Building Tribes: How Administrative Units Shaped Ethnic Groups in Africa

American Journal of Political Science, conditionally accepted

Carl Müller-Crepon*

London School of Economics and Political Science

April 4, 2023

Abstract

Ethnic identities around the world are deeply intertwined with modern statehood, yet the extent to which territorial governance has shaped ethnic groups is empirically unknown. I argue that governments at the national and subnational levels have incentives to bias governance in favor of large groups. The resulting disadvantages for ethnic minorities motivate their assimilation and emigration. Both gradually align ethnic groups with administrative borders. I examine the result of this process at subnational administrative borders across Sub-Saharan Africa and use credibly exogenous, straight borders for causal identification. I find substantive increases in the local population share of administrative units' predominant ethnic group at units' borders. Powerful traditional authorities and size advantages of predominant groups increase this effect. Data on minority assimilation and migration show that both drive the shaping of ethnic groups along administrative borders. These results highlight important effects of the territorial organization of modern governance on ethnic groups.

Keywords: Ethnicity; Sub-Saharan Africa; Territorial Governance; Administrative Geography; GIS;

---

*email: `carl.muller-crepon@politics.ox.ac.uk`. I thank the editor, two anonymous reviewers, Jeremy Bowles, Lars-Erik Cederman, Jessica Gottlieb, Lukas Haffert, Horacio Larreguy, Emmy Lindstam, Moritz Marbach, Yannick Pengl, Pia Raffler, and Yuhua Wang and participants of seminars at ETH Zurich, Harvard University, Washington University in St. Louis, and University of Oxford, as well as the 2020 APSA Annual Meeting and the Swiss Political Science Conference for helpful comments and suggestions.

Ethnicity constitutes one of the most salient political cleavages. It affects public goods provision (Alesina and Ferrara 2005), redistribution (Franck and Rainer 2012; De Luca et al. 2018), and violent conflict (Horowitz 1985; Cederman, Gleditsch and Buhaug 2013). While it is well recognized that ethnicity and ethnic boundaries are socially and politically constructed (Barth 1969; Posner 2004, 2005; Wimmer 2013), less is known about the drivers of that process. In particular, there is only sparse systematic evidence on the transformative effect of modern state governance on ethnic identities highlighted in qualitative studies on Europe (Weber 1977) and Africa (Southall 1970; Young 1985).

This paper addresses this gap and examines how territorial governance, that is governance through spatially bounded administrative divisions, shaped ethnic groups in Sub-Saharan Africa. In doing so, I build on Mamdani's (2001; 2020) and Posner's (2005) seminal works on the relation between ethnicity and the colonial imposition of territorial governance by the state and traditional institutions. I argue that local and regional authorities tends to favor the largest ethnic group in their population, in particular where (neo-)traditional institutions are powerful. This incentivizes local ethnic minorities to assimilate into the majority identity or emigrate to co-ethnic governance units. The resulting change in ethnic demography crystallizes ethnic boundaries along often haphazardly drawn administrative borders and constitutes an important mechanism behind Iliffe's (1979,

p. 324) statement that "Europeans believed Africans belonged to tribes; Africans built tribes to belong to."

Current scholarship traces the origins of (political) ethnicity in Africa to geography (Michalopoulos 2012), colonial-era missionaries, cash crop agriculture (Pengl, Roessler and Rueda 2021) and indirect rule (Ali et al. 2019; McNamee 2019), ethnic coalitions (Posner 2004) and power distributions (Green 2021), as well as political entrepreneurs (Kayira, Banda and Robinson 2019; Robinson 2017).[1] I add a focus on the effects of territorial rule imposed by colonialism, which revolutionized local governance, fostered the "invention of tradition" (Ranger 1997), and thus transformed ethnic identities (Lentz 1995; Southall 1970; Young 1985). This focus also highlights an analogy between subnational 'tribe-building' in Africa and 'nation-building' elsewhere, both powered by material and ideological forces that aimed at increasing the congruence between ethnic groups and political units (Argyle 1969).

I test my argument that administrative units shaped ethnic groups by estimating the change in ethnic groups local population shares at units' boundaries using a regression discontinuity design. I find that the share of regions' (districts') main ethnic group sharply increases by 14 (8) percentage points or 54 (23) percent at borders with units dominated by a different

[1]A related literature assesses individuals' ethnic versus national identification (e.g., Eifert, Miguel and Posner 2010; Robinson 2014).

group. Ruling out omitted variable bias and reverse causality from endogenous borders and their change, these local average treatment effects are robust to restricting variation to within colonial-era settlement areas of ethnic groups, as well as to variation at relatively straight, arbitrarily drawn subnational boundaries. Estimated treatment effects increase in the ethnicization of local governance, proxied by stronger traditional institutions and larger size-advantages of predominant groups.

Assimilation and emigration by local minorities account for these border effects. Adapting the main regression discontinuity design, I find that local minorities assimilate to the majority through language adoption and intermarriage. In addition, census data on 33 million individuals show ethnic sorting between administrative units: local minorities emigrate more and immigrate less frequently than predominant groups in a manner that correlates strongly with the main treatment effects.

Evidencing the effects of administrative borders on local ethnic demography in Africa, the paper highlights the endogeneity of ethnic identities and geography as a larger issue for the study of ethnicity. Ethnic identities are, at least in the long run and within (unknown) limits, partially a result of ethnicized territorial governance. This root of ethnic identities raises the crucial question of when, where, and how else citizens and political elites foster ethnic change to warp the political playing field in their favor (e.g., Brass 1991; Posner 2004).

Theoretical argument

Governance through geographically bounded administrative units transforms ethnic groups because traditional and state authorities tend to ethnically specialize in governing large ethnic groups and discriminate against minorities. Minorities can improve their lot through assimilation or emigration, thereby selecting into a majority. Because ethnically biased governance and minority responses are delimited by administrative borders, the resulting transformation of ethnic groups aligns their geography with administrative boundaries. Minorities can alternatively demand secession, thus becoming a majority in a new governance unit. While important, I here focus on ethnic change, leaving border change mostly as an empirical challenge.

Ethnicized territorial governance

The establishment of administrative divisions in multiethnic states typically creates ethnically diverse units. This is because local ethnic diversity inhibits the drawing of homogeneous but non-overlapping and contiguous divisions. In governing their multiethnic population, local governments – here used broadly, including traditional authorities and state governments – frequently favor large, powerful groups. Extensively analyzed, governments often cater material goods and services to 'their' ethnic constituencies (Franck and Rainer 2012; De Luca et al. 2018) for intrinsic (Chandra 2007) or instrumental reasons (Fearon 1999). Because large ethnic groups hold, on average, most executive power (Bormann 2019), ethnic favoritism leaves ethnic minorities disadvantaged.

Ethnicized governance also emerges where governments 'specialize' in large ethnic groups by using specific languages or drawing on ethnic traditions to foster their legitimacy. Both tools can increase co-ethnics' "quasi-voluntary compliance" (Levi 1988). Facing a multiethnic population and economies of scale, specializing in small groups yields smaller, possibly negative benefits than specializing in large group(s). Governments will therefore specialize most in large groups, leaving minorities worse off.

Similar incentives can lead governments to ethnically homogenize minority populations (Alesina and Spolaore 2005). Governments thus 're-educate' ethnic minorities to learn the majority's language and customs and increase their interaction with the state (Weber 1977; Zhang and Lee 2020).[2] At the extreme, governments violently 'right-people' their population through displacement and genocide (O'Leary 2001).

In sum, I expect governments to favor the largest ethnic group in the population, leaving ethnic minorities, disenfranchised, under-serviced, and potentially subject to homogenization policies. The strength of these dynamics increases with the relative size of the largest ethnic group.

[2]See Fouka (2019) on potential backlash.

Minority responses to ethnicized governance

To improve their lot under ethnicized governance, ethnic minorities may choose to assimilate, emigrate, or mobilize to create their own governance unit. Ethnic assimilation aims at changing one's language, religion, appearance, and taste to be able to better 'pass' as a member of the majority group and claim its benefits (Laitin 1995). As some group characteristics are innate or learned during childhood, assimilation often proceeds intergenerationally through ethnic intermarriage (Kalmijn 1998). Alternatively, emigration offers an exit option for minority members who face discrimination in their governance unit. Migrants may either ethnically sort into units where they belong to the ethnic majority or head to prosperous areas where discrimination is offset by economic opportunity (Docquier and Rapoport 2003).

Assimilation and ethnic sorting through migration increase the relative size of the largest group. As in Schelling's (1971) tipping point model, this will, ceteris paribus, reinforce governments' ethnic biases and reinforces minorities' incentives to assimilate or emigrate. However, parallel *heterogenizing* processes such as non-ethnic migration likely prevent a stable, homogeneous equilibrium.

Spatially concentrated minorities may also mobilize against their discrimination and demand their own governance unit, achieved often by 'upgrading' one or multiple subunits (Green 2008; Grossman and Lewis 2014).

New borders can align administrative and ethnic geographies more closely but create new minorities where cutting through ethnically diverse populations. While certainly important, I here focus on *ethnic* change and address endogenous border change empirically below.

In sum, I argue that the initial geography of governance units and ethnic groups determines ethnic groups' status within each unit. In response to governance biased towards units' largest groups, minority assimilation and ethnic sorting through migration increase the relative size of plurality groups. As this process is spatially bounded, I expect a sharp increase of the local population share of unit's dominant groups at their borders.

Administrative units and ethnicity in Africa

Sub-Saharan Africa provides a suitable testing ground for my argument. European colonialists created multi-ethnic administrative units with borders that often disregarded local ethnic geography. The ethnicization of subnational governance, particularly by 'traditional' authorities, incentivized minorities to assimilate or emigrate. Structuring this process spatially, states' administrative borders thus shaped ethnic groups.

While I expect territorial governance to affect ethnic identities beyond the subnational level in Sub-Saharan Africa, my empirical focus has two benefits. First, African states are highly diverse, and most feature no majority group or homogenization into single ethno-national identities (Laitin

1992). Subnational borders that ‘cement’ ethnic identities contribute to explaining this pattern. Second, African borders were often haphazardly drawn (e.g., Herbst 2000), reducing the risk of reverse causality incurred when studying border effects elsewhere.

The colonial introduction of administrative borders

Defining the state via its territory demarcated by borders is integral to the idea of modern statehood (Weber 1919), but was virtually unknown to most of pre-colonial Africa (Asiwaju 1983). Instead, even administratively centralized polities were unbounded and non-contiguous, their power radiating outwards from the center (Herbst 2000). Political borders were conceptually even more foreign to acephalous societies where the lack of centralized power made separation lines superfluous.

European colonizers radically changed these political topographies. Carving up the continent into colonies, they partitioned each into administrative units to roll out the territorial governance that established purported effective control. This creation of regions, districts, and further subdivisions was as revolutionary as the drawing of international borders (Asiwaju 1983). Both sharply delimited the territorial scope of (sub)national governance by the state and the traditional institutions co-opted by it.

The design of governance units determined their initial ethno-demographic composition and may have been influenced by ethnic geography. Such influence was likely strongest in areas ruled indirectly through precolonial authorities, in particular the British colonies (Crowder 1968; Müller-Crepon 2020). Here as elsewhere, the predominant administrative mindset expected individuals to belong to tribes, "discrete, bounded groups, whose distribution could be captured on an ethnic map" (Young 1985, p. 74). Yet, the idea of 'tribal homelands' as 'natural' governance units (Asiwaju 1970; Crowder 1968) clashed with a reality of interspersed ethnic settlement areas (Cohen and Middleton 1970) and political loyalties that cut across ethno-spatial lines (Lentz 1995, Southall 1970). While perceptions of ethnic geography likely influenced the broad outlines of administrative units, this incompatibility meant that pragmatism coupled with administrative and geographic exigencies determined the precise location of borders (Lentz 2006, p. 53), thus forcing "the round peg of existing authority patterns into the square hole of territory-based administration" (Posner 2005, p. 30).

Administrative geographies and the transformation of ethnicity

The introduction of territorial governance changed the relationship between rulers and their people from governance based on 'ethnic' allegiances to governance based on individuals' place of residence (Herbst 2000). However, ethnic identities remained important, being directly tied to customary law that often disenfranchised local minorities (Mamdani 2001). This ethnicization of local governance powered the shaping of ethnic groups along administrative borders.

