries direct stakes for both exposure to risk and claims to group belonging. Yet how religious identity is publicly expressed, and what shapes that expression among majority and minority groups, remains poorly understood.

This research note documents high levels of religious identity expression in India using personal names data from 505 million voters. Names signal ethnic visibility (Robinson, 2023). They are durable markers that signal social identity and shape individuals' economic opportunities and well-being.[4] Contrary to expectations of asymmetry of identity expression between majorities and minorities, we show that Muslims are nearly as likely as Hindus to express religious identity, and religious expression increases for both groups over time. The final part of the paper documents how, instead of being rooted in growing religious fundamentalism, as is commonly postulated, the

[1] Lowes et al. (2025); Hackett et al. (2025); Pew (2018).

[2] Nearly two-thirds of Hindus say it is very important to be Hindu to be "truly" Indian (Pew Research Center, 2021, p.6).

[3] A substantial body of scholarship documents these dynamics. See, for example, studies of communal violence (Wilkinson, 2004; Varshney, 2003; Allie, 2023; van Noort and Goyal, 2025); discrimination in housing, labor markets, and everyday life (Jaffrelot and Ahmed, 2024; Evans and Sahgal, 2021; Datta and Pathania, 2016; Majumdar, 2023; Nellis et al., 2024); political exclusion and under-representation (Allie, 2025, 2026; Farooqui, 2020; Ahmed, 2019); and intergenerational mobility and economic inequality (Asher et al., 2024).

[4] Names are known to shape access to labor markets, housing, dating, and social life across a range of contexts (Bertrand and Mullainathan, 2004; Rubinstein and Brenner, 2014; Ahmad, 2020; Shadab et al., 2022; Thorat, 2015; Ge and Wu, 2024).

heightened identity expression we document is more closely associated with social factors like signaling and segregation. The broader takeaway is that religious polarization in democracies can deepen through everyday cultural choices well before it appears in elections, and that minorities facing majoritarian pressure may have less room to manage their visibility than standard accounts assume.

We begin from social identity theory, which posits that individuals derive utility from expressing salient identities but must weigh these benefits against social and intrinsic costs (Tajfel and Turner, 1979; Shayo, 2009; Bursztyn et al., 2020; Akerlof and Kranton, 2000). In unequal and polarized contexts, these costs are asymmetric: majorities can express identity at low risk, while minorities face heightened exclusion and exposure to discrimination (Atkin et al., 2021a). As inequality and group-based tensions rise, this asymmetry intensifies. The framework, therefore, predicts high baseline religious identity expression and rising asymmetry.

Building on these microfoundations, our core contribution is to provide a theoretically grounded descriptive analysis of religious identity expression at an unprecedented scale, answering a call from de Kadt and Grzymala-Busse (2025). To assess these asymmetric expectations, we assemble a dataset of personal names from voter rolls comprising over 505 million registered voters across 27 states and 6 Union Territories in India, covering birth cohorts from 1950 to 1995. We use a machine-learning algorithm to infer individuals' religion from their parents' names and then construct a Religious Name Index (RNI) based solely on individuals' personal names. The RNI measures the relative adoption of each child's personal name by Muslim and Hindu parents in a specific state and birth year, ranging from 0 (distinctively Hindu) to 1 (distinctively Muslim).

We find partial support for the theoretical expectations. Consistent with the social identity theory framework, religious identity expression is high among both Hindus and Muslims: more than 80 percent of parents in each group choose religiously distinctive names. However, contrary to expectations of asymmetry, Muslims are nearly as likely as Hindus to express religious identity, and identity expression rises over time for both groups and consolidates across generations rather than declines.

Why does the minority group express its identity to a similar degree as the majority? We draw on two complementary accounts from the study of religion and ethnic politics. The first emphasizes a divine logic: religion is a distinctive identity that confers transcendental meaning and rewards, and rising identity expression through names reflects these growing religious commitments (Grzymala-Busse, 2012; Gill, 2001; Philpott, 2007; McCauley, 2014). The second emphasizes a social logic, in which religious identity operates as a form of ethnic identity, and visible markers therefore facilitate coordination, solidarity, and sanctioning within and between groups (Laitin, 1998; Posner, 2005; Chandra, 2004; Habyarimana et al., 2007). In this view, visible identity markers such as names carry both benefits and costs. In highly segregated and conflict-prone settings, this means that minorities face a precarious calculus: moderating identity may reduce discrimination, but it can also invite punishment from in-group and out-group members who want to maintain group boundaries, while also policing opportunistic or disloyal behavior.

Both perspectives yield testable predictions that can operate simultaneously. The divine logic implies persistently high baseline levels of religious naming and rising adoption of names with religious origins. The social logic predicts a growing preference for common names and stronger Muslim identity expression in more segregated neighborhoods, where coordination and sanctioning are intense and concealment yields limited benefits. We find that although the divine logic accounts for high baseline expression, it cannot explain its rise over time; instead, the increasing adoption of common, non-religious names and stronger minority expression in segregated settings point to the importance of the social logic.

Our findings challenge the theoretical expectation that marginalized minorities may conceal or lower the visibility of their identity expression (Laitin, 1998; Fouka, 2019). But, rather than interpreting heightened identity expression among minorities solely as defiance or resistance, we offer another interpretation: minorities may have limited room to navigate their vulnerability. In increasingly polarized and segregated settings, where attempts to lower visibility are easily detected and sanctioned or perceived as insincere and opportunistic, the marginal returns to lowering visibility may be negligible or even reversed. Under such conditions, moderation ceases to function as a viable strategy, and standard asymmetric expectations about majority–minority behavior may fail to hold. By grounding large-scale descriptive patterns in explicit theoretical expectations, this study demonstrates how theory-driven description can sharpen existing understandings of identity expression (de Kadt and Grzymala-Busse, 2025).

Empirically, this paper broadens the scope of studying minority identity formation beyond the cases of native-immigrant divides in the US and Europe (Adida and Robinson, 2023; Fouka, 2019; Adida et al., 2016; Choi et al., 2022). In this way, we join recent work that has begun to document how minorities may reinforce their identities in precarious majority-minority contexts (Portes and Rumbaut, 2001; Fouka, 2020). Amidst rising ethno-nationalism across the world, not only immigrants but also minority citizens face new challenges to express their identity while being part of an increasingly contested national community (Haas and Lindstam, 2024).

We also contribute to the study of religion and politics (Grzymala-Busse, 2012; McClendon and Riedl, 2019; Tuñón, 2019). By focusing on a majority-minority divide along a religious cleavage, we advance the view from Grzymala-Busse (2012) to take the divine mechanisms of religious identity seriously beyond social components that are often associated with ethnic identity. In offering a tractable indicator of religious identity, we advance previous work that focuses on measuring everyday religious behavior, such as attendance at places of worship, adherence to religious norms, and observance of religious holidays (Atkin et al., 2021b; Dube et al., 2022; Abdelagadir and Fouka, 2020).

Varshney (1998) notes that "the only cleavage that has the potential to rip India apart is the divide between Hindus and Muslims" (Varshney, 1998, 44). We contribute to the study of Indian politics by examining this divide. We document a long-term intensification of religious identity expression during the dawn of right-wing majoritarian politics in India (Varshney, 2003; Wilkinson, 2004; Haas and Lindstam, 2024; Allie, 2025). Our measure provides insights into cultural aspects of majoritarian politics that move beyond the electoral success of Hindu nationalism (Baral et al., 2023; Iyer and Shrivastava, 2018). In the decades preceding the Bharatiya Janata Party's national dominance, visible religious distinctiveness increased for both Hindus and Muslims, highlighting the social roots of India's majoritarianism.

2 Theory

2.1 Asymmetric and divergent religious identity expression

Social identity theory holds that individuals derive utility from identifying with salient groups and conforming to group-prescribed norms, while suffering disutility from deviation (Tajfel and Turner, 1979; Akerlof and Kranton, 2000; Shayo, 2009). Religion is among the most powerful of such identities, with the majority of the world's population identifying with a religious tradition (Hackett et al., 2025). When religious affiliation forms part of the self, visible expression carries intrinsic value even as its external costs vary across groups and contexts.

While many factors shape name choice, such as gender, we focus on religion. Names are public, durable, and costly to change (Fryer and Levitt, 2004; Fouka, 2019), making them a clearer window into identity than malleable practices like dietary choices or grooming (Atkin et al., 2021a; Carvalho, 2013; Abdelagadir and Fouka, 2020). Prior work documents naming as a marker of identity assertion or assimilation across settings (Fryer and Levitt, 2004; Abramitzky et al., 2020; Fouka, 2019; Kuipers et al., 2025; Assouad, 2020).

Applied to India, the framework yields a baseline expectation that both Hindus and Muslims should exhibit high levels of religious naming, since religion confers normative meaning and parents have incentives to conform to community conventions. However, identity expression is also shaped by group status (Shayo, 2009), where majorities enjoy institutional dominance and cultural authority, so visible religious expression reinforces status at low risk. Minorities, by contrast, weigh the psychological benefits against the costs of discrimination, exclusion, and violence (Atkin et al., 2021b). In India, Muslims face documented disadvantages in housing and labor markets, political under-representation, and periodic communal violence (Wilkinson, 2004; Varshney, 2003; Jaffrelot and Ahmed, 2024; Evans and Sahgal, 2021; Farooqui, 2020), and religious names themselves shape perceptions in hiring, housing, and everyday interaction (Shadab et al., 2022; Thorat, 2015). If visibility raises exposure to exclusion, minorities face a trade-off that majorities largely avoid.

We derive two expectations from this logic. First, minorities should exhibit lower visible religious identity expression than the majority. Second, as polarization increases and socio-economic inequalities widen, the cost of minority expression should rise, leading minorities to moderate over time relative to the majority. The framework therefore predicts both asymmetry in levels and increasing divergence over time.

As we will document below, our empirical analysis reveals a different pattern. As expected by theory, religious distinctiveness is high in India. However, contrary to predictions, expression is high for both Hindus and Muslims, and is also increasing over time. That is, rather than observing minority moderation and widening divergence, we find a parallel rise in identity expression. Taking these findings as a point of departure, we turn to the study of religion and ethnic politics to explain these patterns.

2.2 Symmetric and rising religious identity expression

2.2.1 Divine benefits

One mechanism that can sustain high levels of visible identity expression among both majorities and minorities is the divine nature of religion itself. Religious identity is not only ethnic but uniquely *sacred*. Religion makes comprehensive, supernatural claims over believers' lives and ties identity to "supernatural goods of salvation and nirvana" and "eternal damnation" (Grzymala-Busse, 2012, 423). Names often reference divine attributes, prophets, saints, or scriptural figures, and parents may select such names for spiritual or transcendental reasons (Grzymala-Busse, 2012; Wald et al., 2005). Religious teachings can shape understandings of identity transmission and moral obligation (McClendon and Riedl, 2019). In this divine logic, naming reflects devotion and the reproduction of sacred tradition rather than a strategic response to external incentives.

This perspective helps explain high baseline levels of religious naming among both Hindus and Muslims. If naming is grounded in sacred and spiritual commitments, religious identity expression may persist even when visibility carries social risk. Importantly, this logic also generates implications for temporal change. If rising distinctiveness reflects intensifying religiosity or doctrinal emphasis, we should observe growth in names with religious origins, such as those referencing deities, prophets, saints, or scripture.

2.2.2 Social benefits and punishment

Another complementary explanation for high levels of religious identity expression builds on studies of ethnic politics, which treat religion as a socially embedded ethnic identity that structures coordination, trust, and sanctioning (Fearon and Laitin, 1996; Posner, 2005; Chandra, 2004). While social identity theory emphasizes the psychological utility individuals derive from group membership and norm conformity, work in ethnic politics highlights the instrumental and institutional functions of identity as a mechanism for coordination, monitoring, and sanctioning. The former explains why identity expression carries intrinsic value; the latter explains how visible markers structure social interaction. In this framework, visible markers such as names facilitate recognition and reciprocity within dense ethnic networks (Habyarimana et al., 2007; Robinson, 2023). Consequently, conformity signals commitment and solidarity with the in-group, while deviation is observable and capable of being sanctioned.

Crucially, deviations can generate punishment not only from in-group but also out-group members. Within in-group communities, adopting an out-group or ambiguous name may be interpreted as disloyalty, inviting social sanction. At the same time, out-group members may view such moderation as opportunistic or insincere, particularly in polarized contexts where boundaries are actively policed. When majority group actors have incentives to maintain clear group distinctions, ambiguous signaling can provoke suspicion or hostility. Public reporting in India documents incidents in which Muslims using ambiguous or Hindu-sounding names faced hostility when their religious identity was later revealed.[5] These episodes illustrate that attempts to moderate visible identity may invite sanction rather than protection.

These dynamics can intensify under residential segregation (Adukia et al., 2019). Where communities are spatially clustered, monitoring is dense, and identities are easily inferred. In such settings, concealment yields few benefits. If religious identity can be inferred from where you live, adopting ambiguous or out-group names neither meaningfully reduces discrimination nor avoids scrutiny, while still risking punishment and foregoing the benefits of in-group signaling. Visible expression may therefore become the safer and more stable strategy.

This social calculus yields two empirical predictions. First, because coordination and solidarity operate through shared conventions, we should observe increasing adoption of common names that function as recognizable in-group signals. Second, minority identity expression should be strongest in highly segregated contexts, where concealment is ineffective and potentially costly. Under such conditions, asymmetric predictions of minority moderation need not hold; instead, visible identity expression may rise for both groups.

3 Data and Methods

To measure religious identity expression among Hindus and Muslims, we construct a Religious Name Index (RNI) based on the relative frequency of personal names across religious groups. The RNI is similarly computed as the Foreign Name Index in Fouka (2019) and the Black Name Index in Fryer and Levitt (2004). Building this measure requires three steps. First, we assemble and

[5] See, for example: The Guardian.

standardize voter roll data across states. Second, we infer each voter's religion from their parent's name using a supervised machine-learning algorithm. Third, we extract voters' personal names and compute the RNI, which captures how distinctively associated each name is with either religious group. We describe each step in turn.

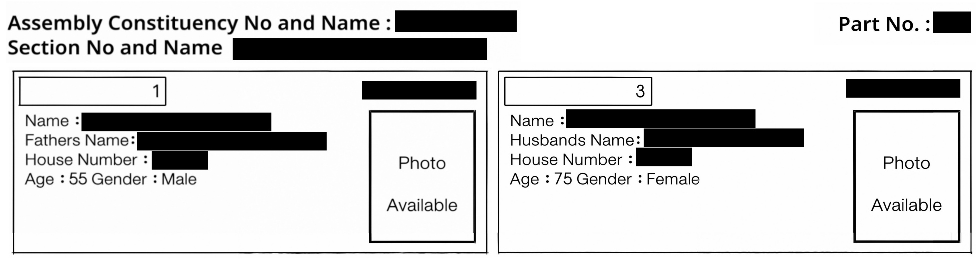
3.1 Voter Rolls

Our data are drawn from publicly available PDF voter rolls published by state Chief Electoral Officers (e.g., `https://electoralsearch.eci.gov.in/`).[6] Figure 1 provides an example of a voter roll and illustrates the key variables available to us, including name and age, the latter of which allows us to calculate birth year.

Voter rolls are compiled by state electoral authorities in regional languages or English, following guidelines set out in the Block Level Officer Handbook (Election Commission of India, 2011). They include each voter's full name, age, gender, location, and father's or husband's name. Our compiled dataset covers 27 Indian states and 6 Union Territories and includes individuals born between 1950 and 1995.

We harmonize variable names across states and standardize formats. To address optical character recognition (OCR)-related errors arising from PDF conversion, we exclude names containing non-alphabetic characters (less than 2 percent of observations). Variation in OCR quality stems primarily from differences in regional written scripts; however, because Hindu and Muslim names within a given state are recorded in the same script, OCR quality is unlikely to be systematically correlated with religion. Moreover, all comparisons are conducted within state, further mitigating concerns that variation caused by written script that vary by state-level biases our estimates. Figure A1 displays the geographic coverage of our data. We do not include the entirety of Chhattisgarh, Ladakh, and Lakshadweep, and most parts of Jammu Kashmir, Karnataka, Assam due to unavailability of voter roll data.

[6] We thank Sood and Dhingra (2018) and Goyal (2021) for sharing the raw scraped data.



Assembly Constituency No and Name : [redacted] Part No. : [redacted]
Section No and Name [redacted]

1 [redacted]
Name : [redacted]
Fathers Name: [redacted]
House Number : [redacted]
Age : 55 Gender : Male
Photo Available

3 [redacted]
Name : [redacted]
Husbands Name: [redacted]
House Number : [redacted]
Age : 75 Gender : Female
Photo Available

Figure 1: **Voter Rolls in PDF Format.** Voter rolls are published as PDF documents containing sensitive personal information on voters and their relatives.

A first concern with our use of the voter rolls is whether they are representative of the adult citizen population. Voter registration rates in India are extremely high. Of the 912 million adults eligible to vote, approximately 869 million are registered, implying a registration rate of roughly 98 percent.[7] Our dataset covers about 56 percent of registered voters, primarily because voter roll records are unavailable for several states (see Figure A1). To further assess representativeness, we compare the demographic composition of our voter roll data to the 2011 Census (Office of the Registrar General and Census Commissioner, India, 2011) (see Appendix B). Figure A2 shows close correspondence across age cohorts and religious groups. Because voter rolls include only individuals aged 18 and above, coverage is lower in the youngest birth cohort. The remaining gap reflects approximately 10 percent of assembly constituencies that are missing in our raw voter roll data. Table A1 shows the state level coverage of ACs and poulation coverage of our sample.

A second concern with our use of the voter rolls is the selective deletion of voters, particularly Muslims, from electoral rolls. First, available evidence suggests that such irregularities are likely geographically concentrated and occur in a small subset of highly competitive constituencies (Das, 2025). Second, even inside these areas, deletion based on identity will be more likely for voters with highly religiously distinctive names. We therefore conclude that the distinctiveness of religious names that we observe below is likely underestimating distinctiveness in a minor way.

[7]Election Commission data indicate that more than 95 percent of eligible adults are registered. Source: *Press Information Bureau* Press Release posted on February 9, 2024.

3.2 Inferring Voter Religion from Parental names

India's voter rolls do not record voters' religion. To address this, we infer religion from the father's full name (or, in the case of married women, the husband's full name). Parental names allow us to avoid mechanically inflating our measure of religious distinctiveness. We can reasonably infer married women's religion from their husbands' religion because religious endogamy in India is extremely high: 99% of Hindus and 98% of Muslims report being married to someone from the same religion (Sahgal et al., 2021).

To classify religion, we adapt the supervised machine-learning algorithm developed by Chaturvedi and Chaturvedi (2022), which exploits linguistically distinct name signatures across religious groups (Ahmad, 2011; Jayaraman and Oak, 2005). We train the model using the nationally representative Rural Economic and Demographic Survey (REDS) dataset (National Council of Applied Economic Research, 2006), which includes individuals' names and self-reported religion.

The algorithm operates in a multi-class framework, predicting the probability that a given name belongs to Hindu, Muslim, Christian, Sikh, Buddhist, or Jain categories. This approach reduces false positives relative to binary classification. Applying the model to our full dataset yields a religious composition of approximately 75% Hindu, 14% Muslim, 4% Christian, 3.4% Sikh, 1.7% Buddhist, and 1.7% Jain. For the purposes of this analysis, we retain only Hindu and Muslim voters, who together comprise roughly 90% of the sample. Figure 2 (A) shows the resulting religious distribution.

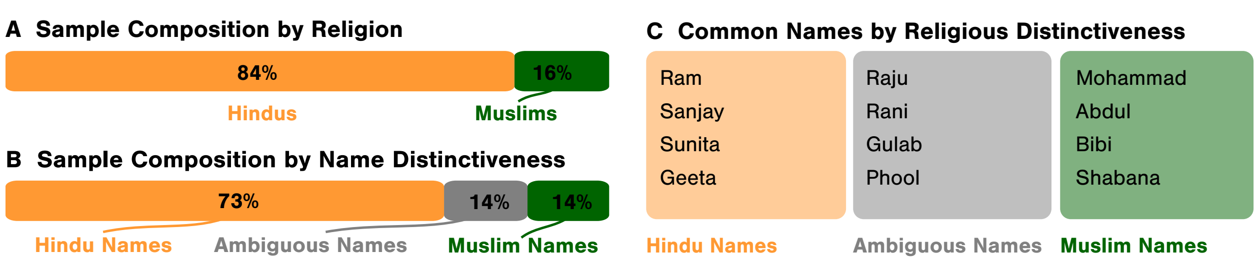


Figure 2: **Sample Summary Statistics. (A)** Religious composition of the sample based on inferred religion. **(B)** Distribution of name distinctiveness categories. **(C)** Examples of common male and female names by distinctiveness category.

How accurate are we in predicting religion? To evaluate classification accuracy, we test the model on an out-of-sample dataset of 20,000 names from rural households following Chaturvedi and Chaturvedi (2022). The algorithm achieves 95.6% overall accuracy. Precision and recall exceed 90% for both Hindu and Muslim classifications. Appendix C.3 presents additional robustness exercises, including restricting the sample to highly confident classifications, which raises accuracy to 98% in the verification dataset.

3.3 Extracting Voter Personal Name

Because last names in India are inherited and reflect lineage, caste, or locality rather than parental choice, we focus on given or *personal* names as the relevant site of identity expression.

	Voter's Full Name	**Father's Full Name**
1 Raw Name Parts	*Mohd Mustak Ali*	*Mohd Aadil Hussain*
2 Remove Overlap	*~~Mohd~~ Mustak Ali*	*~~Mohd~~ Aadil Hussain*
3 Retain First Part	*Mustak ~~Ali~~*	*Aadil ~~Hussain~~*
4 Personal Name	*Mustak*	*Aadil*

Figure 3: **Personal Name Extraction** Illustrates the logic we implement for name extraction in all states but Gujarat, where we retain the second name part as names there are recorded in inverted order.

We extract personal names by removing any overlap between the voter's full name and the parent's name, which may include patronymic, caste, or family identifiers (Figure 3). For example, in some states, the father's or husband's personal name may be present in the voter's name. From the remaining components, we select the first element as the personal name. Most states follow a first–middle–last naming structure (Ahmad, 2011), except Gujarat which follows a family name–personal name ordering. Beyond excluding names with non-letter characters (to address OCR artifacts), we retain all names, including rare and single-letter entries (less than 0.5% of observations), to avoid imposing arbitrary thresholds.

In Appendix C.2, Figure A5 reports summary statistics on name structure across religion and gender. The average number of characters per full name remains stable over time, though Hindu names are slightly longer on average. The number of name components increases modestly over time for both groups. Crucially, these patterns evolve similarly for Hindus and Muslims, alleviating concerns that compositional changes drive our results.

3.4 Religious Name Index (RNI)

We measure religious distinctiveness using a Religious Name Index (RNI), which captures the relative association of a given name with Muslims versus Hindus (Fouka, 2019; Fryer and Levitt, 2004). The RNI is conceptually distinct from the religion classification algorithm: the former measures the name distinctiveness of a voter's name, while the latter assigns religion to individuals based on parental names.

The RNI for name n in state s and birth year t measures the relative share of individuals with that name who are Muslim rather than Hindu, ranging from 0 (distinctively Hindu) to 1 (distinctively Muslim).

First, for each name n (of all N names) in birth year t, we calculate the probabilities of name n in state s being given to a Muslim in year t and a Hindu in year t, $Pr(Name_n|Muslim)_{st}$ and $Pr(Name_n|Hindu)_{st}$ respectively, as:

$$Pr(Name_n|Muslim)_{st} = \frac{\#Muslims\ with\ Name_{nst}}{\#Muslim\ voters_{st}}$$

$$Pr(Name_n|Hindu)_{st} = \frac{\#Hindus\ with\ Name_{nst}}{\#Hindu\ voters_{st}}.$$

We compute these frequencies at the state-by-birth-year level to account for regional variation in naming conventions and religious composition.

Second, we define:

$$RNI_{nst} = \frac{Pr(Name_n|Muslim)_{st}}{Pr(Name_n|Muslim)_{st} + Pr(Name_n|Hindu)_{st}}$$

Third, we use the name-specific $RNI_{n,s,t}$ to calculate the average $RNI_{n,t}$ for all Hindus and Muslims born in year t respectively. This $\overline{RNI}_{religion,t}$, therefore, is weighted by the popularity of name n within the respective religious group.

Intuitively, $\overline{RNI}_{religion,t}$ captures two components of religious distinctiveness: (1) the relative popularity of names across groups and (2) their absolute popularity within groups. For example, the name Munna is somewhat more common among Muslims than Hindus (9:5), resulting in an $RNI_{Munna,1992} = 0.72$. However, because it is rare among both groups (0.05% of Muslims; 0.03% of Hindus), it contributes little to the overall group average.

To enable meaningful interpretation, we label names according to their $RNI_{n,t}$ as follows: values of [0-0.3] are *Hindu names* (0.3-0.7] are *Ambiguous names*, (0.7-1] are *Muslim names* (Jain et al., 2022). While this binning, as shown in Figure 2 (B), affords analytic clarity, we recognize that $RNI_{n,t}$ varies along a spectrum of religious distinctiveness. Figure 2 (C) provides examples of common names in each of these bins.

There might be a concern that our measure of RNI is correlated with the quality of classification for the parents' religion. For example, it might be the case that we measure a lower RNI in cases where Muslim parents' names are predicted with lower certainty. We discuss these concerns extensively in Appendix C.4. Specifically, Figure A11 shows that although there exists a weak correlation between religious prediction confidence and a name's RNI, substantial variation in RNI persists within levels of classification quality, suggesting that measurement error does not mechanically drive our results.

4 Main Results

With data on the names of Indian voters born between 1950-1995, we examine descriptive patterns of Hindu and Muslim naming conventions to determine whether they are in line with predictions of group behavior from social identity theory.

4.1 High and Symmetric Levels of Religious Identity Expression

The partition of the subcontinent along religious lines preserved the Hindu-Muslim cleavage that took shape under British rule and over our 1950–1995 study period, riots, disputes over religious accommodations, and majoritarian mobilization solidified this cleavage as central to Indian society. Consistent with social identity theory and this context, we find high levels of identity expression among both groups. Figure 4 A shows that across the full sample period, 84% of Hindu voters have distinctively Hindu names ($RNI_{n,t}$ between 0 and 0.3) and 81% of Muslim voters have distinctively Muslim names ($RNI_{n,t}$ between 0.7 and 1). This shows that, contrary to the expectation that minorities should display lower rates of visible religious identity, we find that Muslim levels match or exceed Hindu levels. In addition, the share of both groups exhibiting extremely religiously distinct names is also high: 47% of Hindus have names with $RNI_{n,t}$ in $[0, 0.1]$, while 76% of Muslims have names with $RNI_{n,t}$ in $(0.9, 1]$.

A feature of the RNI is that it depends on the joint behavior of both groups, allowing us to probe the asymmetry expectation more finely. First, consistent with the theoretical claim that minorities have incentives to adopt majority markers, we find that Muslims adopt relatively more Hindu names than Hindus adopt Muslim names. This helps explain why Muslim voters may be disproportionately concentrated in the "extremely distinctive" bin as Hindus rarely adopt Muslim names. Second, contrary to the expectation that only minorities have such incentives, both groups adopt ambiguous names at broadly similar rates, with modestly more Hindus doing so. This suggests that identity moderation through ambiguous names may not be a minority-only behavior, but reflects a broader pattern of strategic identity expression across groups.