As the "custmary" became the bedrock of local colonial rule and continued being influential in many post-colonial states, governance by local state and traditional authorities became ethnicized quickly. This is particularly well evidenced with regard to the distribution of the "goods of modernity": property rights, market access, and social services (Bates 1974). Based on ethnically exclusive definitions of the 'customary,' these goods could and can often still today be distributed to local elites and their ethnic constituents (e.g., Bates 1974; Posner 2005; Vaughan 2003). While such favoritism sometimes relates to local public goods such as wells, schools, and roads (Ejdemyr, Kramon and Robinson 2018; Harris and Posner 2019), minorities tend to also be individually disadvantaged by a biased access to (customary) land rights (Boone 2014; Honig 2017), justice (e.g., Choi, Harris and Shen-Bayh 2022), jobs (e.g., Brierley 2021; Hassan 2017),[3] material handouts such as food or seeds (van Hoorn and Rademakers 2021), and education provided in the local vernacular (Pengl, Roessler and Rueda 2021).

Ethnic favoritism is thus not only a feature of African national politics, but also local politics. Afrobarometer (2018) data shows that local ethnic

[3]Local education or health systems dominated by the local majority likely bias service provision towards their ethnic kin.

minorities perceive local authorities as more unresponsive, tend to mistrust them more, and approve of them less than members of plurality groups. These patterns are stronger vis-à-vis traditional authorities than state authorities (Appendix A), which reflects their continuing reliance on the ethnically defined 'customary.'

Administrative borders defined and still define units' ethnic make-up, assign minority or majority status to individuals, and spatially delimit patterns of local ethnic favoritism. As argued above, this incentivizes local minorities to become part of a local majority through assimilation or emigration (see also Posner 2005). Ethnic assimilation is historically frequent across Sub-Saharan Africa, in particular among ethnic 'strangers' (e.g., Cohen and Middleton 1970). For example, Kenyan Kikuyu settlers in a former Maasai reserve assimilated by adopting language and traditions, as well as through intermarriage to secure land rights (Gravesen and Kioko 2019).[4] In turn, emigration of ethnic minorities has been described as a vehicle of 'revolt' against local discrimination by Asiwaju (1976).

Being important structuring elements of local politics, customs, traditions, and ethnic identities did not remain uncontested. Instead, the "invention" (Ranger 1997) of traditions and history became a tool for polit-

[4]See Stahl (1991) for similar evidence from Ghana and Schultz (1984) from Northern Cameroon. Green (2021) relatedly finds citizens attempting to pass as their presidents' co-ethnics.

ical survival still used today (Iliffe 1979; Robinson 2017). Struggles over the customary played out, for example, as Councils of Elders in the Kenyan Taita Hills synthesized lineage practices to control the chiefs (Bravman 1998, p. 157) or when obas, chiefs, and educated elites reconstructed contending versions of traditional authority in Nigeria's Oyo Province (Vaughan 2003, p. 301). Such politically driven cultural and ethnic change is likely again bounded by the administrative borders that define the local political arena and individuals' incentives to adopt the changes fostered from above.

In sum then, prior literature suggests that ethnicized territorial governance spurred ethnic change within the boundaries of administrative units. As a result and throughout the colonial and post-colonial period, minority assimilation and migration aligned ethnic geography with administrative borders. This alignment should increase with stronger traditional institutions, which entail greater bias towards locally predominant ethnic groups. The following quantitative analysis examines this argument systematically.

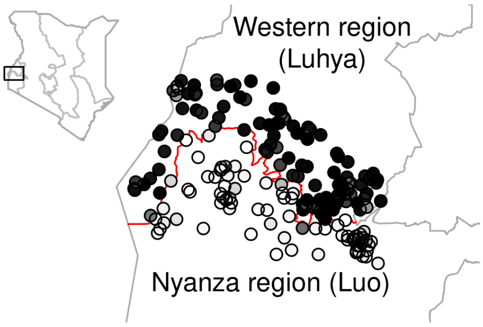
Research design

I investigate whether administrative boundaries have shaped ethnic groups in Sub-Saharan Africa by examining individuals' ethnic identity across Demographic and Health Surveys (DHS 2018) from 25 countries. Building on studies of African border effects (McCauley and Posner 2015), I estimate the impact of administrative borders on the local population share of administrative units' plurality group in a spatial regression discontinuity design (RDD, Keele and Titiunik 2015; Henn 2022).[5] Focusing on credibly exogenous intra-ethnic and straight borders addresses omitted variable bias and reverse causality. Additional analyses shed light on ethnic assimilation and migration as main mechanisms.

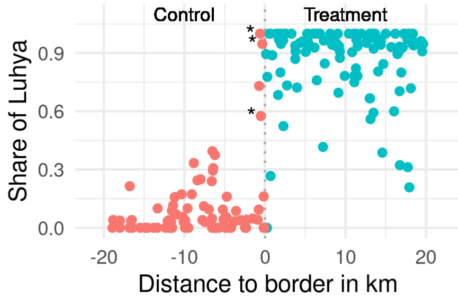
The curious (and extreme) case of the Kenyan Luhya

The extreme and non-representative case of the Luhya in western Kenya illustrates the intuition behind the RDD. At the outset of the 20th century, Bantu-speakers dominated the North Kavirondo district, speaking between 15 and 26 mutually intelligible dialects (MacArthur 2012). Elites from the Wanga held power in the district, which became part of Kenya's Western province after independence. Gold discovered in the early 1930s led local elites to foster a collective identity to fend off settlers. The resulting "umbrella-group" Luhya (or Luyia, 'kinship') quickly became one of Kenya's main tribes with more than 650'000 members in 1948 (MacArthur 2013). In neighboring South Kavirondo, the postcolonial Nyanza province, the Luo exhibited a similar rise to ethnic self-consciousness and political im-

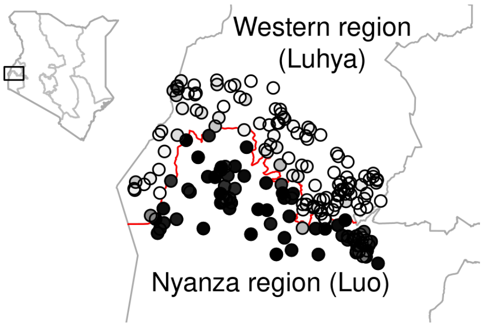
[5] An alternative research design would compare individuals' (localities') ethnic identity (composition) before and after the introduction of administrative borders. This is currently not feasible. Virtually all data on ethnicity is cross-sectional or aggregated to coarse and changing administrative units with low temporal resolution and cross-country coverage.



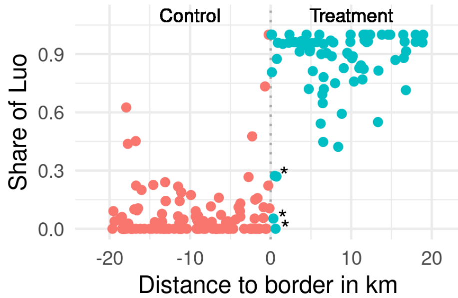
(a) Luhya share



(b) Luhya share



(c) Luo share



(d) Luo share

Figure 1: Luhya and Luo around the Western-Nyanza regional border in Kenya.

Note: Panels (a) and (c): gradient from white to black equivalent to 0 to 100%. The Western region is the 'treated' region in panel (b) and the 'control' region in panel (d). Dots marked with a * in (b) and (d) are geographically attributed to the wrong region and have a flipped treatment/control status. Data comes from the Demographic and Health Surveys (2018).

portance, fostered among others by the Luo Language Committee (Peterson 2018).

How did this transformation and politicization of ethnic identities affect the ethnic demography in the Western and Nyanza provinces? Evidencing the extreme success of the Luhya identity as a regionally bounded construct, Figure 1 shows a sharp change at the border between today's Western and Nyanza provinces. The share of the Luhya population in enumeration areas of the (DHS 2018) drops from an average of roughly 90%

to 5% as one crosses the border from the Western province, dominated by the Luhya, into the Nyanza province which is predominantly Luo. Conversely, the Luo population increases from approximately 10% to more than 90% as one enters Nyanza. The presence of third groups makes the two border effects asymmetric. While the integration and politicization of Luo and Luhya identities likely drove the historical process at the macro-level, the sharp ethnic change at the border must be driven by horizontal change through individual assimilation or migration between the Luyha, Luo, and third groups. Otherwise, we would observe a smooth Luhya-Luo gradient around the border – if the border was drawn in a manner unrelated to local ethnic demography. I will pay particular attention to this assumption.

Estimation strategy and data

The regression discontinuity design (RDD) generalizes the intuition behind Figure 1 for regions and districts across Sub-Sahara Africa. As in the above example, each border between administrative units with differing plurality groups entails two treatments, one for each side. I capture this logic by "stacking" two RDDs per administrative border. This implies that each enumeration area (EA) associated with a point coordinate p enters the analysis twice with two outcomes. It is part of the treatment group of one of the RDDs with the local population share of the plurality group of its own administrative unit as the outcome. And it is part of the control group of the other RDD with the share of the plurality group of the neighboring unit as the outcome.

This research design avoids an arbitrary assignment of treatment and control groups. As Appendix F.1 shows, the two-sided design therefore leads in expectation to the same but more precise point estimate than any one-sided design. In addition, it balances treatment and control groups. Because each EA is part of both groups, geographic covariates and the density function of the running variable are perfectly continuous at the border (Appendix C).

Following Keele and Titiunik (2015), the baseline specification amounts to

$$Y_{p,s,b,t} = \alpha_{b,t} + \gamma_s + \beta_1 T_{u,t} + \beta_2 D_{p,b} + \beta_3 D_{p,b} \times T_{u,t} + \epsilon_{p,u,b,t} \tag{1}$$

where the outcome $Y_{p,s,b,t}$, the fraction of respondents in enumeration area (EA) p and administrative unit u enumerated in survey s that identifies with the local plurality group as defined by the treatment t at border b. Because each border entails two treatments t, the main treatment dummy $T_{u,t}$ takes the values 0 and 1 for each EA. $\alpha_{b,t}$ denotes a border $\times$ treatment-side fixed effect and γ_s a survey-round fixed effect. I add $D_{p,b}$, the distance of p to the respective border segment b, and the interaction of $D_{p,b}$ with the treatment dummy.

I cluster standard errors on the EA level to account for the correlation of the outcomes within them, and the administrative unit u $\times$ treatment

t level to capture the clustering of treatment assignments. In the baseline specification, I only analyze EAs within 20km to the closest border.[6]

To estimate Eq. 1, I combine spatial data on administrative units and ethnic settlement areas with georeferenced (DHS 2018) data collected since the 1990s in 25 African countries (Appendix B.1). I draw on the geographic data on districts and regions in 1990 from FAO's (2014) GAUL database.[7]

I derive the local plurality group of each administrative unit by spatially intersecting it with a map of ethnic "homelands" in the late nineteenth century[8] compiled by Murdock in 1959 (Nunn and Wantchekon 2011). The ethnic group that covers the largest area of a unit is coded as its plurality group.[9] Groups' local plurality status proxies their politically predominant status, assuming that the largest ethnic group on average holds most political power (e.g., Bormann 2019). This approach comes with two caveats.

[6]I show robustness to varying thresholds below. Because increasing the bandwidth adds new borders, optimal bandwidth estimators are inconsistent (see also Henn 2022).

[7]Appendix E and F.8 show robustness to colonial and alternative contemporary border data.

[8]While the exact time at which groups are depicted is unknown, the map is the earliest detailed and complete pan-African map of ethnic groups available.

[9]Identifying local plurality groups from survey data would introduce post-treatment bias.

First, I may mis-identify some groups because of measurement error from the map's low spatial precision and neglect of overlapping settlement areas. Where unsystematic, this should bias my estimates towards zero. Second, Murdock's knowledge of administrative borders may have biased the map. I address this concern through within-ethnic group comparisons, the use of alternative ethnic settlement data from the 1960s (Appendix F.7), and by showing that results do not vary systematically with the alignment between administrative borders and Murdock's map (Appendix D.1).