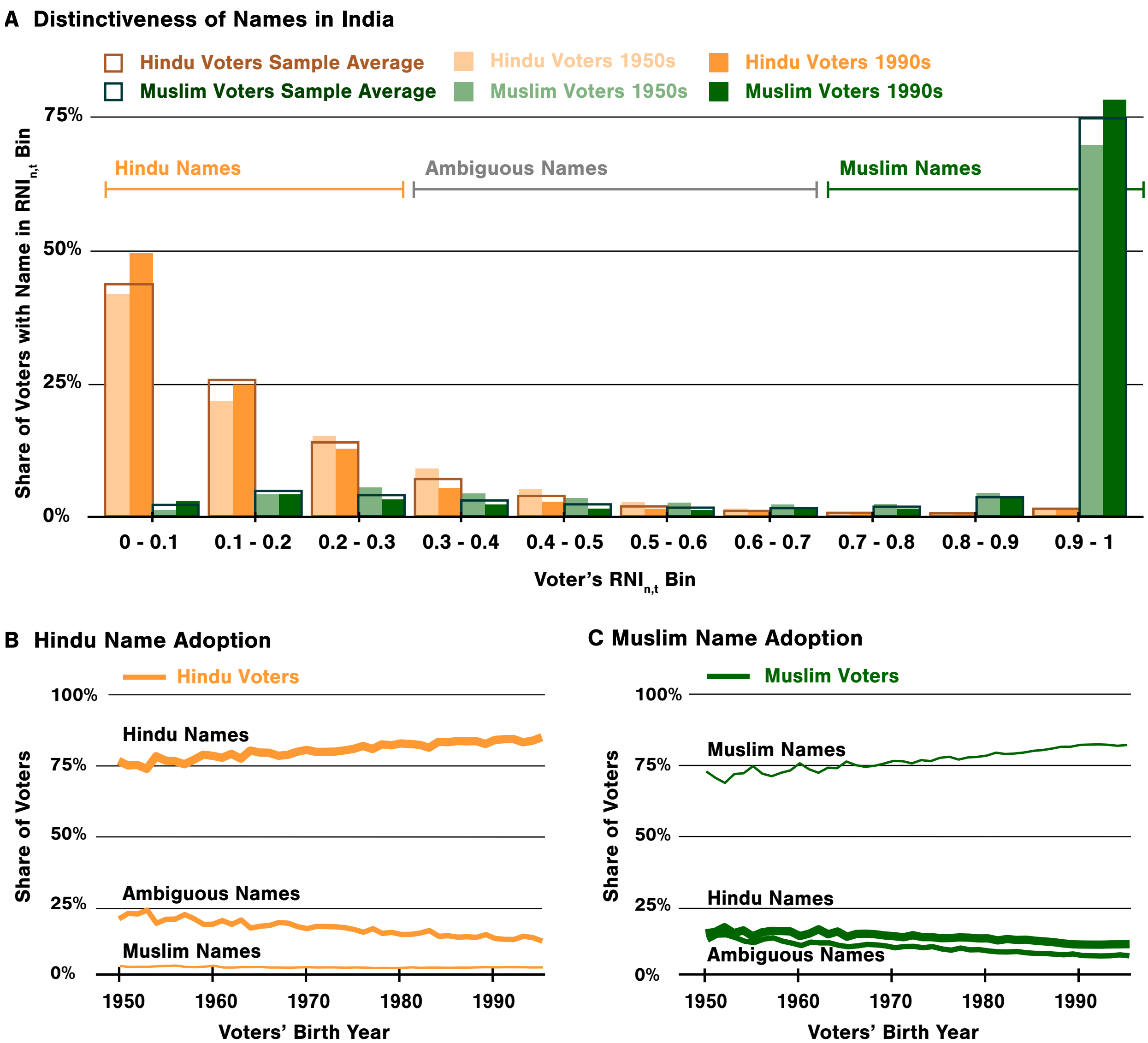


Figure 4: **Religious Distinctiveness of Names in India.** **(A)** The $RNI_{n,t}$ range is divided into 0.1-width bins. The plot depicts the share of Muslims and Hindus whose name falls within each $RNI_{n,t}$ bin for 1950 and 1990. Additionally, we report the averages share of voters in each bin for the full sample period as boxes. More weight towards the extremes indicates more distinctive names (N = 505,309,697). **(B)** The share of voters with a name in our name bins for Hindus (N = 425,093,659). **(C)** The share of voters with a name in our name bins for Muslims (N = 80,216,038).

4.2 Rising Religious Identity Expression

Both Hindus and Muslims increasingly adopt religiously distinctive names over our 1950-1995 study period. A formative phase in postcolonial India, this period, while predating the BJP's national rise, witnessed recurring Hindu-Muslim riots, the spread of Hindu nationalist ideas, and the early mobilization of right-wing movements. Consistent with the expectation that majority groups intensify identity-affirming behavior as cleavages sharpen, Hindus increasingly adopt distinctively Hindu names. However, contrary to the expectation that minorities should reduce visible markers under inter-group tension, we find that Muslims show no decline in distinctively Muslim names.

In Figure 4 A, we document a rise in voters who have extremely religiously distinctive names (meaning they are in the most extreme RNI bin). Such Hindu voters increased from 42% in 1950 to 50% in 1990, while Muslim voters with such names increased from 70% in 1950 to 79% in 1990. This similar rise in the share of voters with extremely distinctive names among both Hindus and Muslims occurs despite asymmetric scope for change across the two groups. Among Hindus, only about half are located in the most distinctive bins, leaving substantial room for further upward movement. Among Muslims, by contrast, voters are already heavily concentrated in these extreme categories, implying more scope for movement away from such high distinctiveness. Given this, while we would expect Hindus to show a sharper increase and Muslims to exhibit some decline, we find that neither temporal pattern appears in the data. Hindus do not disproportionately intensify, and Muslims do not move away from already high levels of distinctiveness.

Panels B and C of Figure 4 show the temporal trajectories separately for Hindu and Muslim names. The same three patterns hold for both groups: the adoption of distinctively in-group names rises, ambiguous names decline gradually, and out-group adoption (Hindus with Muslim names, Muslims with Hindu names) stays low and flat throughout. This parallel rise sharpens the descriptive puzzle. Rising distinctiveness for both groups is hard to reconcile with accounts that rest on minority moderation under pressure.

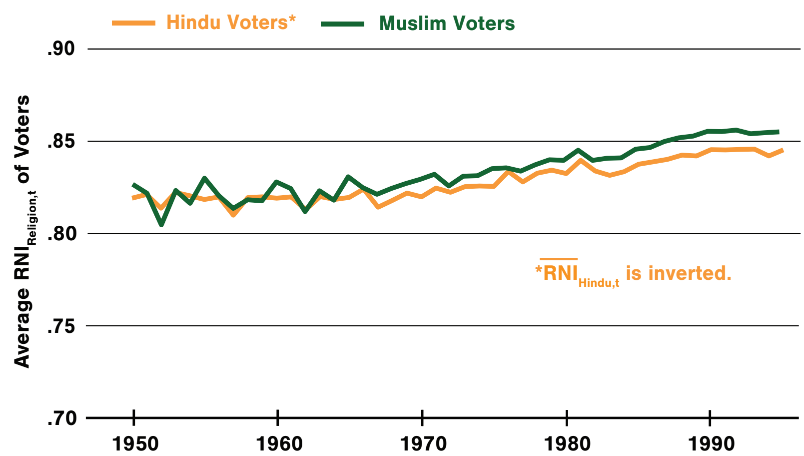


Figure 5: **Religious Identity Expression over Time** The development of $\overline{RNI}_{religion,t}$ over birth years is depicted. $\overline{RNI}_{Hindu,t}$ is inverted to facilitate comparison, i.e., $\overline{RNI}_{Hindu,t}$ of 0.2 becomes 0.8. The vertical axis ranges from 0.7 to 0.9, facilitating the assessment of trends (N = 505,309,697).

In Figure 5, we document a rise in average RNI from 0.80 to 0.85 over more than half a century, a substantively meaningful change that reflects large numbers of voters shifting particularly into more distinctive naming bins. Overall, these results show that over the past half-century, both Hindus and Muslims have increasingly chosen names that signal in-group identity, indicating that rising religious tensions are reflected not only in political behavior but also in everyday cultural choices such as naming.

4.3 Inter-generational Transmission of Identity Expression

To further characterize the rise in religious identity expression among Hindus and Muslims, we take advantage of a feature of our dataset that is rarely available to researchers: the ability to measure both parents' and children's identity expression at scale. Using this, we ask whether parents give their children more, less, or the same levels of identity distinction as themselves and if this differs across the majority and minority groups. Figure 6 compares the $RNI_{n,t}$ of children and their parents.[8]

We find a clear inter-generational increase in identity expression, but with important differences

[8]We exclude married females since only their husbands', rather than their fathers', names are recorded. Appendix C.2 shows that including married women does not affect our conclusions.

across groups. Among Hindus, children are systematically given more religiously distinctive names than their parents, especially at lower levels of parental RNI. As the majority group, Hindus face fewer constraints on expressing identity and may have incentives to increase it, particularly when baseline identity expression is low. At higher levels of parental RNI, Hindu parents tend to give children names with similar levels of distinctiveness, suggesting convergence at the upper end of the distribution.

Among Muslims, the pattern diverges from expectations of lowering identity transmission. At higher levels of parental RNI, Muslim parents give their children names with similar levels of distinctiveness, mirroring their own identity expression. However, at lower levels of parental RNI, Muslim parents, like Hindus, also give their children more distinctive names than their own, although the increase is relatively modest compared to Hindus. Yet, this increase is notable because, as a minority group, Muslims parents might be expected to reduce the expression of identity relative to their own levels given rising marginalization. Instead, we observe persistence at the top and an increase at the bottom of the RNI distribution. The intergenerational evidence reinforces our broader finding that, even within households, both majority and minority groups raise identity expression.

4.4 Robustness

We probe the robustness of our results in Appendix C along four dimensions.

First, we validate our approach using the REDS survey, a sample for which where ground-truth religion is available. We construct the RNI using both the true religion and the predicted religion from our classifier. We find a strong correlation of $r = 0.94$ between the two measures. Because the predicted and true religions agree at high rates, the RNI computed from predicted religion closely tracks the RNI that would have been obtained from the true religion. Any classification error does not meaningfully distort our measure. See Appendix C.1.

Second, we examine whether our findings are driven by males and, more broadly, whether there

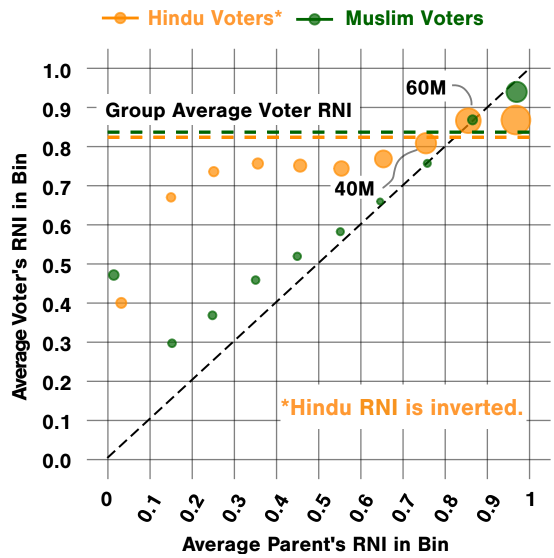


Figure 6: **Intergenerational Transmission of Religious Identity.** Stratifies voters by parent's $RNI_{n,t}$, y-axis shows the average voter $RNI_{n,t}$ in each parent $RNI_{n,t}$ bin, the x-axis shows the average parent $RNI_{n,t}$ in the bins. The sample is restricted to men and unmarried women, since our data does not include the parent name for married women (N = 260,283,237). Perfect intergenerational transmission would be indicated by points on the 45-degree line. Dot size reflects number of voters in each bin.

are gender-based differences we do not account for (see subsection C.2). We assess this in three ways. We begin by showing that naming conventions are similar for men and women, for instance, as measured by name length and the propensity to use two-part names (see Figure A5). We then replicate our main results separately for men and women and find that the substantive patterns hold for both groups (see Figure A6). Finally, we include unmarried women, for whom we can reliably observe parental information, in the intergenerational analysis and the results remain unchanged (see Figure A7).

Third, we assess whether our method of inferring religion from names could be driving the results (see subsection C.3). Misclassification could bias our estimates of identity expression. For example, incorrectly classifying a Hindu as Muslim and observing a low RNI could misleadingly suggest that Muslims are adopting less distinctive names. We address this concern in two ways. First, our out-of-sample accuracy is 95.6%, suggesting that our classification of religion is reliable for most of our sample. Second, the high accuracy is also stable across age cohorts, as it relies on linguistic patterns in names that rarely change over time (see Figure A8). This implies that the dynamic results that we present are not driven by changes in accuracy over time. Third, we replicate our analysis on a subsample of voters for whom the classification certainty is especially high and find that the results are unchanged (Figure A9).

Fourth, we examine whether our measure of religious classification is correlated with RNI (see subsection C.4). A concern is that parents with lower classification accuracy may systematically have children with lower RNI, potentially biasing our findings. To test this, we analyze the relationship between prediction likelihood (our measure of confidence in religious classification) and RNI, and find no correlation (see Figure A10). This is consistent with the fact that our religion classification relies on linguistic features of names, while RNI is constructed from frequency-based patterns, making the two measures conceptually distinct.

5 Explaining the Value of Identity Expression

So far, we have identified high levels of religious identity expression among Hindus and Muslims that increase over time for both groups. While the high levels are consistent with social identity theory, the heightened expression for Muslims over time is unexpected. In this section, we investigate two broad mechanisms that can make sense of these findings: divine benefits and social benefits and punishment.

5.1 Divine Benefits

Perhaps one reason for the deeper identity expression patterns we document above is that people are increasingly turning towards religious doctrine as a source of inspiration for naming their children. We now examine this logic by studying whether names are increasingly likely to have religious origins. We consider names to have religious doctrinal origins if they appear directly in religious scriptures—figures or deities and their epithets or relatives in Hindu sacred texts, and figures or attributes of God in Islamic scripture like the Quran, Hadith, or Seerah (Andersen and Bentzen, 2022). For example, we classify Lakshmi as a Hindu doctrinal name because it refers directly to a Hindu goddess with the same name. On the other hand, while Rajat is a Sanskrit word meaning silver, it does not refer to a God, mythological figure, epithet, or relative of God. Among Muslims, Imran refers to the father of Maryam in Islamic scripture. In contrast, Farhana is a female name of Arabic origin, but it does not refer to a figure in Islamic scripture and is not a name of Allah. To measure whether all names in our dataset have religious doctrinal roots, we use a conventional *Large Language Model* (claude-3-5-sonnet-20240620, Anthropic) to systematically classify names, as outlined in Figure 7, Panel A and Appendix A.3. Manually evaluating 500 randomly sampled names for each group yields that the LLM used factual information in 94.2% of the verification sample, and provided a correct verdict in 94.4% of cases.

Figure 7, Panel B, provides no evidence that names with religious origins are more commonly adopted over time. If anything, doctrinal naming starts from a relatively high baseline and then declines among both groups. In 1950, 43 percent of Hindus and 31 percent of Muslims had names derived from doctrinal sources, indicating that religion was a salient component of identity for substantial shares of both communities, and a high baseline level of religious identity expression is consistent with the logic of divine benefits. Yet Muslims were consistently less likely than Hindus to adopt doctrinal names, suggesting the weaker pull of divine benefits among the minority group. By 1990, these shares had fallen to 37 percent among Hindus and 25 percent among Muslims. The decline among Hindus and Muslims runs against the broader pattern of heightened identity expression and does not support an argument of increasing religious expression through doctrinal naming over time. This decline in the adoption of names rooted in religious doctrine suggests that growing religious identity expression by Hindus and Muslims in India is not supported by a logic of divine benefits.

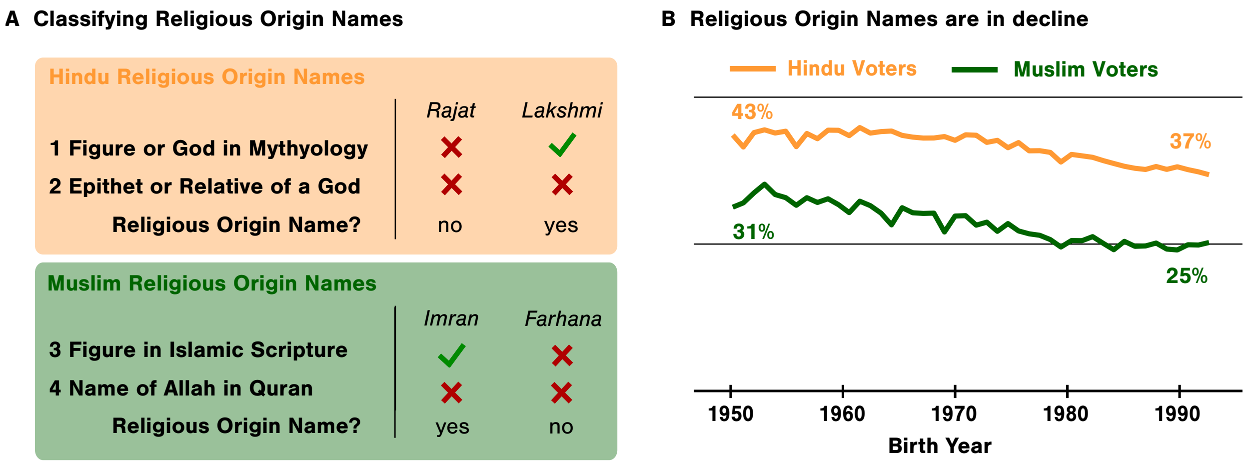


Fig. 3. Religious Origin of Names in India. (A) showcases how religious anmes are classified and presents relevant examples. **(B)** traces the share of voters who have religious origin names for a random sample of 1 million voters.

Figure 7: **Religious Origin of Names in India. (A)** Showcases how religious names are classified and presents relevant examples. **(B)** Traces the share of voters who have religious origin names for a random sample of voters (N = 1,000,000).

5.2 Social Benefits and Punishment

Beyond divine names, we also examine how social benefits and punishments are associated with naming decisions. As discussed above, there are two predictions we bring to the data: first, that the adoption of common names should increase, likely because of coordination and solidarity. And second, minority identity expression should be stronger in segregated settings, where concealment is ineffective and potentially costly and incentives for in-group signaling are greater.

Beginning with common names, we assess whether individuals are converging on a smaller set of widely used names that yield greater social value through recognition and coordination. Following previous approaches (see, for instance, Social Security Administration (2024); Bazzi et al. (2020)), we define a name as common if it ranks among the 1,000 most frequent names in the voter's Assembly Constituency, by birth year, and by religion. Figure 8, Panel A shows that over our sample period, common names become more popular for Hindus and Muslims. This means that parents choose names for their children from a smaller set of group-specific names. However, the rise in common names, which is a specific sub-sample of names, could be driven by those that

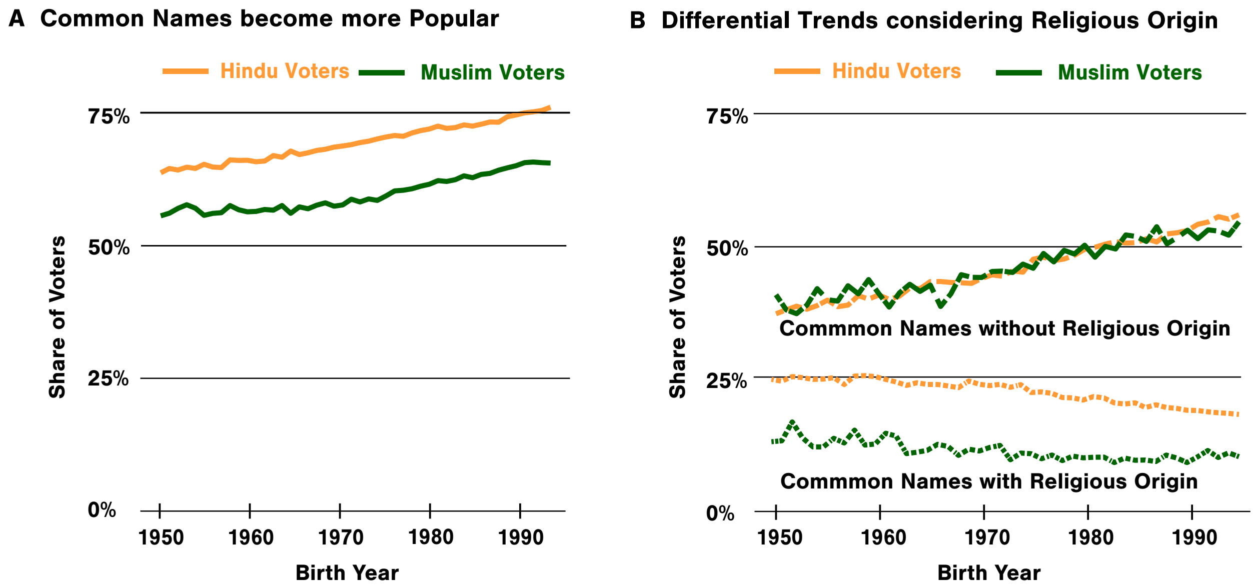


Figure 8: **Common Names in India. (A)** Common names are defined as the top-1000 names in a state at the time of the voter's birth. The plot shows the overall development of the share of voters with common names, and additionally stratifies this by the name's religious origin (N= 505,309,697). **(B)** Splits common names into those with religious origin and such without. In sum the shares of these names result in the overall share of common names (N= 505,309,697).

also have religious origins. Assessing common and religious origin names jointly, we find that the rise in use of common names is concentrated among those without a religious origin. In fact, names that are common and have a religious origin are declining in popularity.

An additional observable implication of the social logic is differential identity expression for majority and minority groups based on the spatial concentration of religious groups in their residential contexts. We implement a disaggregated *Dissimilarity Index* which measures segregation based on cross-sectional population data. Positive (negative) values indicate higher Muslim (Hindu) concentration in the neighborhood relative to the regional average. For instance, a neighborhood with 30% Muslims in a region with 10% Muslims will have a value of +20%. One important limitation of using cross-sectional data is that we implicitly assume naming decisions were made in the same locations where individuals are observed in 2017. While this would generally be a nontrivial assumption, it is less concerning in our context. Our sample spans births from 1950 to 1995, with the vast majority occurring prior to 1991, the onset of economic liberalization in India, which marks the beginning of more substantial inter-state and labor-driven migration (Tumbe, 2018). Before this period, mobility was more limited and predominantly local or marriage-related. Consistent with this, the key patterns we document are already clearly discernible in cohorts born before the early 1990s. Moreover, even among families that do migrate, a large share is likely to retain voter IDs in their place of origin, as these documents remain important for managing ancestral land and maintaining administrative ties to native communities.

Turning to the results in Figure 9, we see that both Hindus' and Muslims' names have higher $RNI_{n,t}$ in neighborhoods where their own group is more concentrated and lower $RNI_{n,t}$ in areas where their group's concentration is lower. This suggests that naming patterns are correlated with the social context. However, one important difference emerges. Muslim $RNI_{n,t}$ declines sharply as one moves from neighborhoods that are Muslim concentrated to Hindu concentrated. In contrast, moving from Hindu to Muslim concentrated neighborhoods is correlated with a reduction in Hindu $RNI_{n,t}$ relatively gradually. Examining these patterns over time, we show in Figure A3 in Appendix B.1 that identity expression for Muslims is increasing more strongly in Hindu concentrated neighborhoods over time, while for Hindus, the temporal increase in distinctiveness is not correlated with spatial segregation.

The findings emphasize how the social cost-benefit calculus is related to religious identity expression. For both majority and minority groups, the increasing adoption of common names that are distinctive for their group can help make in-group coordination easier and boost solidarity among them. While minorities may have incentives to lower their identity expression, segregated contexts can reverse this expectation. When minorities lived in segregated communities, lowering their identity expression risks being viewed as disloyal by their ingroup members. Moreover, segregated contexts make it easier to infer identity based on location, which risks punishment from majority group members if a minority seeks an ambiguous or majority group name. Both of these processes are consistent with the patterns we observe, where Hindu and Muslim naming patterns vary by residential contexts, with Muslim variation across contexts being more pronounced.

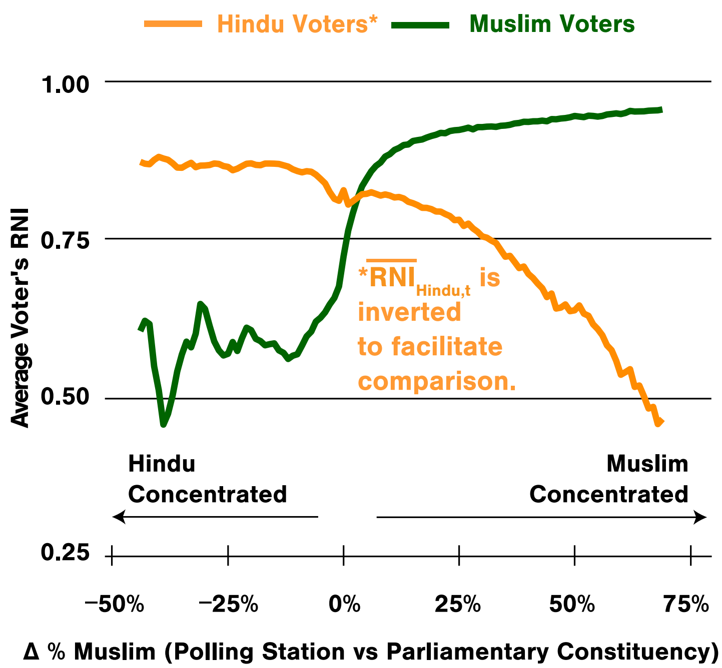


Figure 9: **Segregation and Expression of Religious Identity.** The x-axis depicts standard measure of segregation comparing the share of Muslims in a Parliamentary Constituency (PC) to the one on the local polling station, 0 indicates non-concentrated Polling Station, higher absolute values mean more concentrated. The y-axis depicts average $RNI_{n,t}$ in Polling Stations based on their level of segregation (N= 505,309,697).

6 Conclusion

Our paper documents rising identity expression among the majority *and* the minority group, even before the electoral rise of majoritarianism. With a focus on 505 million Indian voters along one of the country's defining identity cleavages, we document rising identity expression. Hindu voters increasingly adopt more distinctively Hindu names, while Muslim voters also move toward more distinctively Muslim names, contrary to expectations for minority behavior. While the high baseline prevalence of religiously distinctive names is consistent with the divine value attached to such names, the temporal patterns are better captured by a social calculus.

These findings challenge a standard expectation in social identity theory that marginalized minorities will conceal or moderate visible identity markers. Instead, we show that in polarized and residentially segregated settings, where minorities are likely to be readily identified, moderation may cease to function as a viable strategy. In studying religious groups, we broaden the empirical scope of research on minority identity beyond the familiar native–immigrant and racial divides of the United States and Europe. In India's Hindu–Muslim divide, minorities are not newcomers negotiating assimilation into a host society, but centuries-old citizens confronting rising majoritarian pressures within their own democracy, where identity expression is shaped less by the pursuit of assimilation than by the demands of everyday survival. Our paper also contributes to the study of religion and politics in India by documenting a long-run intensification of religious identity expression during the rise of right-wing majoritarian politics (Varshney, 2003; Wilkinson, 2004; Allie, 2025). In the decades preceding the Bharatiya Janata Party's national dominance, visible religious distinctiveness increased for both Hindus and Muslims, suggesting that the cultural foundations of majoritarian consolidation were already taking shape before its full electoral consolidation became apparent.

More broadly, our findings underscore how culturally and socially rooted acts of identity expression can both reflect and reveal larger political transformations. While our findings document one important dimension of minority identity expression at scale, understanding the causal forces that generate these patterns requires future research. This may involve studying how both focal events, such as riots, trade shocks, policy reforms, and social movements, and the cumulative pressures of everyday interactions in communities, rental markets, and labor markets shape these dynamics.

References

Abdelagadir, A. and V. Fouka (2020). Political secularism and muslim integration in the west: Assessing the effects of the french headscarf ban. *American Political Science Review 114*(3), 707–723.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2020). Do immigrants assimilate more slowly today than in the past? *American Economic Review: Insights 2*(1), 125–141.

Adida, C. L., D. D. Laitin, and M.-A. Valfort (2016, 1). *Why Muslim Integration Fails in Christian-Heritage Societies*. Harvard University Press.

Adida, C. L. and A. L. Robinson (2023). Why (some) immigrants resist assimilation: Us racism and the african immigrant experience. *Quarterly journal of political science 18*(3), 295.

Adukia, A., S. Asher, P. Novosad, and B. Tan (2019). Residential Segregation in Urban India.

Ahmad, A. (2020). When the name matters: An experimental investigation of ethnic discrimination in the finnish labor market. *Sociological Inquiry 90*(3), 468–496.

Ahmad, R. (2011, 6). Urdu in Devanagari: Shifting orthographic practices and Muslim identity in Delhi. *Language in Society 40*(3), 259–284.

Ahmed, H. (2019). *Siyasi Muslims: A story of political Islams in India*. Penguin Random House India Private Limited.

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

Allie, F. (2023, 12). Sacred Time and Religious Violence: Evidence from Hindu-Muslim Riots in India. *Journal of Conflict Resolution*.

Allie, F. (2025). India's new minority politics. *Journal of Democracy 36*(2), 78–91.

Allie, F. (2026). The Representation Trap: How Minority Power Shapes Electoral Politics in India.

Andersen, L. H. and J. Bentzen (2022). In the Name of God! Religiosity and the Transition to Modern Growth. *CEPR Discussion Paper No. DP16938*.

Asher, S., P. Novosad, and C. Rafkin (2024, 4). Intergenerational Mobility in India: New Measures and Estimates across Time and Social Groups. *American Economic Journal: Applied Economics 16*(2), 66–98.

Assouad, L. (2020). Charismatic leaders and nation building.

Atkin, D., E. Colson-Sihra, and M. Shayo (2021a). How do we choose our identity? a revealed preference approach using food consumption. *Journal of Political Economy 129*(4), 1193–1251.