In the next step, I delineate all borders between administrative units with differing plurality groups and assign each EA from the DHS to its administrative district and region. I only keep borders between administrative units with at least one EA closer than 20km to either side of the border. If an EA is closer than 20km to one or more remaining borders b, it is assigned to the closest border. In a last step, I compute the two outcomes as the shares of the respondents in each EA that identifies with the plurality groups in its own and neighboring units. Ethnic identities are enumerated by the DHS mostly as respondents' ethnic group, tribe, and language. I use the ethnic link from Müller-Crepon, Pengl and Bormann (2022) to match ethnic labels from the DHS to those on Murdock's map.[10]

In combination, these data come with two important caveats. First, the

[10] I link groups if they share at least one common dialect.

precise locations of administrative boundaries are uncertain.[11] This perturbs observed treatment assignment and biases estimates of β_1 towards zero. Second, EAs' coordinates are randomly displaced to preserve respondent's privacy, dislocating 99% (1%) of all rural clusters by up to 5km (10km) (Burgert et al. 2013).[12] As noted in Figure 1, some survey clusters are therefore assigned to the wrong administrative unit and treatment status. A number of robustness checks address these issues.

For an unbiased estimate of β_1 in Eq. 1 borders must be drawn as-if-randomly at the local level. They must not line up with sharp precolonial ethnic boundaries (reverse causality) or any other geographical feature that causes spatial discontinuities in ethnic geographies (omitted variable bias). As discussed above, some administrative borders *roughly* followed ethnic geographies as perceived at the time. This entails the risk that they may have *exactly* lined up with sharp ethnic boundaries by chance or design. Other borders were drawn more haphazardly, as straight lines cutting across geographical features and ethnic settlement areas. I will use such borders below to improve causal identification.

If the main identifying assumption holds, the RDD identifies the local

[11] 50% of border locations in the GAUL and GADM data differ by less than 100 meters but 25% diverge by more than 1000 meters.

[12] Coordinates are displaced within the 'right' region (but not district) on the basis of border data that does not always align with the GAUL data.

effect of the change in a unit's predominant group on that group's population share at the border. Because borderlands are peripheral, the analysis draws on a population with a low plurality group share of 33% and 39% for regions and districts, respectively. The sample is more rural, older, less educated, and materially poorer than other respondents in the same units. Its historical, ethnic, or environmental characteristics do not differ systematically (Appendix C).

Empirical analysis

Figure 2 plots the sharp increase in the share of local plurality groups at the 323 regional and 1'019 district borders in the data. Closely coinciding, Table 1 presents the results of estimating Eq. 1 in Models 1 and 4. At the regional level, the share of treatment units' plurality group increases by 14 percentage points as one crosses from control into treatment units. This amounts to an increase of 54 percent from a plurality share at the border of 26 percent in the control group to 40 percent in the treatment group.[13] The effect of district borders amounts to 8 percentage points, bringing about an average increase of 23 percent in the plurality share at the border from 35 to

[13]In line with Eq. 1, the control group outcome at the border is computed as the weighted average of all fixed effects.

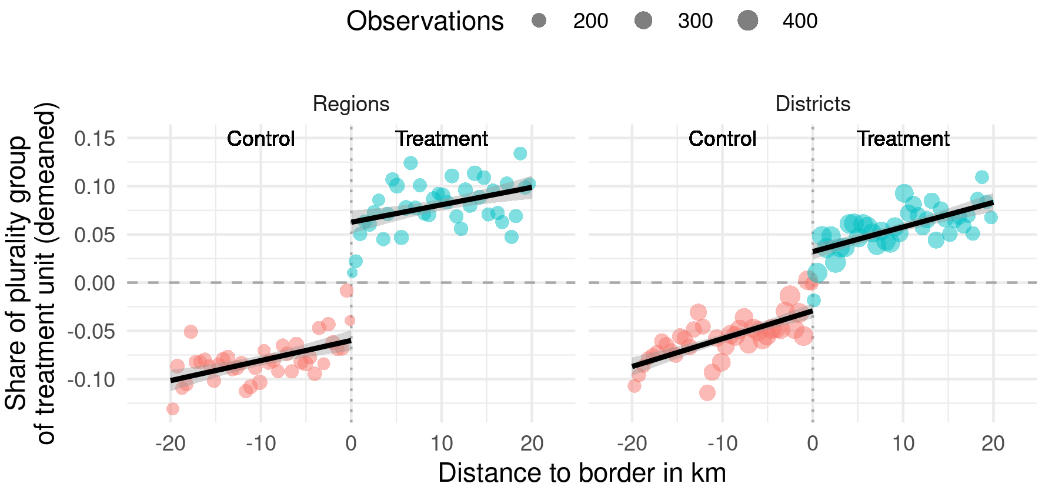


Figure 2: Local population shares of units' plurality groups increase at their borders.

Note: Shows the demeaned percentage of units' plurality groups within 20km of borders between treatment and control units, and linear trends on each border side.

43 percent.[14] These effects are precisely estimated.

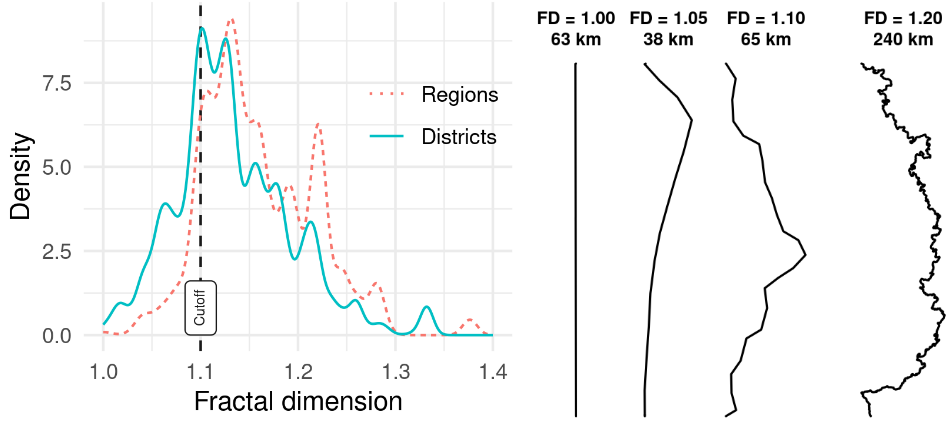
I assess the estimates' causal interpretability by zooming in on plausibly exogenous borders. Addressing potential reverse causality from ethnic geography, the first test focuses on survey clusters separated by administrative borders but located in the same ethnic settlement area on Murdock's (1959) map. Econometrically, this "within-group" analysis exchanges the previously border fixed effect with a border-ethnic group intercept. This precludes that the results are driven by an alignment of ethnic boundaries and administrative borders. The results in Models 2 and 5 closely coincide

[14]Effects of district borders within regions amount to 4 percentage points, while effects of district borders aligned with regional ones correspond to Model 1 in Table 1 (Appendix D.2). Border effects thus increase in the scale of administrative units.

Table 1: Effect of administrative borders on the population share of local plurality groups

	Outcome: Plurality group share (0-1)					
	Regions			Districts		
	Base	W/in grps.	Frac. dim.	Base	W/in grps.	Frac. dim.
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.143***	0.136***	0.103***	0.079***	0.085***	0.091***
	(0.018)	(0.021)	(0.031)	(0.011)	(0.015)	(0.020)
Distance to border	0.003***	0.001	0.001	0.004***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Distance × Treated	−0.0003	0.001	0.004*	−0.0005	−0.002	0.0004
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Cutoff	20km	20km	20km	20km	20km	20km
Max fractal dimension	2	2	1.1	2	2	1.1
Running var linear	yes	yes	yes	yes	yes	yes
Survey FE:	yes	yes	yes	yes	yes	yes
Border FE:	yes	–	yes	yes	–	yes
Group-Border FE:	no	yes	no	no	yes	no
Mean DV:	0.33	0.34	0.31	0.39	0.4	0.37
Control DV at border:	0.26	0.27	0.24	0.35	0.36	0.32
Borders:	323	785	92	1019	1283	512
Observations	15,396	10,186	2,562	23,180	13,240	9,250
Adjusted R^2	0.595	0.666	0.679	0.648	0.699	0.655

Notes: OLS linear models following Eq. 1. The unit of analysis is the survey cluster. The outcome is the share of respondents in a cluster from the treatment unit's ethnic plurality group. The treatment coefficient captures the increase in the share of administrative units' plurality groups at their borders. Standard errors clustered on the EA and administrative unit × treatment levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01



(a) Distribution (b) (c) (d) (e)

Figure 3: Fractal dimension of borders.

Note: Distributions and examples are based on observations in the baseline analysis.

with the baseline results.

Second, a set of relatively straight borders further addresses reverse causality and potential omitted spatial features that cause administrative borders and sharp discontinuities in ethnic population shares. Assuming that straight borders are least likely caused by ethnic geography or omitted spatial features, I measure borders' straightness as their fractal dimension (Alesina, Easterly and Matuszeski 2011), the degree to which they fill a two-dimensional plane.[15] Straight lines have a fractal dimension of 1 and wiggly lines approach a value of 2. Figure 3 plots the distribution of the fractal dimension of observed borders, as well as four example lines. I limit the sample to EAs along borders with a low fractal dimension of less than 1.1. This corresponds to retaining only 16% (40%) of survey clusters along 92 (512) regional (district) borders. Shown in Subfigure 3d, these borders con-

[15]See Appendix B.2.

sist of few, straight line segments. The effects of straight borders in Models 3 and 6 are consistent with the baseline results. The estimated effect of regional borders slightly decreases to 10.3 percentage points while that of district borders slightly increases to 9.1 percentage points. Results are robust when varying the fractal-dimension cutoff and using borders' alignment with rivers or watersheds as arbitrariness measures (Appendix D.1). Taken together, this shows that the main estimates are not substantively affected by reverse causality or omitted variable bias.

In principle, Table 1 reveals little information about the *global* effect of groups' plurality status on ethnic demographics beyond the *local* effect at units' borders. Two additional analyses suggest that the main estimates generalize to administrative units' interior. First, the treatment and control trends of plurality groups' population share towards administrative borders do not significantly differ in slope (see Figure 2 and Table 1).[16] This absence of effect bunching or reversal is suggestive of substantive increases in plurality groups' population share across the analysed 20km bandwidth. A second, correlational analysis in Appendix F.10 compares minority and plurality group shares among *all* DHS respondents. It suggests that units' plurality group shares are approximately twice as large as expected from

[16]Model 3 suggests a slightly steeper slope for the treatment group ($p <$ .1), which could imply growing effects towards the interior of the treated unit.

groups' territorial settlement shares alone, an increase which is larger than the main treatment effects.

In sum, the results support the hypothesis that administrative borders have significantly shaped ethnic groups. As one enters a region (district), the share of its largest ethnic group increases by approximately 14 (8) percentage points. This local treatment effect is consistent along credibly exogenous borders and likely generalizes to units' interior. I next assess whether treatment effect heterogeneity corresponds to observable corollaries of the theoretical argument and test the estimates' robustness. I then investigate assimilation and migration as mechanisms driving the results.

Treatment effect heterogeneity

If minority discrimination by local governance drive the shaping of ethnic groups along administrative borders, treatment effects should increase with stronger traditional institutions that tend to be ethnically exclusive. Effects should also increase in the population share and margin of the plurality group, as both incentivize greater ethnic specialization of governance.

I test these arguments by linearly interacting the treatment indicator and all other RDD-terms in Eq. 1 with variables that operationalize these moderators: (1) a constitutionalization index of traditional institutions that proxies for their influence;[17] (2) the share of a unit's territory settled by the plurality

[17] A principal component that explains 88% of variation in Holzinger et

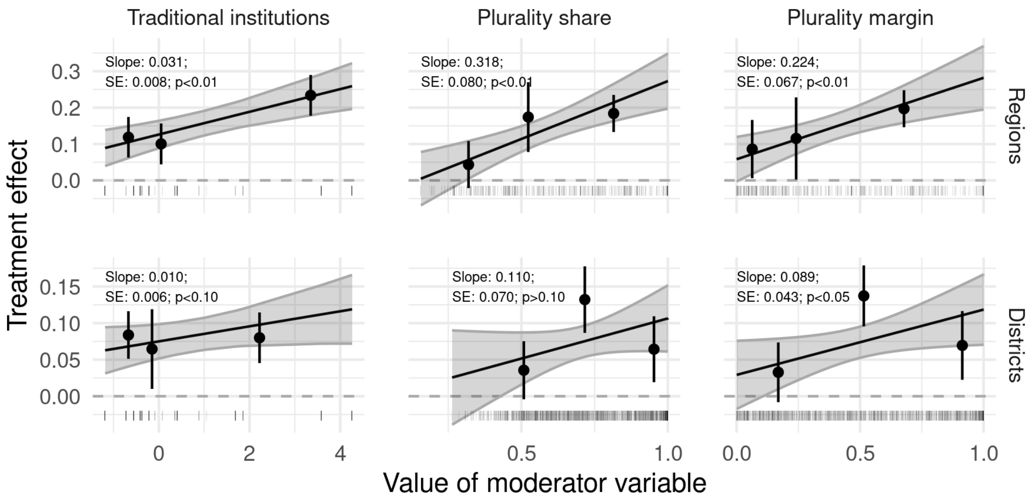


Figure 4: Treatment effect heterogeneity

Note: Results from linear interaction models and estimates by tercile of the moderator (Hainmueller, Mummolo and Xu 2019). Bars denote sample observations.

group; and (3) its margin over the second largest group-share.