Atkin, D., E. Colson-Sihra, and M. Shayo (2021b, 4). How Do We Choose Our Identity? A Revealed Preference Approach Using Food Consumption. *Journal of Political Economy 129*(4), 1193–1251.

Baral, S., G. Nellis, and M. Weaver (2023). The Electoral Consequences of Mass Religious Events: India's Kumbh Mela. *Center for Effective Global Action*.

Bazzi, S., M. Fiszbein, and M. Gebresilasse (2020). Frontier Culture: The Roots and Persistence of "Rugged Individualism" in the United States. *Econometrica 88*(6), 2329–2368.

Bertrand, M. and S. Mullainathan (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review 94*(4), 991–1013.

Bursztyn, L., M. Callen, B. Ferman, S. Gulzar, A. Hasanain, and N. Yuchtman (2020). Political identity: Experimental evidence on anti-americanism in pakistan. *Journal of the European Economic Association 18*(5), 2532–2560.

Carvalho, J.-P. (2013). Veiling. *The Quarterly Journal of Economics 128*(1), 337–370.

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

Chaturvedi, R. and S. Chaturvedi (2022). It's All in the Name: A Character Based Approach to Infer Religion. *SSRN Electronic Journal*.

Choi, D. D., N. Sambanis, and M. Poertner (2022). Native bias: Overcoming discrimination against immigrants.

Das, S. (2025, September). Electoral irregularities in the world's largest democracy. Available at SSRN: `https://ssrn.com/abstract=4512936`.

Datta, S. and V. Pathania (2016). *For whom does the phone (not) ring? Discrimination in the rental housing market in Delhi, India*, Volume 2016. UNU-WIDER.

de Kadt, D. and A. Grzymala-Busse (2025). Good description.

Dube, O., J. Blumenstock, and M. Callen (2022, 11). Measuring Religion from Behavior: Climate Shocks and Religious Adherence in Afghanistan. Technical report, National Bureau of Economic Research, Cambridge, MA.

Election Commission of India (2011). *Handbook for Booth Level Officers*.

Evans, J. and N. Sahgal (2021). Key findings about religion in India. Technical report, Pew Research Centre.

Farooqui, A. (2020, 3). Political representation of a minority: Muslim representation in contemporary India. *India Review 19*(2), 153–175.

Fearon, J. D. and D. D. Laitin (1996). Explaining interethnic cooperation. *American political science review 90*(4), 715–735.

Fouka, V. (2019, 5). How Do Immigrants Respond to Discrimination? The Case of Germans in the US During World War I. *American Political Science Review 113*(2), 405–422.

Fouka, V. (2020, 1). Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I. *The Review of Economic Studies 87*(1), 204–239.

Fryer, R. G. and S. D. Levitt (2004, 8). The Causes and Consequences of Distinctively Black Names. *The Quarterly Journal of Economics 119*(3), 767–805.

Ge, Q. and S. Wu (2024). How do you say your name? difficult-to-pronounce names and labor market outcomes. *American Economic Journal: Economic Policy 16*(4), 254–279.

Gill, A. (2001). Religion and comparative politics. *Annual Review of Political Science 4*(1), 117–138.

Goyal, T. (2021). *Representation from Below: How Women Mobilize Inside Parties*. Phd thesis, University of Oxford.

Grzymala-Busse, A. (2012). Why comparative politics should take religion (more) seriously. *Annual Review of Political Science 15*(1), 421–442.

Haas, N. and E. Lindstam (2024, 11). My History or Our History? Historical Revisionism and Entitlement to Lead. *American Political Science Review 118*(4), 1778–1802.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007, 11). Why Does Ethnic Diversity Undermine Public Goods Provision? *American Political Science Review 101*(4), 709–725.

Hackett, C., M. Stonawski, Y. Tong, S. Kramer, A. F. Shi, and D. Fahmy (2025). How the global religious landscape changed from 2010 to 2020. pew research center. *Published (June 9, 2025) Accessed July 28*.

Iyer, S. and A. Shrivastava (2018, 3). Religious riots and electoral politics in India. *Journal of Development Economics 131*, 104–122.

Jaffrelot, C. and H. Ahmed (2024, 11). Indian Muslims: (Self-)Perceptions and Voting Trends in 2024. *Studies in Indian Politics*.

Jain, V., T. Enamorado, and C. Rudin (2022). The importance of being Ernest, Ekundayo, or Eswari: an interpretable machine learning approach to name-based ethnicity classification. *Harvard Data Science Review* (4.3).

Jayaraman, R. and M. Oak (2005, 11). The Signalling Role of Municipal Currencies in Local Development. *Economica 72*(288), 597–613.

Kuipers, N., T. B. Pepinsky, and J. C. Kuipers (2025). Observing the transformation of religious identity through indonesian naming practices.

Laitin, D. D. (1998). *Identity in formation: The Russian-speaking populations in the near abroad*. Cornell University Press.

Lowes, S., B. Marx, and E. Montero (2025). Religion in emerging and developing regions. Technical report, National Bureau of Economic Research.

Majumdar, R. (2023). Reducing Prejudice and Support for Religious Nationalism Through Conversations on WhatsApp. *SSRN Electronic Journal*.

McCauley, J. F. (2014). The political mobilization of ethnic and religious identities in africa. *American Political Science Review 108*(4), 801–816.

McClendon, G. H. and R. B. Riedl (2019). *From pews to politics: Religious sermons and political participation in Africa*. Cambridge University Press.

Mehrotra, R. R. (1982, 3). Impact of Religion on Hindi Personal Names. *Names 30*(1), 43–47.

National Council of Applied Economic Research (2006). Rural Economic & Demographic Survey (REDS).

Nellis, G., A. Gosh, P. Kundu, and M. Lowe (2024). Creating Cohesive Communities: A Youth Camp Experiment in India. *Review of Economic Studies [Conditionally Accepted]*.

Office of the Registrar General and Census Commissioner, India (2011). Census of india 2011: National population register and socio-economic and caste census. Technical report, Office of the Registrar General and Census Commissioner, India (ORGI). Accessed 2025-10-13.

Pew, R. (2018). The age gap in religion around the world. *Epub ahead of print 13*.

Pew Research Center (2021, June). Religion in india: Tolerance and segregation.

Philpott, D. (2007). Explaining the political ambivalence of religion. *American Political Science Review 101*(3), 505–525.

Portes, A. and R. G. Rumbaut (2001). *Legacies: The story of the immigrant second generation*. Univ of California Press.

Posner, D. N. (2005). *Institutions and Ethnic Politics in Africa*. New York: Cambridge University Press.

Rahman, T. (2013). Personal Names and the Islamic Identity in Pakistan. *Islamic Studies 52*, 239–296.

Rizwan, A., V. Kulikov, and N. Iqbal (2023, 9). Muslim personal names in Urdu: structure, meaning, and change. *International Journal of the Sociology of Language 2023*(283), 161–188.

Robinson, A. L. (2023). Ethnic visibility. *American Journal of Political Science 68*(4), 1234–1251.

Rubinstein, Y. and D. Brenner (2014). Pride and prejudice: Using ethnic-sounding names and inter-ethnic marriages to identify labour market discrimination. *Review of Economic Studies 81*(1), 389–425.

Sahgal, N., J. Evans, A. M. Salazar, K. J. Starr, and M. Corichi (2021, June 29). Religion in india: Tolerance and segregation. `https://www.pewresearch.org/religion/2021/06/29/religious-segregation/`. Pew Research Center, "Religious segregation in India".

Shadab, R., V. Sharan, and D. Lahiri (2022). Hiring Bias Employment for Muslim women at entry-level roles. Technical report, LEDBY Foundation.

Sharma, M. L. (1969, 9). Origin and Meaning of Some Indian Names. *Names 17*(3), 208–213.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review 103*(2), 147–174.

Social Security Administration (2024). Change in Name Popularity.

Sood, G. and A. Dhingra (2018). Indian Electoral Rolls PDF Corpus.

Tajfel, H. and J. C. Turner (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader 56*(65), 9780203505984–16.

Thorat, Sukhadeo; Banerjee, A. M. V. K. R. F. (2015). Urban Rental Housing Market: Caste and Religion Matters in Access. *Economic and Political Weekly*, 47–53.

Tumbe, C. (2018). *India moving: A history of migration*. Penguin Random House India Private Limited.

Tuñón, G. (2019). *When the Church Votes Left: The Electoral Consequences of Progressive Religion*. Ph. D. thesis, University of California, Berkeley.

van Noort, S. and T. Goyal (2025). Is ethnic violence self-perpetuating? quasi-experimental evidence from hindu-muslim riots in india. *Journal of Conflict Resolution 0*(0), 00220027251383563.

Varshney, A. (1998, 6). Why Democracy Survives. *Journal of Democracy 9*(3), 36–50.

Varshney, A. (2003). *Ethnic conflict and civic life: Hindus and Muslims in India*. Yale University Press.

Wald, K. D., A. L. Silverman, and K. S. Fridy (2005). Making sense of religion in political life. *Annu. Rev. Polit. Sci. 8*(1), 121–143.

Wilkinson, S. I. (2004, 7). *Votes and Violence*. Cambridge University Press.

ONLINE APPENDIX

A Data and Methodological Approach

A.1 Data Sources

Our analysis is based on Indian voter lists which include sensitive private information on the names and addresses of hundreds of millions of citizens. Therefore, we will not include individual level data in the public replication archive. Instead, the replication archive will contain aggregated data level code. Scholars can obtain the full datasets from Sood and Dhingra (Sood and Dhingra, 2018) at `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVNOG47IV`, Goyal (Goyal, 2021) at `https://ora.ox.ac.uk/objects/uuid:6bb8a8f6-7e8a-45c8-a4d8-073331a7fac5` and the data we use to train the religion inference algorithm from REDS (National Council of Applied Economic Research, 2006) at `https://www.ncaer.org/project/additional-rural-incomes-survey-aris-rural-economic-demographic-survey-reds/`.

A.2 Data Coverage

Figure A1 maps the coverage of our sample across India. States shaded in grey are entirely absent from the data, of which Chhattisgarh is the only case. States shaded in light blue contain at least some assembly constituencies in our sample but are missing others; this is most pronounced in Gujarat, Karnataka, and Jammu & Kashmir, each of which has substantial gaps in AC coverage. States shaded in dark blue are fully or near-fully represented in the sample.

While our results are internally consistent on the population of registered voters, there might be concerns that they might weaken for the full population. To investigate this possibility, we compare the number of observations in our data with the 2011 Census (Office of the Registrar General and Census Commissioner, India, 2011). Before doing so, we remove Chhattisgarh, Ladakh, and Lakshadweep, since we are lacking these states in our voter data. We find that our sample achieves to cover around 80% of the Indian population in the remaining states.

Table A1 depicts our sample coverage of the Indian population. Our sample draws on digitized electoral rolls from the 2017 state assembly elections and covers 3,674 of the 4,096 assembly constituencies (ACs) across India, representing 90% of all ACs and approximately 79% of the 2011 Census population. When benchmarking against the Census, we exclude the 14–18 age group: while

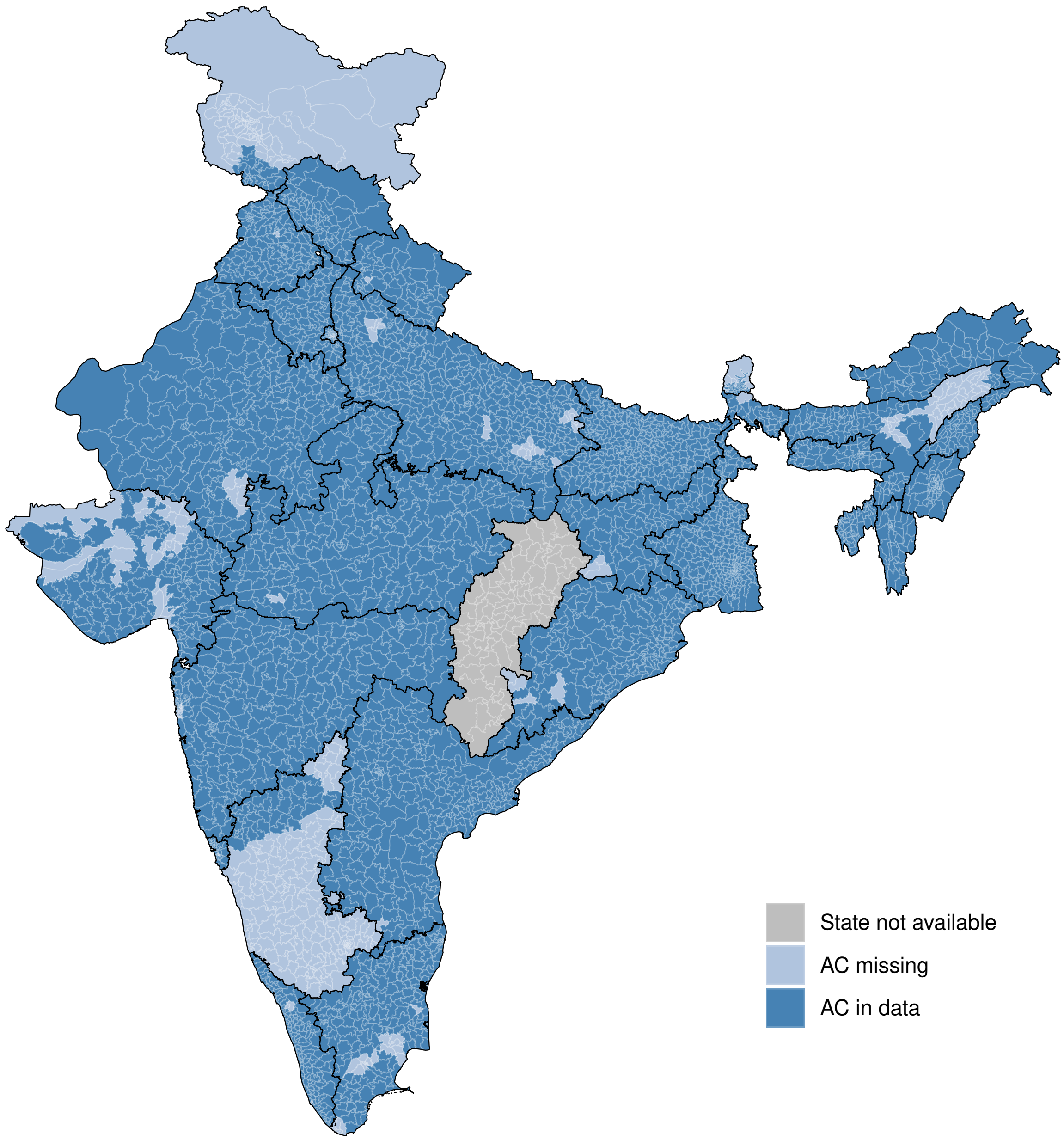


Figure A1: **Map of Data Coverage** This map shows our AC-level data coverage.