Figure 4 shows that treatment effects consistently increase with more constitutionalized and thus powerful traditional authorities.[18] Effects are also stronger in administrative units with larger plurality groups and ones that enjoy larger population advantages over the second-largest group. These patterns are stronger at the regional than at the district level where the first interaction is estimated noisily ($p < .1$) and the latter two feature a slight non-linearity. In sum, administrative units with powerful traditional institutions and large plurality groups exhibit effects that are 1.5 to

al.'s (2019) data on the constitutional acknowledgment, regulation, and positive integration of traditional institutions.

[18]Appendix D.3 shows consistent differences between former British and French colonies.

2 times larger than average treatment effects. Additional analyses suggest that treatment effects do not increase with groups' post-colonial inclusion in national governments and are larger where groups have a history of ethnic civil war (Appendix D.4).

Effect timing

Probing the temporal dynamics underlying the main results, two additional analyses show that the main treatment effects are likely the result of colonial and post-colonial ethnic change. Figure 5 first assesses whether treatment effects increase among later-born individuals as effects accumulate across generations. Region-level treatment effects more than double between individuals born in the 1940s and 1990s while district-level effects show no increase. Assuming that most assimilation and migration happens early in individuals' life, substantive treatment effects among individuals born in the 1940s evidence that the main effects are not purely driven by post-colonial dynamics.

A second analysis in Appendix E accounts for post-colonial border change and compares treatment effects among (post-)colonial administrative borders that either changed or remained stable. I find somewhat smaller, yet substantive effects at the borders of colonial regions and districts, in particular those that survived until 1990, but not along colonial borders that have disappeared. In turn, effects at 1990 borders of colonial

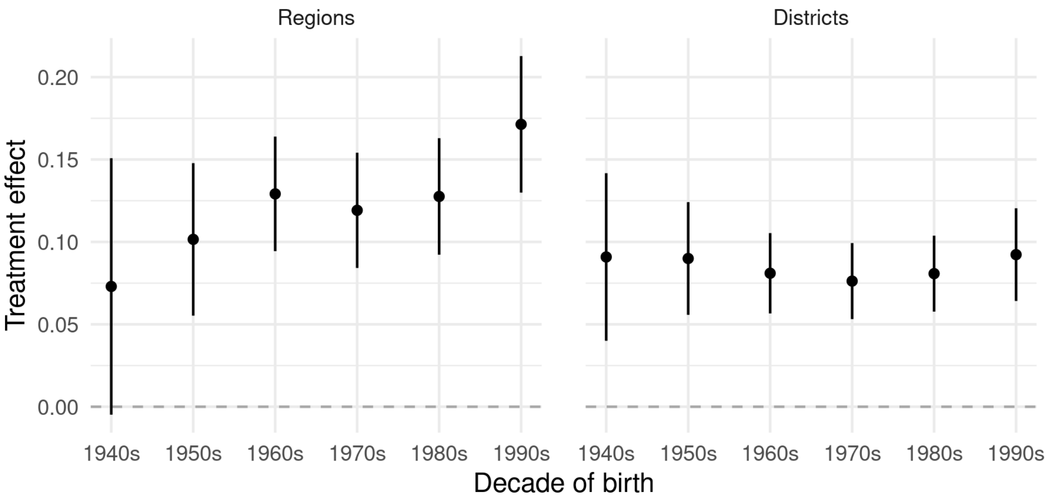


Figure 5: Border effect estimates by birth decade.
Note: Estimation follows the baseline specifications in Models 1 and 4 in Table 1.

origin are substantively similar to effects at newly drawn post-colonial borders. This again suggests that colonial and post-colonial developments contributed to the overall effects of borders on ethnic demography.

Robustness tests

This section presents the main robustness checks summarized in Figure 6 and discussed further in Appendix F. Results for models within ethnic groups and across straight borders coincide with the baseline specification discussed below.

Running variable: Closely examining Figure 2 we see a non-linearity in the outcomes very close (<5km) to administrative borders which likely stems from noise in the spatial attribution of survey clusters to administrative units. To examine however a conservative scenario in which these

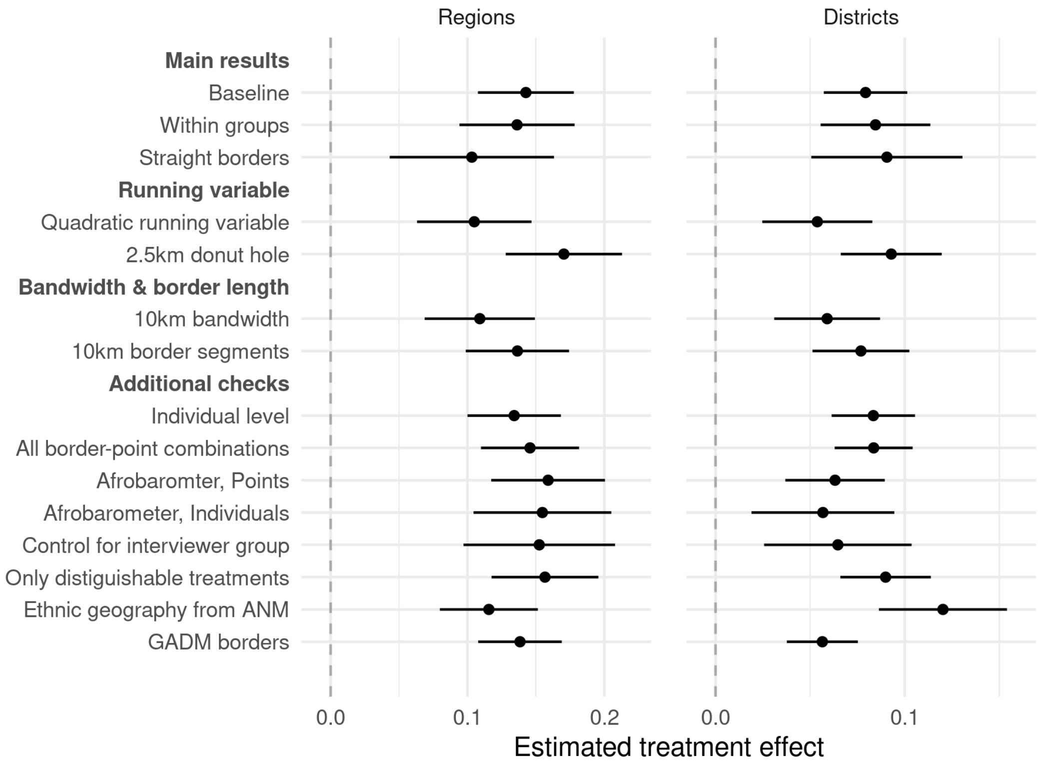


Figure 6: Summary of robustness checks.

non-linear dynamics are indeed real, I control for a linear and quadratic trend towards the border in treatment and control groups. This decreases the estimated effect of regional (district) borders by 4 (2) percentage points. I also examine the liberal scenario in which these deviations are only due to measurement error. I do so by estimating a 'donut'-RD, dropping all EAs closer than 2.5km[19] to the border. Doing so increases the estimated border effects by 2 to 3 percentage points. This difference shrinks with lower minimum distance cutoffs (Appendix F.2). In sum, the baseline results are well-centered between the conservative and liberal estimates.

Bandwidth and -breadth: The spatial setup of the RDD entails two other influential parameters, the first being the bandwidth of 20km. Appendix F.3 shows that results remain mostly stable when subsetting the data to EAs closer to the border. With a 10km bandwidth, effect estimates decrease on par with the quadratic specification discussed above. A second test avoids potentially undue influence of distant survey clusters located at opposite ends of a long border. To avoid such cases from driving the results, I define sub-border segments of a length of 10km. Limiting variation to such short (or shorter) border segments does not substantively change the results.

Additional robustness checks: Further testing the stability of the results, I find that analyzing individual- rather than EA-level DHS data or data

[19]This is half the displacement radius of rural DHS clusters.

from the Afrobarometer (2018) does not affect the results. The results are robust to controlling for social desirability bias where respondents want to appear co-ethnic to interviewers from the local plurality group. Dropping observations where DHS's ethnic categories do not distinguish between the two plurality groups across a border increases effect estimates.[20] I lastly re-implement the RDD with alternative ethnic settlement data from the *Atlas Narodov Mira* (ANM; Bruk and Apenchenko 1964) and administrative borders from GADM, and conduct a country-by-country jackknife (Appendix F.9). These tests show stable results.

The permutations of the research design evidence the robustness of the main results. The following analysis of assimilation and ethnic migration patterns further supports the theoretical argument by testing its micro-foundations.

Mechanisms: Assimilation and migration

The RDD estimates show discontinuous changes in ethnic demography at borders between administrative units with differing plurality groups. In the following, I test the argument that assimilation and ethnically biased migration patterns among minorities drive these results.

[20]In these 15 (25) percent of the region (district) sample, effects are 0 since the same group shares appear as outcomes in treatment and control conditions due to one-to-many links between DHS's and Murdock's groups.

Ethnic assimilation

Measuring individual-level assimilation in the presence of individual-level migration and absent panel data on individuals' (changing) ethnic identity is challenging. Assimilation can occur within one's own lifetime and over generations as frequently observed in immigrant populations (e.g., Fouka 2019).

Illuminating individual assimilation, the Afrobarometer enumerates respondents' *ethnic self-identification*, their main spoken *language*, and, in round 4, *all languages*, thus capturing frequent multilingualism (Buzasi 2016). Focusing on linguistic assimilation, we can test whether self-identified minority members speak the local plurality language as an important assimilation outcome (Cohen and Middleton 1970).

Ethnic assimilation is also fostered by marriage between local minority and majority members, which increases children's identity choice set (Cohen and Middleton 1970; Fouka 2019). Following Bandyopadhyay and Green (2021) who describe frequent interethnic marriages across Africa, I use DHS records of spouses' ethnic identities to measure whether married female respondents have a local plurality husband.[21]

For these three measures of assimilation, Table 2 implements the baseline RDD specification but adds a dummy variable for whether respondents' ethnic self-identification in the Afrobarometer or DHS data corresponds to

[21] Appendix Table A8 shows similar results for men and their first spouses.

the local plurality (Plur. group member) in interaction with the treatment dummy. The treatment dummy then captures the change of the respective outcomes among self-identified ethnic minorities at region and district borders.

Table 2: Minority assimilation to local plurality groups

	Regions			Districts		
	Speak lang.	Main lang.	Intermarr.	Speak lang.	Main lang.	Intermarr.
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.089*	0.066**	0.018**	0.068**	0.028	0.026***
	(0.049)	(0.029)	(0.008)	(0.034)	(0.020)	(0.007)
Plur. group member	0.430***	0.716***	0.745***	0.352***	0.724***	0.673***
	(0.075)	(0.045)	(0.015)	(0.063)	(0.022)	(0.012)
Treated × P.G.	−0.109	−0.021	0.014	−0.129*	−0.017	0.004
	(0.103)	(0.061)	(0.013)	(0.075)	(0.039)	(0.009)
Source	AB	AB	DHS	AB	AB	DHS
Cutoff	20km	20km	20km	20km	20km	20km
Running var linear	yes	yes	yes	yes	yes	yes
Survey FE:	yes	yes	yes	yes	yes	yes
Border FE:	yes	yes	yes	yes	yes	yes
Mean DV:	0.56	0.37	0.35	0.6	0.41	0.4
Borders:	97	250	246	125	489	743
Observations	16,088	96,608	44,680	18,220	127,038	70,056
Adjusted R^2	0.800	0.781	0.777	0.768	0.807	0.750

Notes: OLS linear models. The unit of analysis are individuals. The outcomes capture assimilation with administrative units plurality groups as indicated in the column headers: Speaking the plurality language, using the plurality language as one's main language, and being married to a plurality group member. The treatment coefficient captures the increase in assimilation among local minority group members at administrative borders. Standard errors clustered on the point and administrative unit × treatment levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01

I find generally significant and meaningful border effects on minorities' assimilation. The estimated effect on self-identified minorities speaking the plurality language amount to 8.9 and 6.8 percentage points at regional (Model 1) and district (Model 3) borders, respectively. I observe a slightly smaller increase in listing the plurality language as minorities' main language, with the district-level estimate being statistically insignificant. These estimates suggest that minorities tend to linguistically assimilate to units' predominant groups.[22] Models 3 and 6 find similar patterns of interethnic marriage. Female minority members chance of marrying a plurality group member increases by 1.8 (2.6) percentage points at regional (district) borders with units dominated by a different group. Additional analyses in Appendix G suggest that post-treatment change in the 'supply' of plurality men may explain this effect.

Part of the effect of administrative borders on plurality groups' population share thus likely works through minority assimilation. However, results in Table 2 cannot be *causally* interpreted as conditioning on 'plurality group membership' introduces post-treatment bias. The existence of former minority members that have assimilated and now fully self-identify with the plurality group biases estimated treatment effects downwards. The estimates thus likely constitute conservative estimates of ethnic assimilation among minority members. Migrants, a source of additional selection bias, are the subject of the following analysis.

Ethnic migration patterns

Ethnic sorting through migration constitutes the second mechanism behind the sharp decrease in the share of units' plurality group at administrative borders. Such sorting comprises (1) higher emigration rates of local minor-

[22]Additionally, Appendix Table A6 shows that plurality status increases ethnic vs. national identification.

ity members and (2) higher immigration rates of plurality members. Theoretically, border effects could also be driven by minority (plurality) members moving towards the interior (periphery) of a unit. Given data limitations, I only test for ethnic sorting through migration between administrative units.

To assess ethnically 'biased' subnational migration patterns, I rely on census data samples from eleven countries in Sub-Saharan Africa[23] provided by IPUMS (Minnesota Population Center 2018). The records contain the region of birth and residence of 33 million individuals. The data from Burkina Faso, Mali, Senegal, Sierra Leone, and Zambia additionally contain the same variables for districts. The information on birth and residence units allows me to derive the full lifetime migration matrix of the population enumerated in each census. To assess distinct migration patterns of local minority and plurality members, I draw on IPUMS' geographic data on administrative units and again take the group from Murdock's (1959) map with the largest spatial intersection as their plurality group. I draw on Müller-Crepon, Pengl and Bormann (2022) to link the latter to IPUMS' ethnic labels and differentiate between local plurality and minority groups in the censuses.

With the resulting data, I conduct three analyses (Table 3). I first estimate the extent to which local plurality status reduces individuals' emigra-

[23]Burkina Faso, Ghana, Liberia, Malawi, Mali, Rwanda, Senegal, Sierra Leone, South African, Uganda, and Zambia. See Table A1.

tion from their regions and districts of birth (Models 1 and 4). Regions' (districts') plurality group members show an emigration rate that is 12 (17) percentage points lower than that among local minorities. Second, I estimate the effect of local plurality status on immigration into migrants' co-ethnic regions and districts (Models 2 and 5). Again, the extent of such co-ethnic migration bias is substantive. Migrants move with a 6.3 (8.1) percentage points higher probability towards regions (districts) dominated by their ethnic kin than to other units. Both analyses account for fixed effects at the ethnic group and administrative unit levels.

Table 3: Ethnic migration patterns

	Share of migrants					
	Regions			Districts		
	Emigrants	Immigrants	Dyadic	Emigrants	Immigrants	Dyadic
	(1)	(2)	(3)	(4)	(5)	(6)
Eth. plur. source	-0.119^{***}		-0.007^{***}	-0.170^{***}		-0.003^{***}
	(0.018)		(0.002)	(0.020)		(0.001)
Eth. plur. target		0.063^{***}	0.022^{***}		0.081^{***}	0.007^{***}
		(0.013)	(0.004)		(0.012)	(0.001)
Unit of analysis	Source× group	Target× group	Dyad× group	Source× group	Target× group	Dyad× group
Group FE	yes	yes	yes	yes	yes	yes
Source FE:	yes	–	–	yes	–	–
Target FE:	–	yes	–	–	yes	–
Dyad FE:	–	–	yes	–	–	yes
Mean DV:	0.26	0.035	0.014	0.36	0.026	0.0063
Observations	6,519	7,255	169,632	9,942	15,161	628,983
Adjusted R^2	0.640	0.599	0.696	0.487	0.603	0.306

Notes: OLS linear models. Observations are weighted according to the number of individuals they include. Standard errors clustered on the migration source units in Models 1 and 4, target units in 2 and 5, and both in 3 and 6. Significance codes: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

The third analysis in Models 3 and 6 is fully dyadic, the unit of analysis being the ethnic group nested in directed birth to residence unit dyads. The outcome consists in the share of an ethnic group in a birth unit that has

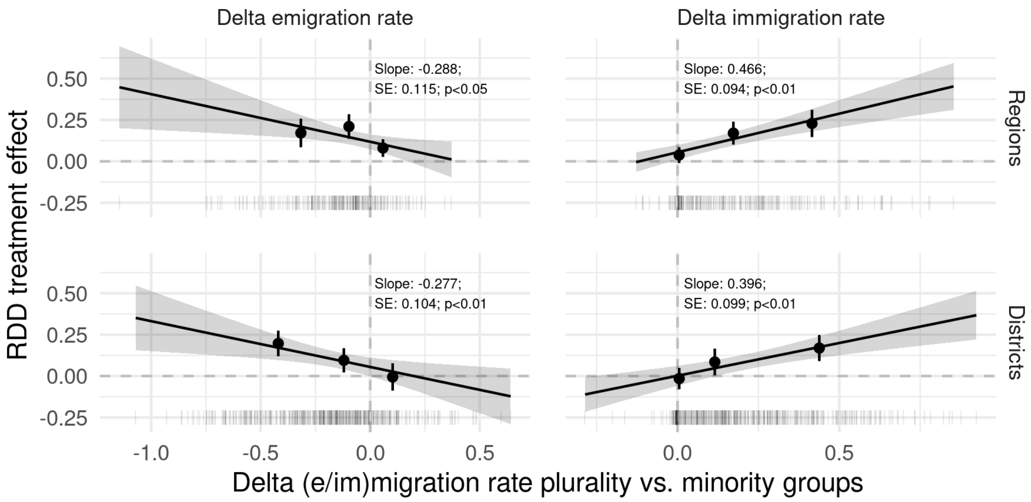


Figure 7: Effect of borders on ethnic identities increase in ethnic differences in e- and immigration.

Note: Results from four linear interaction models and by tercile of the moderator (Hainmueller, Mummolo and Xu 2019). Grey bars denote sample observations.

migrated towards a given residence unit. Controlling for dyad and ethnic group fixed effects, the models assess the degree to which plurality groups differentially move between administrative units. The average migration rate between two regions (districts) amounts to 1.4 (.63) percent of the population of the source unit. Belonging to the predominant group in one's birth region (district) decreases this rate by .7 (.3) percentage points. In turn, being a member of the plurality in the dyad's target unit increases it by 2.1 (.7) percentage points. In size similar to the average dyadic migration rate, these effects are substantive, robust to different specifications, and remain stable across birth cohorts going back to the 1900s (Appendix H.1).

Local minority members are thus more likely to exit, and migrants preferentially move to co-ethnic administrative units. But do such ethnic migration patterns explain the effects of administrative borders on ethnic demographics? To answer this question, I derive unit-level migration biases as the coefficients in Models 1 and 2 (4 and 5) estimated separately for each region (district) in the IPUMS data. I take the resulting unit-level measures and test whether they moderate the main RDD treatment effects from above in the 11 (5) countries for which I have region (district) level migration data.[24]

Figure 7 shows that the effect of administrative borders on plurality groups' population share strongly increases with (1) the extent of local plurality members lower emigration rate (*Delta emigration rate*) and (2) the migrants increased immigration rates into co-ethnic units (*Delta immigration rate*). For example, a decrease of plurality group members' differential emigration rate by 10 percentage points corresponds to a 3 percentage points larger border discontinuity of the plurality group share. Bearing in mind that this correlation between migration biases and ethnic discontinuities is not causally identified, this result nevertheless strongly suggests that ethnically biased migration contributed to the administrative borders effects on ethnic demography.

Conclusion

John Iliffe (1979) argued that "Europeans believed Africans belonged to tribes; Africans built tribes to belong to." This paper has analyzed the ef-

[24] I do so by adding them as an interaction term to Eq. 1. For consistency, I here couple the DHS data with IPUMS' border data to estimate the RDD.

fect of territorially bounded administrative units on ethnic demography as an important mechanism behind this argument. Colonialists devised administrative borders that frequently cut across ethnic geography, but local governance was nevertheless ethnicized on the bedrock of partly invented, partly preexisting 'customary' institutions. In turn, local minorities created by administrative boundaries reacted to their politically diminished status, often assimilating to the local majority or emigrating.

My analyses support this account of the administrative shaping of ethnic groups. I find sharp spatial discontinuities in ethnic demographics at administrative borders: the share of regions' (districts') predominant ethnic group locally increases by about 14 (8) percentage points or 54 (23) percent at borders to units dominated by a different group. Suggestive evidence shows that groups' plurality status has a substantively similar effect inside administrative units. Consistent with historical evidence, borders' effects increase with strong traditional institutions and larger dominant groups. I find that ethnic assimilation and ethnically biased migration patterns drive this phenomenon.

Taken together and acknowledging that the *global* effect of territorial governance on ethnicity cannot be empirically known absent a valid counterfactual, my argument and evidence offer an instrumentalist interpretation of constructivist accounts of the colonial transformation of ethnic identities in Africa. Ethnic identities and geographies are not prehistorically given but shaped by individual responses to ethnicized governance bounded by administrative borders. Once politicized, ethnic assimilation and migration patterns left ethnic identities crystallized along administrative borders, contributing to the alignment of administrative units and ethnic geographies.

While my results highlight the impact of administrative borders on ethnicity, other relevant, parallel, and potentially intersecting processes have shaped ethnic groups and individuals' identification with them. For example, states' nation-building efforts can likely counteract strong local identities rooted in subnational administrative units. Ethnic competition and conflict, in turn, may increase the salience of some ethnic differences but decrease that of others. And national-level institutions beyond the constitutionalization of traditional institutions analyzed here may affect the localization of ethnic identities. Complementary to my findings, these highlight the need to better understand the foundations of one of the most important political cleavages in Africa and beyond.

References

Afrobarometer. 2018. "Afrobarometer Data." *Available at http://www.afrobarometer.org* .

Alesina, Alberto and Eliana La Ferrara. 2005. "Ethnic diversity and economic performance." *Journal of economic literature* 43(3):762–800.

Alesina, Alberto and Enrico Spolaore. 2005. *The size of nations*. Cambridge, MA: MIT Press.

Alesina, Alberto, William Easterly and Janina Matuszeski. 2011. "Artificial states." *Journal of the European Economic Association* 9(2):246–277.

Ali, Merima, Odd-Helge Fjeldstad, Boqian Jiang and Abdulaziz B. Shifa. 2019. "Colonial Legacy, State-building and the Salience of Ethnicity in Sub-saharan Africa." *Economic Journal* 129(619):1048–1081.

Argyle, William John. 1969. European Nationalism and African Tribalism. In *Tradition and Transition in East Africa. Studies of the Tribal Element in the Modern Era.*, ed. P H Gulliver. Berkeley: University of California Press p. 41..58.

Asiwaju, Anthony I. 1970. The Aleketu of Ketu and the Onimek of Meko: The Changing Status of two Yoruba Rulers under French and British rule. In *West African Chiefs. Their Changing Status under Colonial Rule and Independence*, ed. Michael Crowder and Obaro Ikime. Ile-Ife: University of Ife Press pp. 134–161.