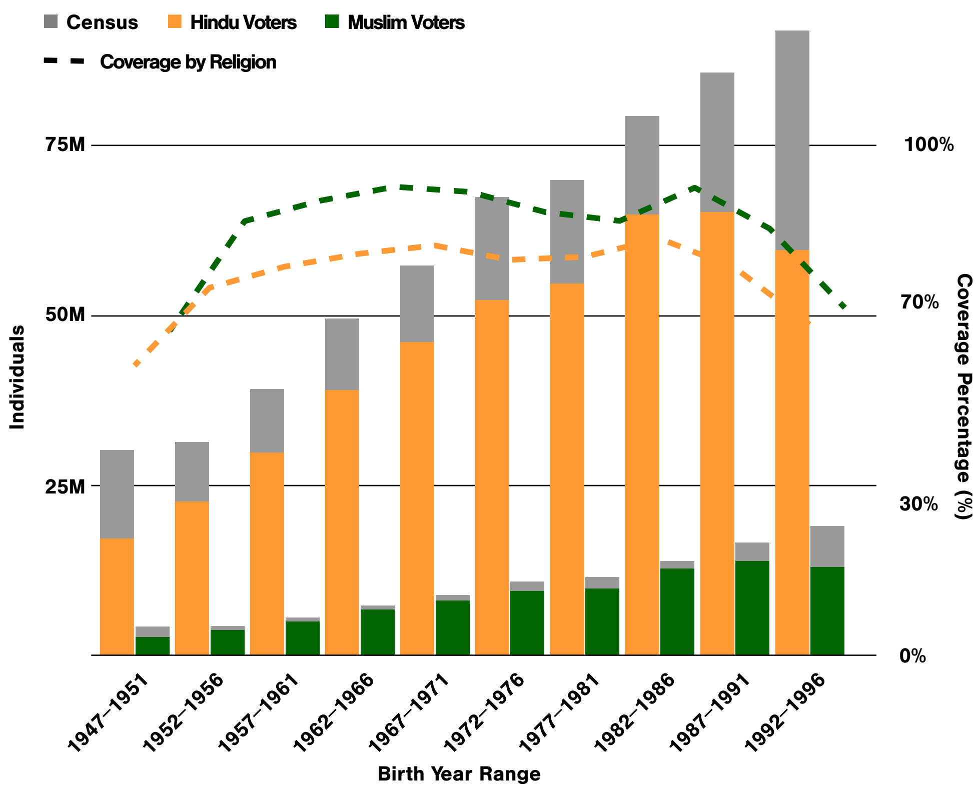


Figure A2: **Comparing Voter Roll Data and 2011 Census.** Plots the number of individuals in our data against the 2011 census number of individual in standardized birth year ranges. This comparison is restricted to our sample period from birth years 1950 to 1995. Chhattisgarh, Ladakh and Lakshadweep were removed from the census data since they are not included in our voter data (N= 505,309,697).

State	ACs in Sample	ACs in State	Population in Sample	Population in Census (2011)	AC Coverage	Population Coverage
Total	***3674***	***4096***	***662,087,046***	***838,410,861***	***90%***	***79%***
Uttar Pradesh	383	403	121,604,865	128,504,075	95%	95%
Maharashtra	288	302	83,307,658	82,457,118	95%	101%
West Bengal	293	294	58,943,602	66,538,640	100%	89%
Andhra Pradesh	293	296	26,703,576	62,789,985	99%	43%
Bihar	243	243	69,139,633	62,378,264	100%	111%
Tamil Nadu	211	234	29,681,354	55,139,527	90%	54%
Madhya Pradesh	225	226	47,374,852	48,324,567	100%	98%
Karnataka	39	225	7,745,093	45,070,423	17%	17%
Rajasthan	197	201	42,645,166	44,823,011	98%	95%
Gujarat	140	172	23,397,947	42,994,079	81%	54%
Odisha	143	147	33,876,166	29,897,796	97%	113%
Kerala	138	141	21,541,828	25,575,087	98%	84%
Jharkhand	92	95	20,246,000	21,097,016	97%	96%
Assam	92	133	7,059,275	20,956,677	69%	34%
Punjab	116	117	18,366,664	20,658,388	99%	89%
Haryana	90	90	16,443,061	17,821,508	100%	92%
Chhattisgarh	NA	NA	NA	17,361,362	NA	NA
Delhi	68	70	12,908,399	12,222,622	97%	106%
Jammu & Kashmir	32	107	2,299,105	8,300,592	30%	28%
Uttarakhand	69	70	6,883,138	6,957,284	99%	99%
Himachal Pradesh	68	68	NA	5,089,217	100%	NA
Tripura	60	60	2,411,425	2,655,926	100%	91%
Manipur	68	68	1,906,083	1,994,106	100%	96%
Meghalaya	59	60	1,735,360	1,788,947	98%	97%
Nagaland	60	60	1,143,136	1,299,470	100%	88%
Goa	39	41	1,077,783	1,140,385	95%	95%
Puducherry	30	30	935,319	949,561	100%	99%
Arunachal Pradesh	61	61	722,790	890,366	100%	81%
Chandigarh	1	1	579,878	788,938	100%	74%
Mizoram	40	40	740,277	741,204	100%	100%
Sikkim	34	38	335,363	444,640	89%	75%
Andaman & Nicobar	1	1	215,791	287,906	100%	75%
Dadra & Nagar Haveli	NA	NA	NA	235,896	NA	NA
Daman & Diu	1	2	116,459	188,262	50%	62%
Lakshadweep	NA	NA	NA	48,016	NA	NA

Table A1: **AC and Population Coverage.** The population and AC coverage figures are derived by comparing data from the 2017 state assembly elections with the 2011 Census, as this is the closest available census to the election year. Consequently, population coverage values exceeding 100% reflect genuine population growth between 2011 and 2017 rather than measurement error. AC coverage is calculated by comparing the number of assembly constituencies present in our sample against the total number of constituencies in each state.

our data records individual birth years, the Census reports this cohort as a single bracket, making a reliable age-by-age disaggregation infeasible. Population coverage figures exceeding 100% in some states reflect real population growth between 2011 and 2017 rather than measurement error. Missingness in our sample arises from three distinct sources. First, Chhattisgarh is absent entirely from our data. Second, several states are missing a subset of ACs, for instance, Karnataka has only 39 of its 225 constituencies in the sample. Third, even within included ACs, data can be incomplete: in Gujarat, for example, only the first page of each voter roll was digitized, which accounts for the state's low population coverage of 54% despite 81% AC coverage. Together, these sources of missingness should be borne in mind when interpreting our results.

A.3 Classifying Names' Religious Origin

Names with clear religious origins, such as Lakshmi, Mohammad, Ram, and Fatima, share one common feature: they appear directly in religious scriptures. This criterion is central to our definition which derives from understandings of religious names in existing literature (Rizwan et al., 2023; Rahman, 2013; Sharma, 1969; Mehrotra, 1982). In our version, we consider names to have a religious origin if they are the names of gods or figures in religious scriptures, such as the Vedas or the Quran, or if they are epithets of relatives of such figures, or if the name is one of the 99 names of Allah. However, to our knowledge, extensive dictionaries of the universe of names in Islamic or Hindu religious scriptures do not exist.

Model Configuration

We implement this classification through a Large Language Model (LLM) based approach on a carefully defined set of criteria for what constitutes a religious name. We use claude-3-5-sonnet-20240620 through Anthropic's API including message-batches-2024-09-24 for batching our requests and prompt-caching-2024-07-31 for caching the prompt. We implement caching and batching to control cost during the coding. Moreover, it allowed us to receive the classification results quickly.

Classification Approach and Prompt

The prompt asks the LLM to classify each name into eight categories, four for Islam and four for Hinduism. On the Islamic side, the first category captures names of figures directly mentioned in the main scriptures Quran, Hadith, or Seerah. This includes prophets such as Adam, Musa, and

Muhammad, companions of the Prophet, and other key individuals in early Islamic history. The second category captures concepts found in the Quran, meaning abstract ideas, moral principles, or religious notions such as Iman (faith) or Sakina (tranquility). The third category captures names of Allah or names containing one of the 99 names of Allah, such as Rahman (The Most Merciful) or Malik (The King). On the Hindu side, the categories follow an analogous structure. The first captures figures or gods appearing in main scriptures, the Vedic, Epic, and Puranic texts, such as Rama, Arjuna, or Sita. The second captures concepts from these scriptures, such as Dharma (righteousness) or Moksha (liberation). The third captures epithets or relatives of gods, such as Govinda for Krishna or Parameshwara for Shiva.

For both religions, a final category records whether part of a name falls into any of the religious categories. This is necessary because many Indian names are compounds. A name like Ramaya, for instance, is not itself a figure in Hindu scripture, but it contains Rama, who is. The partial category helps the LLM to distinguish such cases from names that are direct references to religious figures.

We found during testing that the distinction between figures and concepts was important for classification accuracy. The LLM sometimes incorrectly coded names that carry vague religious meanings, for instance, words that appear in scriptures but whose meaning is general enough that they are not necessarily religious. This problem is particularly relevant in our setting because many names used in India derive from Sanskrit and Arabic words that may not carry deep religious connotations but could be misclassified due to their linguistic association with sacred texts. By including a separate concept category, we found that the LLM became better at distinguishing names that clearly reference scriptural figures from those with only an indirect religious association through positive connotations.

Beyond these categories, the prompt instructs the LLM to follow five rules. These rules were developed iteratively based on cases where human coders and the LLM showed disagreement and where manual testing revealed to be ambiguous. Rule A states that feminine versions of a name receive the same classification as the masculine version, and vice versa. The reasoning is that a parent naming their daughter Imrana likely intends the allusion to the Quranic figure Imran, just as a parent naming their daughter Krishnaa likely intends a reference to Krishna. Rule B extends this logic to morphological variants more broadly. If a name in its basic form would be classified as religious, then all versions of that name that are valid morphologies under Sanskrit or Arabic grammar receive the same classification. For example, Safiyah is classified as a Quranic concept because it is a valid morphological form of Safi, and Lakshmika is classified like Lakshmi because it is a valid Sanskrit derivation. We do not claim expertise in the exact etymology of names or the grammatical rules governing their formation in both languages. Rather, we instruct the LLM to apply its knowledge of Sanskrit and Arabic grammar, on the assumption that parents giving their children a grammatically recognizable variant of a name likely intend the religious allusion. Rule C provides a definition of what makes a word a religious concept. A word qualifies if it is mentioned in a sacred text with a religious or spiritual meaning, if it refers to a theological or philosophical idea, if it is tied to religious practice or ritual, or if it describes divine attributes or metaphysical principles. Rule D governs the partial name categories. If a name is coded as religious in one of the main categories, it is not automatically coded the same way when it only forms part of a larger name. Instead, it is assigned to the corresponding partial category. Rule E provides an exception for names containing a name of Allah with the prefix Abd (meaning servant of). Because Names of Allah cannot traditionally be used as standalone personal names, compounds like Abdulrahman are coded directly in the Name of Allah category rather than only in the partial category.

Finally, the prompt includes a set of worked examples. In our testing, we found that including examples substantially improved the accuracy of the LLM's classifications relative to a manually verified benchmark. These examples were not chosen randomly but selected to highlight difficult or very common cases where the LLM was otherwise prone to error.

A.3.1 LLM Verification

The main challenge for our identification of doctrinal names are sampling and validating the LLM's assessments. To make this analysis feasible, we randomly sampled 1 million voters, and extracted their around 200 thousand names. While this sample is small compared to our main sample, we think that it's size is sufficient to draw correlational inference.

While it may seem feasible to determine the religious origins of particular names through individual search and evaluation, a dictionary-based approach to conduct this task at scale — and for validating large language models (LLMs) — presents significant challenges. There is no single authoritative text that comprehensively captures all religious or religious-adjacent sources (such as folklore), nor is there clarity on which texts should be included. Given the diversity and multiplicity of religious traditions, it is impractical, if not impossible, to compile an exhaustive dictionary of religious names.

We considered online databases as potential sources for more extensive lists of names and their religious affiliations. Platforms like quranicnames.com and indiaparenting.com offer detailed lists that include meanings and origins, often categorizing names by religious, linguistic, or cultural criteria. However, the methodologies these sites use for classification are typically opaque, making it difficult to ascertain which aspects of religion are captured or overlooked. Additionally, many entries on these websites appear to be generated by AI, raising further concerns about the reliability and accuracy of their classifications. These limitations underscore the inherent difficulties in relying on dictionary-based approaches or online resources to definitively determine the religious origins of names.

While a dictionary-based approach to determining the religious origins of names faces challenges in scalability and comprehensiveness, an LLM-based approach offers significant advantages for handling this task at scale. Unlike static dictionaries, LLMs can process vast datasets and dynamically classify a large sample of names based on an evidence-based prompt. Although LLMs offer scalability and flexibility in classifying the religious origins of names, they come with some limitations. LLMs are prone to "hallucinations," where they generate plausible but incorrect information, which can lead to misclassification of names. Additionally, the accuracy of LLM outputs depends on the quality of the prompt and examples. These challenges underscore the need for careful validation and human oversight when using LLMs for such tasks.

We do this by implementing a human coder to evaluate the LLM's performance. Two human coders evaluated the around 1000 randomly sampled names to guard against above pitfalls of using LLMs. This verification yields the conclusion that the LLM used factual information in 94.2% of the verification sample, and provided a correct verdict in at least 94.4% of cases.

B Additional Results

B.1 Name Distinctiveness by Segregation over Time

Figure A3 disaggregates the relationship between spatial concentration and religious identity expression by decade. For Muslim voters, the gradient across segregation levels steepens over time, as later cohorts show markedly higher $RNI_{n,t}$ in Hindu-concentrated neighborhoods, while differences across decades are smaller in Muslim-concentrated areas. The rise in Muslim identity expression is thus not uniform but concentrated in areas where Muslims are a local minority. For Hindu voters, by contrast, the curves remain tightly bundled across decades. The temporal increase in Hindu name distinctiveness does not vary meaningfully with local demographic composition. These patterns reinforce the idea that minority naming decisions are more responsive to social context, consistent with the stronger pressures Muslims face to demonstrate in-group loyalty and reduce out-group sanctioning.