Asiwaju, Anthony I. 1976. "Migrations as Revolt: The Example of the Ivory Coast and the Upper Volta before 1945." *The Journal of African History* 17(04):577.

Asiwaju, Anthony I. 1983. "The Concept of Frontier in the Setting of States in Pre-colonial Africa." *Présence Africaine* 127:43.

Bandyopadhyay, Sanghamitra and Elliott Green. 2021. "Explaining interethnic marriage in Sub-Saharan Africa." *Journal of International Development* 33(4):627–643.

Barth, Fredrik. 1969. *Ethnic groups and boundaries: The social organization of culture difference*. Boston: Little, Brown.

Bates, Robert H. 1974. "Ethnic competition and modernization in contemporary Africa." *Comparative Political Studies* 6(4):457–484.

Boone, Catherine. 2014. *Property and political order in Africa: Land rights and the structure of politics*. Cambridge University Press.

Bormann, Nils-Christian. 2019. "Uncertainty, cleavages, and ethnic coalitions." *The Journal of Politics* 81(2):471–486.

Brass, Paul R. 1991. *Ethnicity and nationalism: theory and comparison*. Newbury Park, CA: Sage.

Bravman, Bill. 1998. *Making ethnic ways: communities and their transformations in Taita, Kenya, 1800-1950*. Social history of Africa Portsmouth, NH: Heinemann.

Brierley, Sarah. 2021. "Combining patronage and merit in public sector recruitment." *The Journal of Politics* 83(1):182–197.

Bruk, Solomon I. and V. S. Apenchenko. 1964. *Atlas Narodov Mira*. Moscow: Glavnoe upravlenie geodezii i kartografii gosudarstvennogo geologicheskogo komiteta SSSR and Institut etnografii im. H. H. Miklukho-Maklaia, Akademiia nauk SSSR.

Burgert, Clara R, Josh Colston, Thea Roy and Blake Zachary. 2013. "Geographic Displacement Procedure and Georeferenced Data Release Policy for the Demographic and health Surveys." *DHS Spatial Analysis Report 7* .

Buzasi, Katalin. 2016. "Linguistic situation in twenty sub-Saharan African countries: A survey-based approach." *African Studies* 75(3):358–380.

Cederman, Lars-Erik, Kristian Skrede Gleditsch and Halvard Buhaug. 2013. *Inequality, Grievances, and Civil War*. Cambridge: Cambridge University Press.

Chandra, Kachan. 2007. *Why ethnic parties succeed: Patronage and ethnic head counts in India*. Cambridge: Cambridge University Press.

Choi, Donghyun Danny, J Andrew Harris and Fiona Shen-Bayh. 2022. "Ethnic Bias in Judicial Decision Making: Evidence from Criminal Appeals in Kenya." *American Political Science Review* pp. 1–14.

Cohen, Ronald and John Middleton. 1970. *From Tribe to Nation in Africa: studies in incorporation processes*. Chandler Pub. Co.

Crowder, Michael. 1968. *West Africa under Colonial Rule*. London: Hutchinson.

De Luca, Giacomo, Roland Hodler, Paul A Raschky and Michele Valsecchi. 2018. "Ethnic favoritism: An axiom of politics?" *Journal of Development Economics* 132:115–129.

DHS. 2018. "Demographic and Health Surveys." *Integrated Demographic and Health Series (IDHS), version 2.0, Minnesota Population Center and ICF International. Available at http://idhsdata.org* .

Docquier, Frédéric and Hillel Rapoport. 2003. "Ethnic discrimination and the migration of skilled labor." *Journal of Development Economics* 70(1):159–172.

Eifert, Ben, Edward Miguel and Daniel N. Posner. 2010. "Political competition and ethnic identification in Africa." *American Journal of Political Science* 54(89):494–510.

Ejdemyr, Simon, Eric Kramon and Amanda Lea Robinson. 2018. "Segregation, Ethnic Favoritism, and the Strategic Targeting of Local Public Goods." *Comparative Political Studies* 51(9):1111 –1143.

FAO. 2014. "Global Administrative Unit Layers." *Available from: http://data.fao.org/map?entryId=f7e7adb0-88fd-11da-a88f-000d939bc5d8&tab=metadata* .

Fearon, James D. 1999. "Why ethnic politics and 'pork' tend to go together." *Unpublished Working Paper. Available from: https://web.stanford.edu/group/fearon-research/cgi-bin/wordpress/wp-content/uploads/2013/10/Pork.pdf* .

Fouka, Vasiliki. 2019. "Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I." *The Review of Economic Studies* 87(1):204–239.

Franck, Raphaël and Ilia Rainer. 2012. "Does the Leader's Ethnicity Matter? Ethnic Favoritism, Education, and Health in Sub-Saharan Africa." *American Political Science Review* 106(2):294–325.

Gravesen, Marie Ladekjær and Eric Mutisya Kioko. 2019. "Cooperation in the midst of violence: land deals and cattle raids in Narok and Laikipia, Kenya." *Africa* 89(3):562–585.

Green, Elliott D. 2008. "Understanding the limits to ethnic change: lessons from Uganda's "lost counties"." *Perspectives on Politics* 6(3):473–485.

Green, Elliott D. 2021. "The politics of ethnic identity in Sub-Saharan Africa." *Comparative Political Studies* 54(7):1197–1226.

Grossman, Guy and Janet I. Lewis. 2014. "Administrative Unit Proliferation." *American Political Science Review* 108(01):196–217.

Hainmueller, Jens, Jonathan Mummolo and Yiqing Xu. 2019. "How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice." *Political Analysis* 27(2):163–192.

Harris, J Andrew and Daniel N Posner. 2019. "(Under What Conditions) Do Politicians Reward Their Supporters? Evidence from Kenya's Constituencies Development Fund." *American Political Science Review* 113(1):123–139.

Hassan, Mai. 2017. "The strategic shuffle: Ethnic geography, the internal security apparatus, and elections in Kenya." *American Journal of Political Science* 61(2):382–395.

Henn, Sören. 2022. "Complements or Substitutes? How Institutional Arrangements Bind Chiefs and the State in Africa." *American Political Science Review, FirstView* .

Herbst, Jeffrey. 2000. *States and Power in Africa*. Princeton: Princeton University Press.

Holzinger, Katharina, Roos Haer, Axel Bayer, Daniela M. Behr and Clara Neupert-Wentz. 2019. "The Constitutionalization of Indigenous Group Rights, Traditional Political Institutions, and Customary Law." *Comparative Political Studies* 52(12):775–1809.

Honig, Lauren. 2017. "Selecting the state or choosing the chief? The political determinants of smallholder land titling." *World Development* 100:94–107.

Horowitz, Donald L. 1985. *Ethnic groups in conflict*. Berkeley: University of California Press.

Iliffe, John. 1979. *A modern history of Tanganyika*. Cambridge: Cambridge University Press.

Kalmijn, Matthijs. 1998. "Intermarriage and homogamy: Causes, patterns, trends." *Annual review of sociology* 24(1):395–421.

Kayira, Gift Wasambo, Paul Chiudza Banda and Amanda Lea Robinson. 2019. “Ethnic associations and politics in contemporary Malawi.” *Journal of Eastern African Studies* 13(4):718–738.

Keele, Luke J. and Rocío Titiunik. 2015. “Geographic Boundaries as Regression Discontinuities.” *Political Analysis* 23(1):127–155.

Laitin, David D. 1992. *Language repertoires and state construction in Africa*. Cambridge University Press.

Laitin, David D. 1995. “Marginality: a microperspective.” *Rationality and Society* 7(1):31–57.

Lentz, Carola. 1995. “‘Tribalism’and ethnicity in Africa.” *Cah. Sci. hum* 31(2):303–328.

Lentz, Carola. 2006. *Ethnicity and the Making of History in Northern Ghana*. Edinburgh: Edinburgh University Press.

Levi, Margaret. 1988. *Of rule and revenue*. Berkeley: University of California Press.

MacArthur, Julie. 2012. “The making and unmaking of African languages: oral communities and competitive linguistic work in western Kenya.” *The Journal of African History* 53(2):151–172.

MacArthur, Julie. 2013. “When did the Luyia (or any other group) become a tribe?” *Canadian Journal of African Studies/La Revue canadienne des études africaines* 47(3):351–363.

Mamdani, Mahmood. 2001. “Beyond settler and native as political identities: Overcoming the political legacy of colonialism.” *Comparative studies in Society and History* 43(4):651–664.

Mamdani, Mahmood. 2020. *Neither Settler nor Native*. Cambridge, MA: Harvard University Press.

McCauley, John F. and Daniel N. Posner. 2015. “African Borders as Sources of Natural Experiments Promise and Pitfalls.” *Political Science Research and Methods* 3(02):409–418.

McNamee, Lachlan. 2019. “Indirect colonial rule and the salience of ethnicity.” *World Development* 122:142–156.

Michalopoulos, Stelios. 2012. "The Origins of Ethnolinguistic Diversity." *American Economic Review* 102(4):1508–1539.

Minnesota Population Center. 2018. *Integrated Public Use Microdata Series, International: Version 7.1*. Minneapolis: University of Minnesota.

Müller-Crepon, Carl. 2020. "Continuity or Change: (In)direct Rule in British and French Colonial Africa." *International Organization* 74(4):707–741.

Müller-Crepon, Carl, Yannick Pengl and Nils-Christian Bormann. 2022. "Linking Ethnic Data from Arica." *Journal of Peace Research* 59(3):425–435.

Murdock, George Peter. 1959. *Africa. Its Peoples and Their Culture History.* New York: McGraw-Hill Book Company.

Nunn, Nathan and Leonard Wantchekon. 2011. "The slave trade and the origins of Mistrust in Africa." *American Economic Review* 101(7):3221–3252.

O'Leary, Brendan. 2001. "The elements of right-sizing and right-peopling the state." *Right-sizing the state: The politics of moving borders* pp. 15–73.

Pengl, Yannick, Philip Roessler and Valeria Rueda. 2021. "Cash Crops, Print Technologies and the Politicization of Ethnicity in Africa." *American Political Science Review, First View* .

Peterson, Derek R. 2018. Vernacular Language and Political Imagination. In *Tracing Language Movement in Africa*, ed. Ericka A Albaugh and Kathryn M de Luna. Ovford: Oxford University Press pp. 165—-186.

Posner, Daniel N. 2004. "The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi." *The American Political Science Review* 98(4):529–545.

Posner, Daniel N. 2005. *Institutions and Ethnic Politics in Africa*. Cambridge: Cambridge University Press.

Ranger, Terence. 1997. The invention of tradition in colonial Africa. In *Perspectives on Africa: A reader in culture, history, and representation*, ed. Roy Richard Grinker, Stephen C. Lubkemann and Christopher B. Steiner. Oxford: Wiley-Blackwell pp. 597–612.

Robinson, Amanda Lea. 2014. "National versus ethnic identification in Africa: Modernization, colonial legacy, and the origins of territorial nationalism." *World Politics* 66(4):709–746.

Robinson, Amanda Lea. 2017. "The Political Logic of Cultural Revival: Ethnic Visibility and Electoral Politics in Malawi." *Unpublished Working Paper. Available at: https://www.iast.fr/sites/default/files/IAST/conf/pol_science/robinson_paper.pdf* .

Schelling, Thomas C. 1971. "Dynamic models of segregation." *Journal of mathematical sociology* 1(2):143–186.

Schultz, Emily A. 1984. "From Pagan to Pullo: ethnic identity change in northern Cameroon." *Africa* 54(1):46–64.

Southall, Aidan W. 1970. "The illusion of tribe." *Journal of Asian and African Studies* 5(1-2):28–50.

Stahl, Ann B. 1991. "Ethnic style and ethnic boundaries: A diachronic case study from West-Central Ghana." *Ethnohistory* pp. 250–275.

van Hoorn, André and Robbert Rademakers. 2021. "Decentralization and Local Ethnic Favoritism: Micro-Level Panel Evidence from Indonesia." *Available at SSRN 3822945* .

Vaughan, Olufemi. 2003. "Chieftaincy politics and communal identity in western Nigeria, 1893–1951." *The Journal of African History* 44(2):283–302.

Weber, Eugen. 1977. *Peasant into Frenchmen. The Modernization of Rural France 1870-1914*. London: Chatto & Windus.

Weber, Max. 1919. *Politik als Beruf*. Berlin: Duncker & Humblot.

Wimmer, Andreas. 2013. *Ethnic boundary making: Institutions, power, networks*. Oxford: Oxford University Press.

Young, Crawford. 1985. Ethnicity and the Colonial and Postcolonial State in Africa. In *Ethnic Groups and the State*, ed. Paul R. Brass. London: Croom Helm pp. 57–93.

Zhang, Nan and Melissa M Lee. 2020. "Literacy and State–Society Interactions in Nineteenth-Century France." *American Journal of Political Science* 64(4):1001–1016.

Online Appendix

Building Tribes: How Administrative Units Shaped Ethnic Groups in Africa

Table of Contents

A Survey data on local ethnic favoritism

Contemporary Afrobarometer (2018) survey data offer descriptive evidence on the political disenfranchisement of local ethnic minorities. Figure A1 shows that survey respondents who are members of the largest ethnic group within their region or district report having more trust in and interaction with local governance institutions than minority members do. In particular, plurality members tend to perceive their local authorities as more responsive, and hence trust in and approve of them more. Consistent with the argument that traditional authorities are particularly ethnicized, these associations are stronger with respect to traditional authorities as compared to local governments.

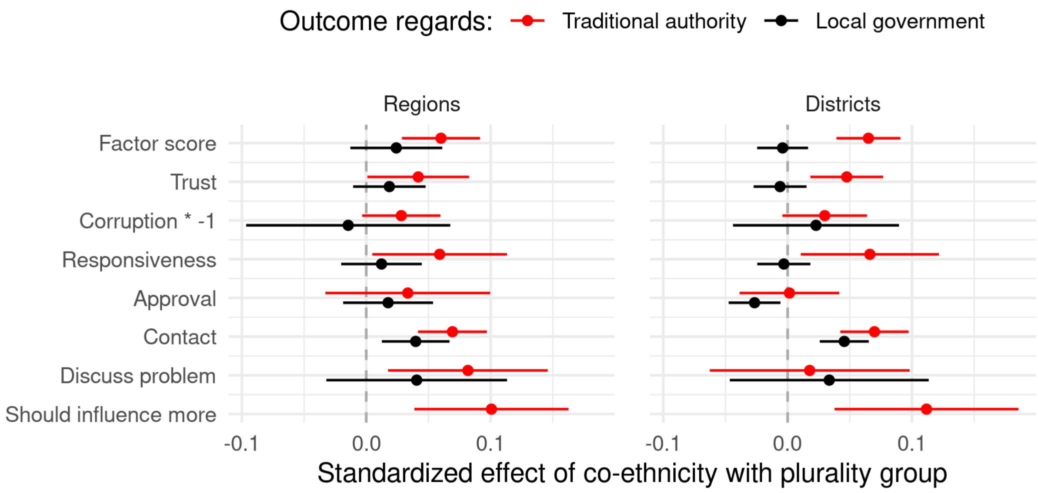


Figure A1: Increased interaction between local plurality members and traditional authorities as well as local governments.

Note: Estimates result from linear regressions of the variables indicated on the y-axis on a dummy that captures whether an Afrobarometer respondent is a member of the local plurality group derived from Murdock's map, individual-level covariates, and administrative unit as well as language group fixed effects. The 'Factor score' is the principal component of the separate, multiply imputed, and standardized survey items.

B Data Appendix

B.1 Data summary

Table A1: Samples across data sources

Country	Col. units	DHS	Afrobarometer	IPUMS
Benin	yes	3.1, 4.1, 6.1	3, 4, 5, 6	
Botswana			1, 2, 3, 4, 5, 6	
Burkina Faso	yes	2.1, 3.1, 4.1, 6.1, 7.1	4, 5, 6	2006
Burundi			5, 6	
Cameroon		2.1, 4.1, 6.1	5, 6	
Chad	yes	7.1		
Côte d'Ivoire	yes	3.1, 3.2, 6.1	5, 6	
Eswatini		5.1	5, 6	
Ethiopia	yes	4.1, 5.1, 6.1, 7.1		
Gabon	yes	6.1	6	
Ghana	yes	3.1, 4.1, 4.2, 5.2, 7.1, 7.2	1, 2, 3, 4, 5, 6	2000, 2010
Guinea	yes	4.1, 5.1, 6.1	5, 6	
Kenya	yes	4.1, 5.1, 7.1, 7.2	2, 3, 4, 5, 6	
Lesotho		4.1, 6.1, 7.1	1, 2, 3, 4, 5, 6	
Liberia	yes	0.1, 5.1, 5.2, 6.1, 6.2, 7.1	4, 5, 6	1974, 2008
Madagascar			3, 4, 5, 6	
Malawi	yes	4.1, 4.2, 6.1, 6.2, 7.1, 7.2, 7.3	1, 2, 3, 4, 5, 6	2008
Mali	yes	3.1, 4.1, 5.1, 6.2, 7.1	1, 2, 3, 4, 5, 6	1987, 1998, 2009
Mozambique	yes	5.1, 6.1, 7.1	2, 3, 4, 5, 6	
Namibia	yes	4.1, 5.1, 6.1	1, 2, 3, 4, 5, 6	
Niger	yes	2.1, 3.1	5, 6	
Nigeria	yes	2.1, 4.2, 5.1, 6.1, 6.2, 7.1	1, 2, 3, 4, 5, 6	
Rwanda				1991, 2002
Senegal	yes	2.1, 3.1, 4.2, 5.2, 6.1, 6.2, 7.2, 7.3	2, 3, 4, 5, 6	2002
Sierra Leone	yes	5.1, 6.1, 7.1	5, 6	2004
South Africa			1, 2, 3, 4, 5, 6	2001, 2011
Tanzania			1, 2, 3, 4, 5, 6	
Togo	yes	0.1, 3.1, 6.1	5, 6	
Uganda	yes	4.1, 5.1, 5.2, 6.1, 6.2, 7.1, 7.2	1, 2, 3, 4, 5, 6	1991, 2002
Zambia	yes	5.1, 6.1	1, 2, 3, 4, 5, 6	1990, 2000, 2010
Zimbabwe			1, 2, 3, 4, 5, 6	

Notes: Some survey rounds (e.g., those from Lesotho, Swaziland, Botswana, Gabon, and Burundi) do not contribute variation to the RDD estimates since they lack observations at administrative borders with differing plurality groups at either side. The table only lists DHS and Afrobarometer rounds as well as IPUMS census data with ethnic information.

B.2 Fractal dimension computation

The fractal dimension of a spatial line can be quantified via the boxcount method as $D = \lim_{\epsilon \to \infty} log(N(\epsilon))/log(1/\epsilon)$ where ϵ is the resolution of a square grid and $N(\epsilon)$ is the number of grid cells covered by a line. Straight lines have a fractal dimension of 1 and lines that cover the plane approach a value of 2. In practice, a number of parameters matters for the estimation of $N(\epsilon)$, in particular the orientation of a

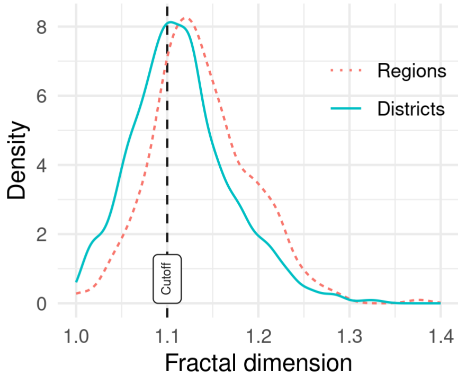


Figure A2: Distribution of fractal dimension values among borders in the baseline analysis. Each border receives equal weight.

line and the alignment of the square grid it is intersected with. In order to achieve consistent results, I implement the following algorithm:

1. Turn each line around the centroid of its bounding circle so that we minimize number of intersecting grid cells of a square grid (x and y resolution ϵ of 1/513 cells) that covers its bounding circle.
2. Define a series of square grids $g \in G$ with a x and y resolution ϵ_g of $1/(2^{\{1,\ldots,8\}} + 1)$ that share the same centroid as the lines bounding circle and a bounding box that touches the line at (at least) one point.
3. Count the number of grid cells $N(\epsilon_g)$ that the line intersects with in each grid $g \in G$.
4. Compute the fractal dimension of the line as the coefficient $\beta_1$of the regression $log(N(\epsilon_g) = \beta_1 log(1/\epsilon_g)$. Note that this regression does not include an intercept since each line intersects the singe cell in grids with $\epsilon_g = 1$ by definition. Therefore, the regression line must go through the origin to be valid.[25]

C Regression discontinuity analysis: Descriptive statistics

[25]Practically, the inclusion of an intercept produces inconsistent fractal dimension estimates which at times take theoretically impossible values lower than 1.

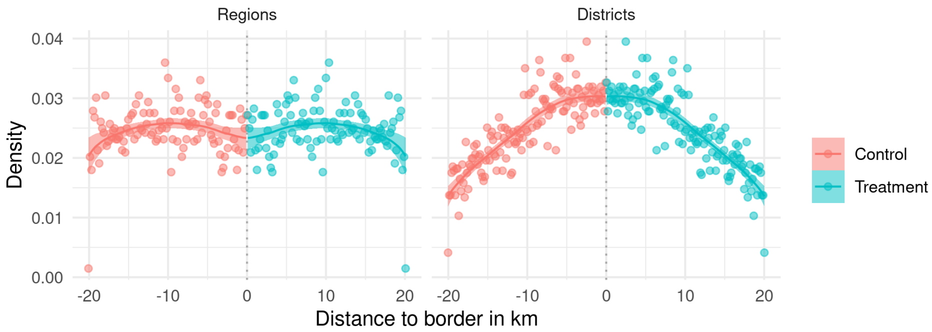


Figure A3: McCrary (2008) test.

Note: Because each observation is part of the treatment and control groups, the distribution of the running variable is perfectly symmetric across the threshold.

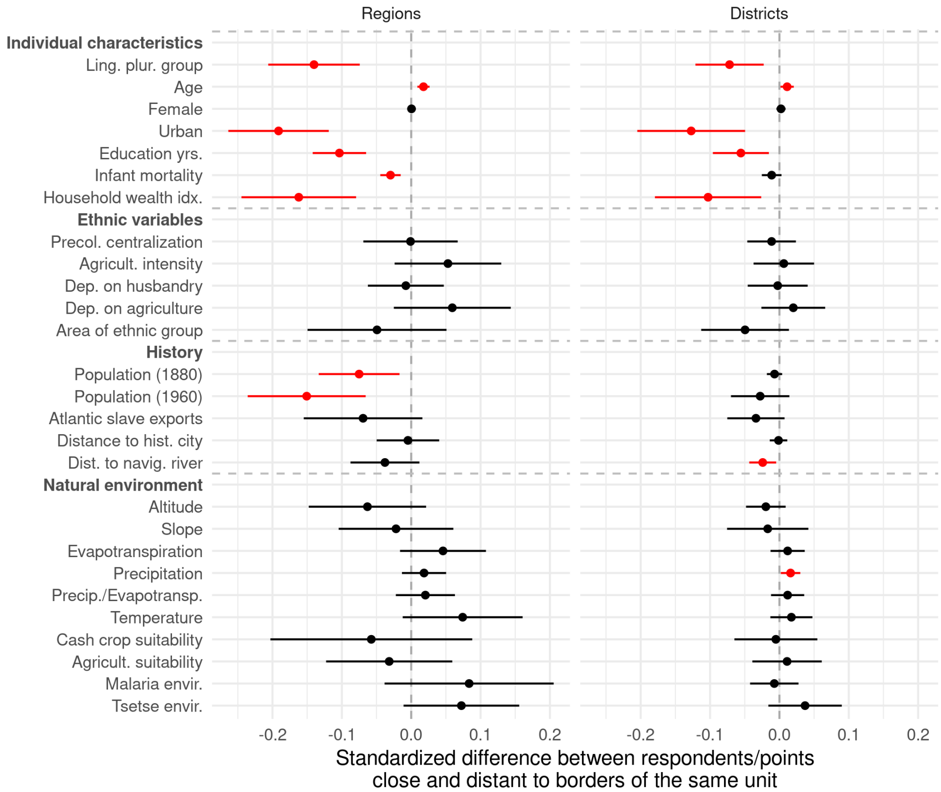


Figure A4: Comparison of DHS respondents and survey clusters close (<20km) to borders that divide units with differing plurality ethnic groups with average observations from the same units.