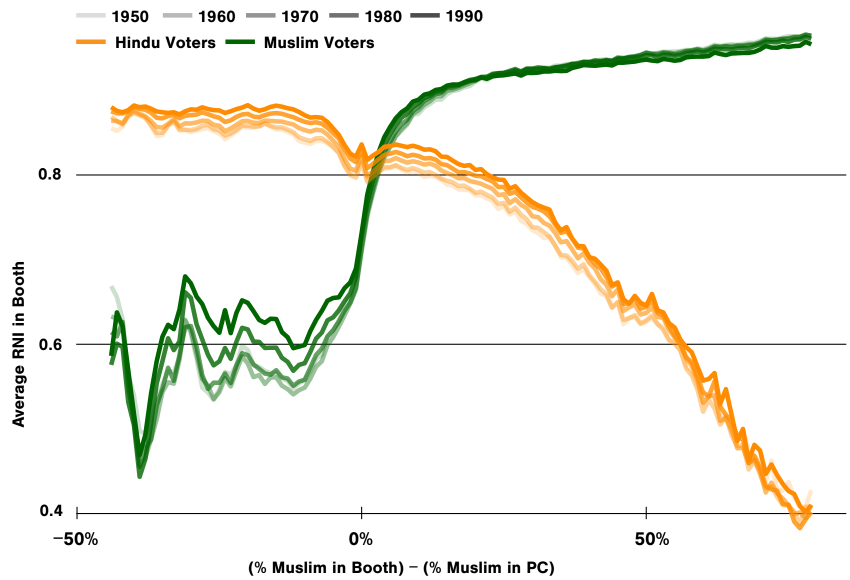


Figure A3: **Segregation and Expression of Religious Identity over Time.** Plots the average $\overline{RNI}_{religion,t}$ by Muslim concentration for the decades in our sample (N= 505,309,697).

C Robustness Analysis

C.1 Does Classification Error Drive Our Results?

A potential concern is about our results being driven by religion misclassification errors that get propagated into the RNI measure. To check this possibility, we turn to the REDS survey, which records each respondent's true religion. We use the holdout sample, the random sub-sample of REDS not used to train the religion classifier, and apply the same transformations as in our main dataset. Then we follow the usual steps: First, we predict each person's religion using their parents' name. Second, we extract the personal name from the full name. Finally, we compute the RNI for all names occurring more than five times in the sub-sample in two ways: once using the true (self-reported) religion and once using the religion predicted from parents' name. The two resulting measures agree strongly with a correlation of $r = 0.94$.

We also use the full REDS sample to examine the distribution of RNI. Figure A4 shows that the distribution based on true religion closely matches the one from our main results in Figure 4. This consistency suggests that our results are not artifacts of misclassification error or confounding between religion prediction and the RNI measure.

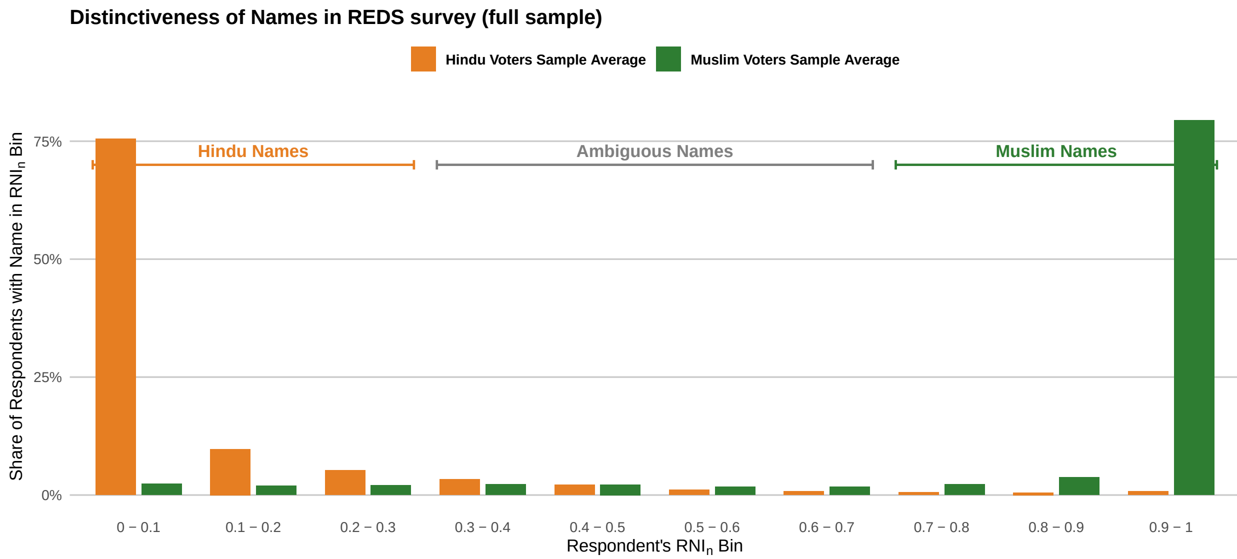


Figure A4: RNI validation using REDS survey

C.2 Do men drive our results?

Are our results driven by naming decisions only for men? It could be the case that parents prefer to transmit more religiously distinct names to only male children. If this were the case than we would see a growing *gender* polarization in RNI in the cross-section and over time. To check for this, we split our sample by gender (as reported in the voter rolls). We then report characteristics like the number of characters in a name (Figure A5 (A)) and the number of name parts in voters' full name (Figure A5 (B)). We conclude that while there exist temporal trends towards a prevalence of two-part names, there are no differences between male and female naming conventions.

Additionally, since male and female voter records differ by the relative included, we rerun Figure 4 split by gender in Figure A6. We note that there no discernible differences between the polarization patterns we document between men and women, allowing us to conclude that our findings are not differentially explained by naming of men.

Moreover, when studying the intergenerational transmission of names in Figure 6, we restricted the sample to men and unmarried women, since married women's voter roll shows their husband instead of their parents' name. As a robustness exercise, Figure A7 shows the same plot but includes married women, documenting the same patterns. This is likely because of the high degree of religion, class, and ethnic endogamy practiced in India.

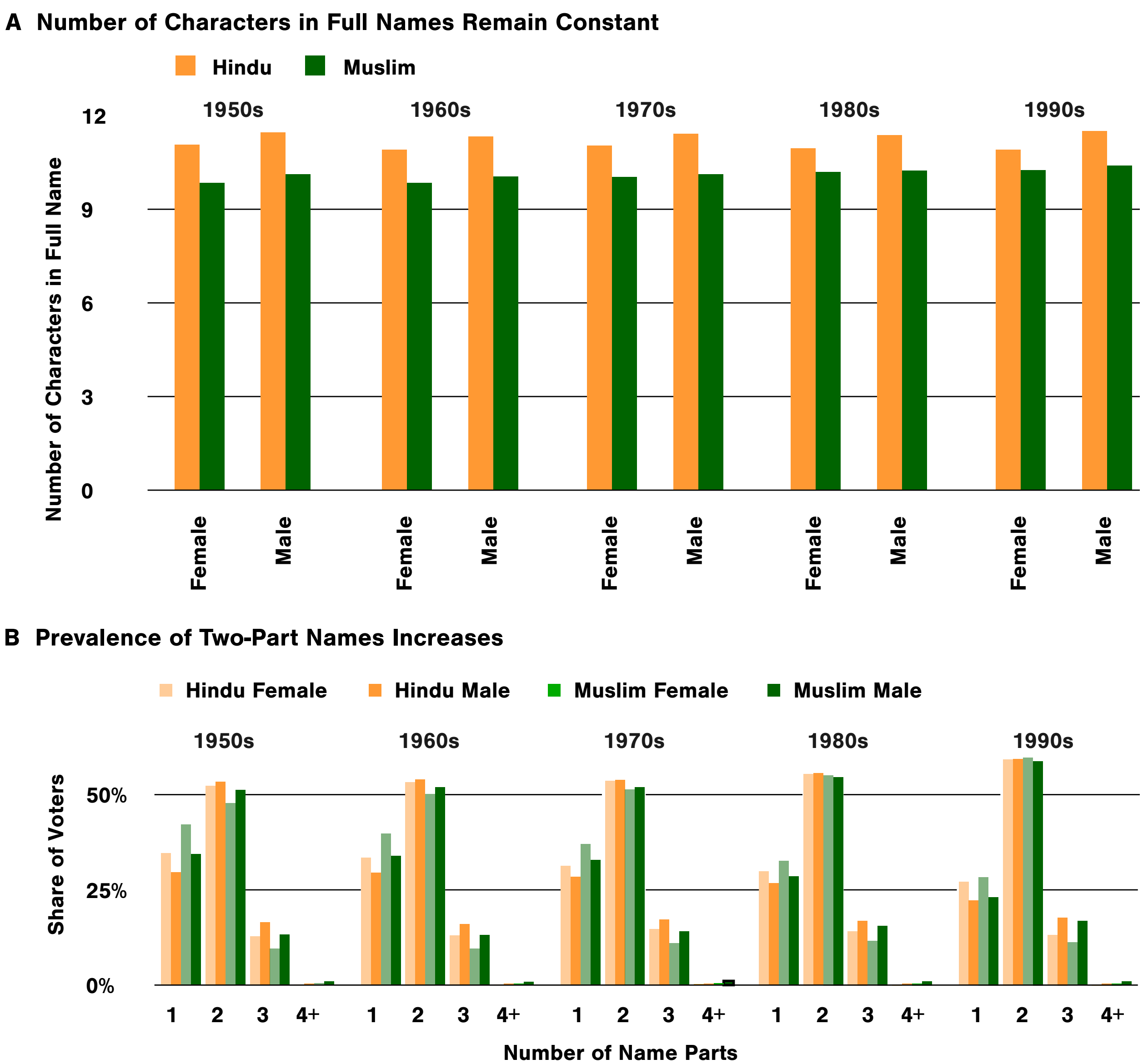


Figure A5: **Name Characteristics across Religion and Gender. (A)** Plots the average number of characters in voters' full names by gender, religion, and decade (N= 505,309,697). **(B)** Plots the average number of name parts in full name by gender, religion, and decade (N= 505,309,697).

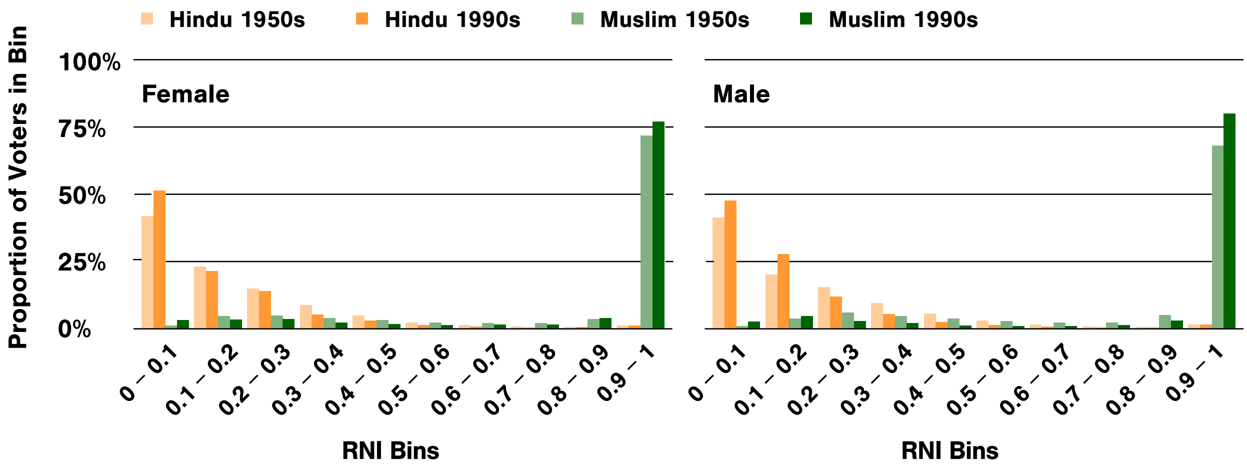


Figure A6: **Robustness for Gender.** Depicts Figure 4 divided by gender (N= 505,309,697).

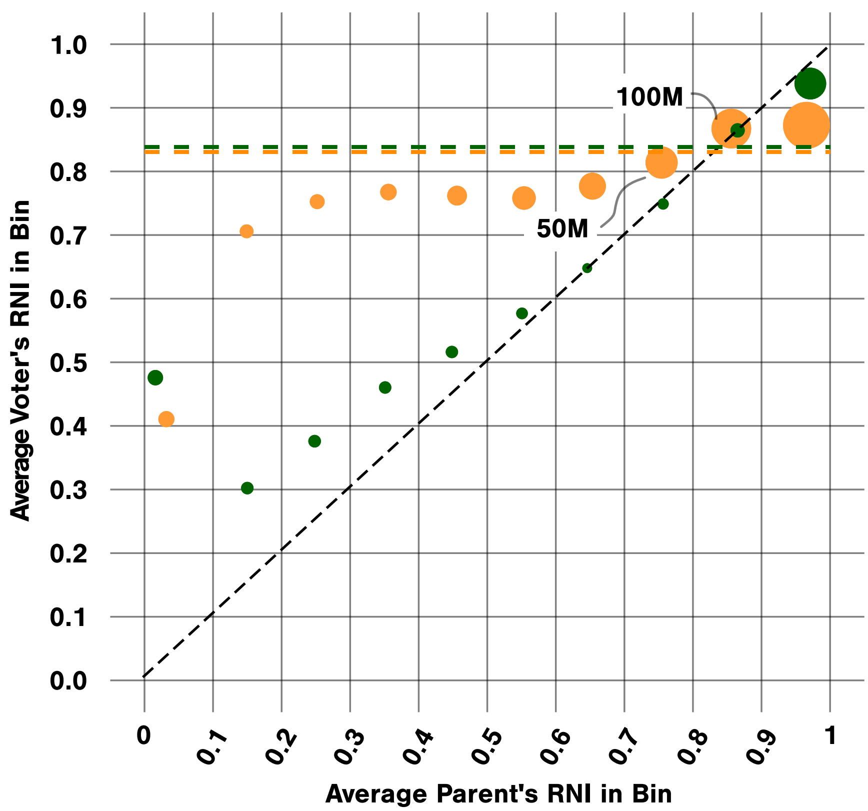


Figure A7: **Intergenerational Transmission of Name Distinctiveness including Married Women.** Replicates 6 including married females (N= 505,309,697).

C.3 Does the way we infer religion drive our results?

A concern with the analysis is that the classification of religion using the machine learning algorithm produces measurement error that explains the results. First we note that the religion prediction errors are uncorrelated with time, as shown in Fig A8. This means that overtime changes in accuracy do not drive changes in polarization over time. To probe this further, we carry out three exercises:

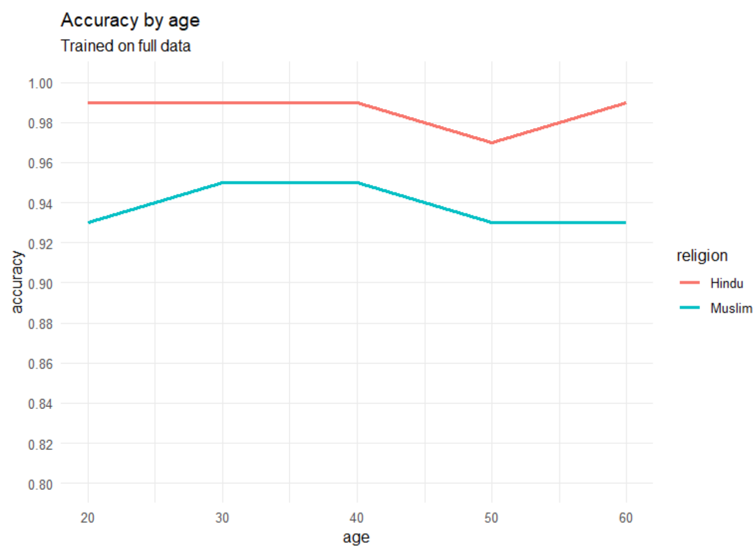


Figure A8: **Religion Prediction Accuracy over Time** Plots the accuracy of the religion prediction algorithm by the age of the voter.

First, we restrict the sample to observations where the prediction probability is greater than or equal to 0.75 as opposed to no restriction in the full sample, thereby removing observations where the religion inference is more precise. In fact, this restriction increases the accuracy in our validation dataset to 98%. Figure A11 shows the results. We see that the broad patterns we observe in this restricted dataset remain consistent: there is evidence of increasing polarization for both Hindus and Muslims as well as space for ambiguous names in the middle. However, we note that the restriction results is a small reduction in ambiguous names, while slightly strengthening the patterns of polarization.Therefore, the main results provide, if anything, a conservative assessment of cultural polarization in India.

Second, we probe the robustness of the classification algorithm by using a different algorithm. In our preferred specification in the main text, we show results with a multi-class algorithm that predicts, for each person on the voter roll, which of all the major religious groups in India the person belongs to. As an alternative, Figure A9 (A) reports results with a different religion classification model that predicts only whether a person is a Muslim or non-Muslim, where non-Muslims are taken as the majority group for the sake of this exercise. We find that this alternative coding decreases the

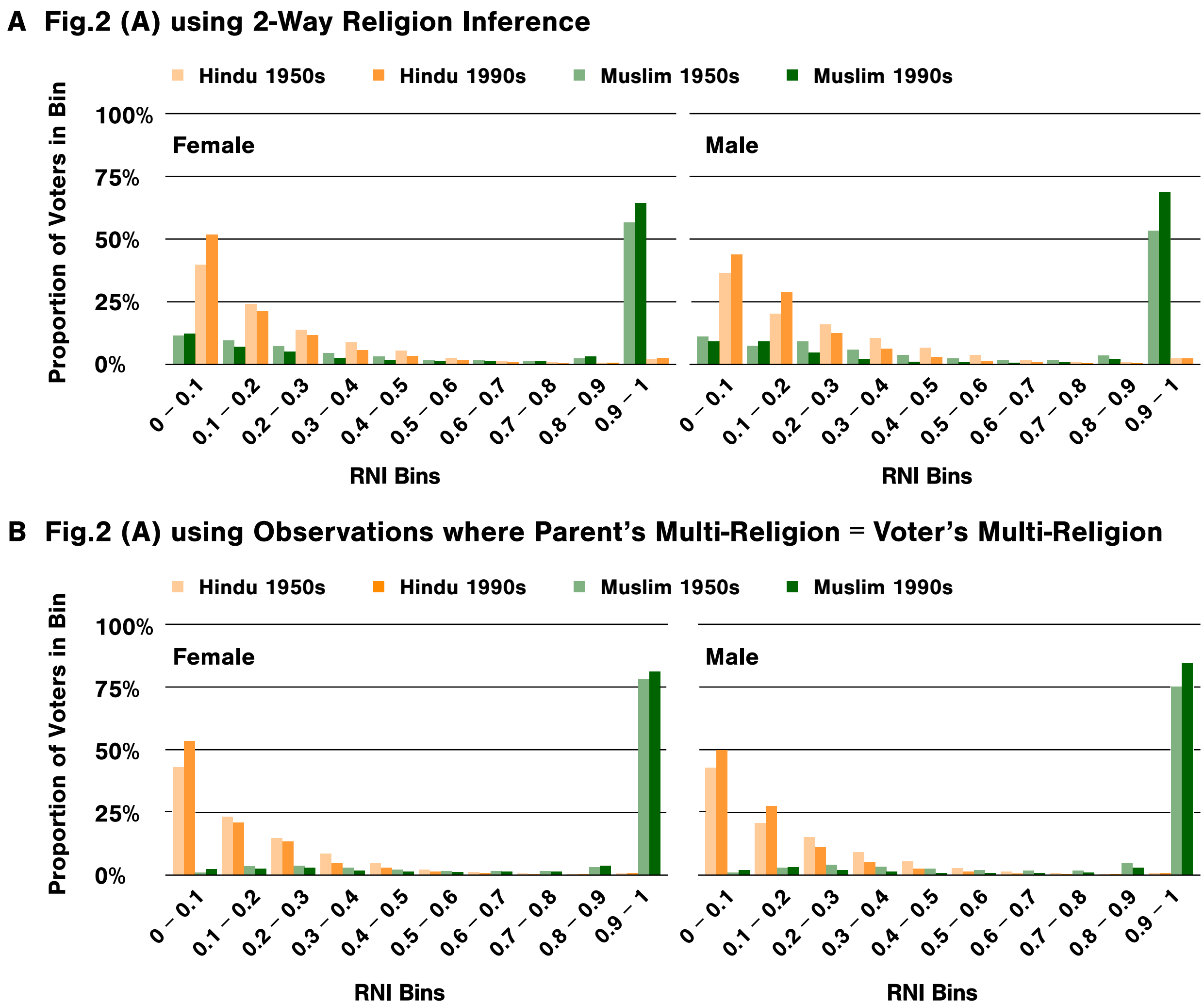


Figure A9: **Robustness for Religion Classifications. (A)** Uses an alternative (two-way) model for inferring voters' religion (N= 505,309,697). **(B)** Restricts the sample to observations where the inference about the voter's religion is the same as the one for the parent's religion (N = 129,698,342).

observed polarization of names. This is likely due to the inclusion of observations from religions other than Hindu or Muslim that would lead to an underestimation of the distinctiveness of names since these names would be counted as ambiguous, assuming that there is no strong selection of the other names into one of the two religions we study.

Finally, another check we can do is vary the name we use to infer the religion of the voter. In our main analysis, we implement the religion inference algorithm through the parents or husbands names of voters. We can also do the same exercise by implementing the religion classification algorithm based on the voter's full name itself and use only the data where both inferences match. Figure A9 (B) restricts the sample to cases where the inferred religion of the voter matches that of the parent. This restriction has little impact on our findings.

C.4 Does the endogeneity of RNI and religion prediction drive our results?

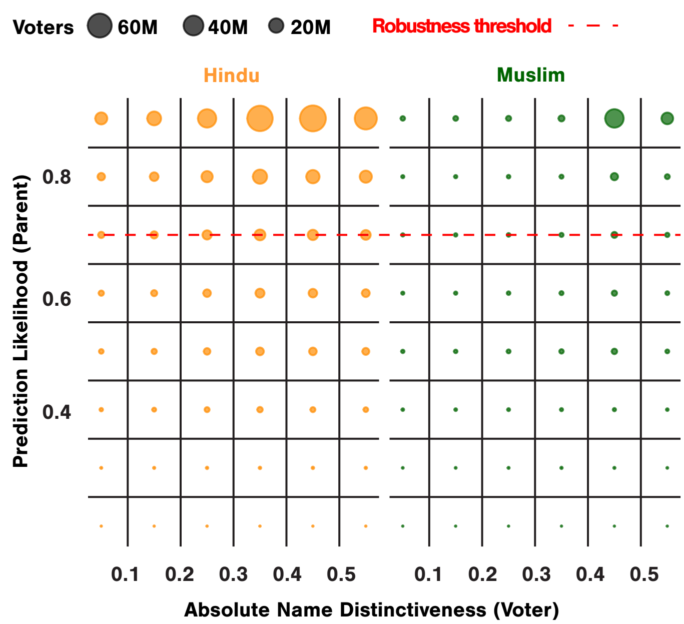


Figure A10: **Correlation between Religion Inference and RNI.** The x-axis plots the distance of a voter's $RNI_{n,t}$ from 0.5, i.e. a normalized measure of distinctiveness, laid out in bins of 0.1 width. The y-axis plots the prediction likelihood, a measure of how well the parent name (or husband name for married women) fit the model parameters when religion is inferred by the Chaturvedi and Chaturvedi (2022) algorithm. This measure is binned into 0.1 bins as well. There is a natural cutoff where no values under 0.2 exist. The size of the dots indicated how many voters fall into each bin (N = 505,309,697). The red line indicates the level of our robustness in A11.

Another potential concern relates to the endogeneity of the algorithm-based religion inference and the frequency-based measure a of name's religious distinctiveness (RNI). First, it is imporatnt to remember that the direct endogeneity of this is blocked through the use of the parent's/husband's name for religion classification, and the use of the voter's own name for the calculating of the RNI.

Second, to illustrate why this is not a major concern in our case, we first discuss how the religion inference algorithm, based on a voter's parent's/husband's name, does not deterministically predict religion based on the first name. Consider the example for the very common Hindu name: *Ram*. Having *Ram* in a voter's the parent's/husband's name does not deterministically map the voter's religion to be Hindu, despite the fact that Ram is a common Hindu name. To see why, note that the religious classification algorithm outputs time-invariant probabilities from 0 to 1 that are calculated with information on the *full* name of the voters' parents, including the last names. This creates important variation in our data: is a voter's parent called *Ram Ali*, they will get mapped as Muslim, while another parent called *Ram Singh* will get mapped as Hindu.

Third, in contrast to the religion prediction algorithm that uses the full names of parents/husbands, the frequency-based $RNI_{n,t}$ measure uses only the personal name of the voter. Additionally, $RNI_{n,t}$ is a time-varying score for a given name that is deterministic for a name and year. As a result, conditional on classifier probability from the religion inference algorithm, the voters' $RNI_{n,t}$ varies over a wide range. Figure A10 illustrates this. For instance, for names in the top row, those with high model fit, where the majority of our data lie, as shown by the size of the bubbles, we can see that there is substantial variation in names absolute distinctiveness (measured as the absolute deviation of a name's $RNI_{n,t}$ from 0.5).

Taken together, several factors allay concerns about the endogeneity of $RNI_{n,t}$ and religion prediction: religious prediction is based on parent's/husband's name while $RNI_{n,t}$ is based on voter's name, $RNI_{n,t}$ varies over time but religious prediction of name does not; and restricting our observations to names with better religion inference does not change the substantive findings. For the full sample, the correlation between the model fit for a name and its $RNI_{n,t}$ is 0.12, dissuading concerns raised by using names for both steps of our analysis.

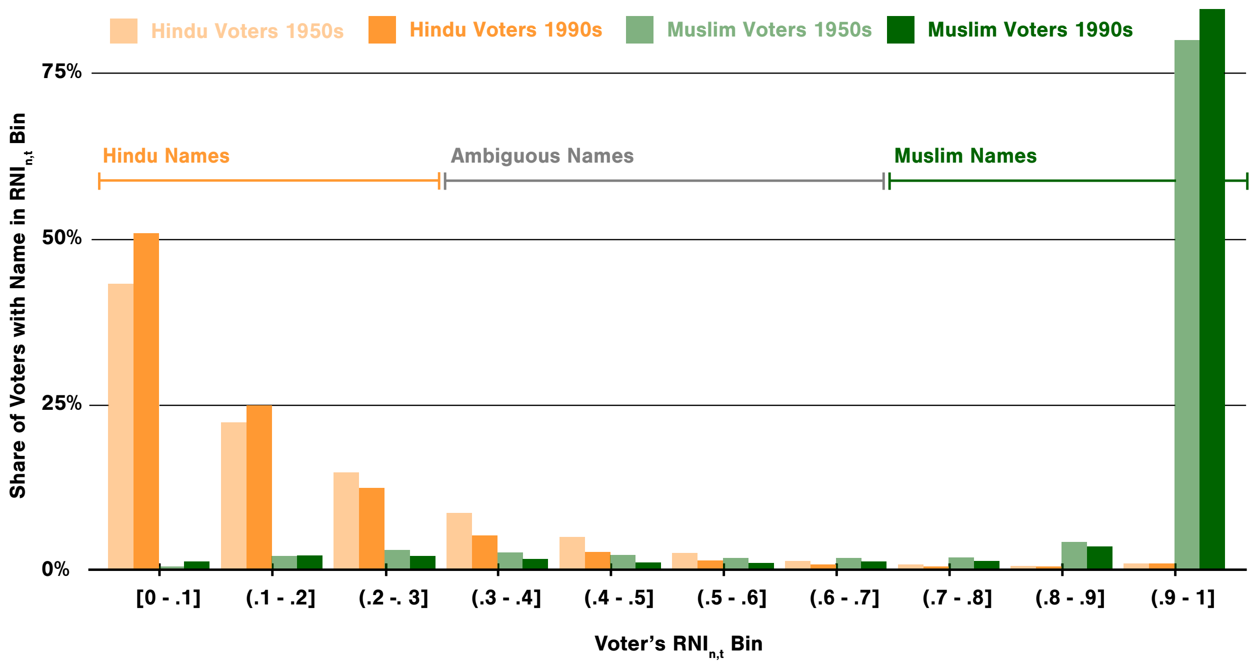


Figure A11: **Robustness of Figure 4 restricted to observations with better model fit in the religion inference step.** The $RNI_{n,t}$ range is divided into 0.1-width bins. The plot depicts the share of Muslims and Hindus whose name falls within each $RNI_{n,t}$ bin for 1950 and 1990. More weight towards the extremes indicates more distinctive names. The sample is restricted to observations with a stricter threshold applied to the religion inference (N = 145,083,698).