Note: Estimates with a p-value of $< .05$ in red. Estimated based on a simple regression of dependent variables (y-axis) on a dummy of closer than 20km to a border with differing plurality groups on either side and administrative unit $\times$ survey fixed effects. The first set of 'individual level characteristics' is based on individual-level data, while the remaining estimates are based on EA-level aggregates from the DHS.

D Regression discontinuity analysis: Heterogeneity

D.1 Fractal dimension and other measures of border arbitrariness

In order to gauge the effect of the particular and somewhat arbitrary fractal dimension cutoff of 1.1 used in the main analysis, this robustness check introduces alternative cutoff values and measures borders' arbitrariness via their alignment with rivers and watersheds. First, Figure A5 re-estimates the 'straight-border' analysis with cutoffs for the maximum fractal dimension of borders varying between 1.025 and the maximum value of 1.375 observed in the data. the results indicated that effect sizes, if at all, increase with straighter borders, thus supporting the claim that potentially endogenous border drawing does not drive the results.

As an alternative measure of borders' arbitrariness, I also measure the extent borders align with rivers and watershed. For each border, I measure the fraction of points on the borders that is within 5km of a major watershed and river.[26] I then split the resulting alignment variables, as well as, for comparative purposes, the fractal dimension values, into quartiles, with higher quartiles assigned to more natural borders, i.e. squiggly lines that align with rivers or watersheds. Finally, I re-estimate the baseline model for each quartile separately, resulting in the estimates presented in Figure A6. The results show no clear correlation of the naturalness of borders with their effect on ethnic identities – effect sizes decrease with greater alignment with watersheds but remain stable with greater river alignment. In sum, the results suggest that the baseline estimates are unlikely caused by endogenous border drawings.

Lastly, I test whether treatment effects vary with the overall alignment of administrative geographies with ethnic settlement patterns as an additional test of potential reverse causality. To that intent, I first measure for each country the alignment between regions/districts and Murdock's map of ethnic groups using a Mutual Information metric that captures the amount of information the administrative

[26] Data on rivers comes from the Natural Earth data: https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-rivers-lake-centerlines/. Data on watersheds comes from from Lehner, Verdin and Jarvis (2008), including all watersheds above a level of 4 on the Pfaffstetter coding system.

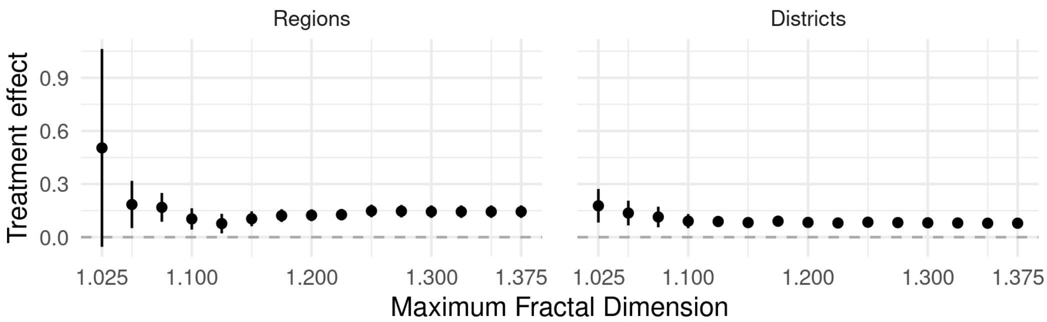


Figure A5: Estimate of border effect on local ethnic identities with different fractal dimension cutoffs.

Note: Specifications are the same as in baseline Models 3 and 6 in Table 1 with the exception of the cutoff of the fractal dimension of borders used (as indicated on the x-axis).

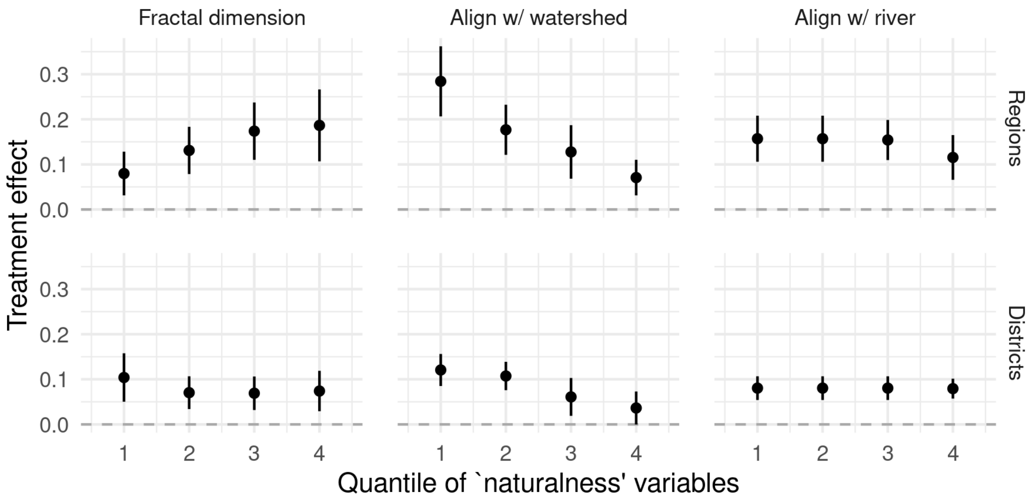


Figure A6: Estimate of border effect on local ethnic identities by degree of naturalness of border.

Note: Split sample estimates. Specifications are the same as in baseline Models 1 and 4 in Table 1. Samples are split along the quartiles of each variable indicated in the rows of the plot to the left.

partitioning of a country contains on its ethnic partitioning (and vice-versa) (Vinh, Epps and Bailey 2010). I compute the normalized mutual information (MI) metric on the basis of the centroids of the cells of a grid with .0833 decimal degrees resolution (ca. 10km at the equator). For each country, I encode the vectors of centroids' administrative unit membership A and ethnic settlement membership B and then compute MI as:

$$MI(A, B) = H(A) - H(A|B) \tag{2}$$

$$MI_{norm}(A, B) = \frac{MI(A, B)}{(E\{A\} * E\{B\})^{.5}} \tag{3}$$

where $H(A)$ and $H(A|B)$ are the (conditional) entropies of the administrative group memberships A. MI returns the quantity of information in A on B in bits. In turn, MI_{norm} adjusts that information with the entropy of partitionings A and B, yielding a measure that varies between 0 (no mutual information) and 1 (full mutual information) and is comparable across countries and types of administrative units.

Following the same approach as above, I use the MI_{norm} metric to split the data into quartiles and re-estimate the main specification for each. The results in Figure A7 show no clear relationship between MI_{norm} and treatment effect estimates which are not driven by countries in which administrative units are either strongly or weakly aligned with ethnic settlement patterns. The main treatment effects are substantively and statistically significant in each quantile and do not differ systematically in size.

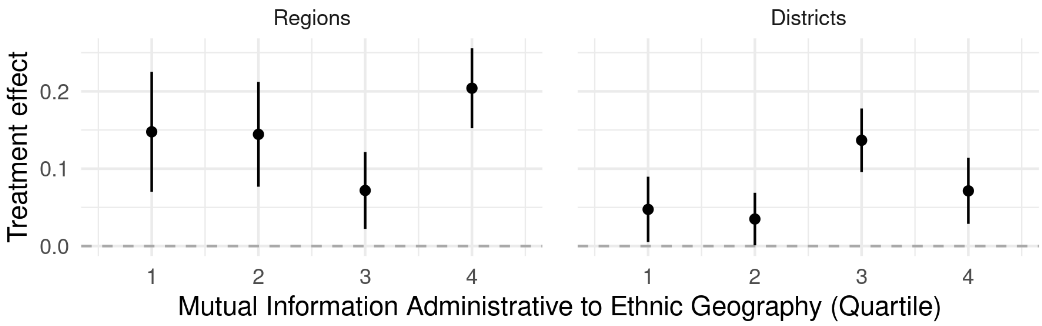


Figure A7: Estimate of border effect on local ethnic identities by Mutual Information between countries administrative geography and their ethnic settlement patterns.

Note: Split sample estimates. Specifications are the same as in baseline Models 1 and 4 in Table 1. Samples are split along the quartiles of MI_{norm}.

D.2 Distinguishing between regional and district borders

I here distinguish between pure district borders and those which are at the same time also regional borders. Table A2 shows that the bulk of the district-level effects are driven by borders that are at the same time regional borders (Models 1-3). However, I find substantial effects along pure district borders as well, effects that amount to 30 to 50% of the size of those associated with regional borders (Models 4-6). This suggests that effect sizes mirror the hierarchy hat structures territorial governance units.

Table A2: Robustness check: Distinguishing pure district borders

	Outcome: Plurality group share (0-1)					
	Aligned w/ region border			Not aligned w/ region border		
	Base	W/in grps.	Frac. dim.	Base	W/in grps.	Frac. dim.
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.144***	0.127***	0.143***	0.043***	0.064***	0.046*
	(0.019)	(0.024)	(0.028)	(0.013)	(0.018)	(0.025)
Cutoff	20km	20km	20km	20km	20km	20km
Max fractal dimension	2	2	1.1	2	2	1.1
Running var linear	yes	yes	yes	yes	yes	yes
Border + survey FE:	yes	–	yes	yes	–	yes
Group-Border + survey FE:	no	yes	no	no	yes	no
Mean DV:	0.36	0.35	0.35	0.4	0.42	0.38
Borders:	482	539	241	655	853	305
Observations	10,604	5,178	4,472	16,218	9,350	5,420
Adjusted R^2	0.616	0.637	0.603	0.678	0.729	0.694

Notes: OLS linear models. Standard errors clustered on the point and administrative unit × treatment levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01

D.3 Heterogeneity by former colonial power

Adding to the analysis of heterogeneous treatment effects along degrees of power of traditional institutions, I here gauge whether the effects vary significantly between former colonizers, in particular the British, who ruled more indirectly, and the French, who tended to establish more direct forms of colonial rule. To do so, I split the sample between the two former colonial powers and re-estimate the main RDD specification. Given its stronger indirect rule, it is not surprisingly the former British colonies show larger effects than former French colonies (Table A3. While of substantive size (ca. 5 to 6 percentage points), the difference between effects in former French and British colonies are are only statistically significant at $p < .1$.

Table A3: Former British vs. French colonial territories

	Outcome: Plurality group share (0-1)					
	Regions			Districts		
	British	French	Both	British	French	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.171***	0.119***	0.174***	0.087***	0.050**	0.091***
	(0.026)	(0.031)	(0.025)	(0.015)	(0.020)	(0.015)
Treated x French			−0.066*			−0.048*
			(0.040)			(0.027)
Cutoff	20km	20km	20km	20km	20km	20km
Max fractal dimension	2	2	2	2	2	2
Running var linear	yes	yes	yes	yes	yes	yes
Survey FE:	yes	yes	yes	yes	yes	yes
Border FE:	yes	yes	yes	yes	yes	yes
Mean DV:	0.34	0.36	0.34	0.4	0.38	0.39
Borders:	142	139	283	498	373	876
Observations	9,268	4,140	13,728	12,586	7,454	20,148
Adjusted R^2	0.635	0.520	0.594	0.664	0.655	0.656

Notes: OLS linear models following Eq. 1. The unit of analysis is the survey cluster. The outcome is the share of respondents in a cluster from the treatment unit's ethnic plurality group. The treatment coefficient captures the increase in the share of administrative units' plurality groups at their borders. Standard errors clustered on the EA and administrative unit × treatment levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01

D.4 Heterogeneity by national ethno-political status and ethnic civil war

In addition to heterogeneity of treatment effects along the dimensions discussed in the main text, I analyse how the *national-level* political status of plurality groups affects treatment effects. While plurality groups past access to national power does not moderate effects, their historical involvement in ethnic civil wars increases treatment effects.

The pattern of effect heterogeneity along access to national power is theoretically unclear. It could be expected from the ethnic favoritism literature (e.g. De Luca et al. 2018; Franck and Rainer 2012) that access to national power increases the resources and favoritism of coethnic local governments and thus observed treatment effects. Yet, three mechanisms cut against this expectation. First, not all discrimination of local minorities depends on resources from the central government, as for example, exclusionary land rights or the provision of primary education in the local vernacular. Second, ethnic favoritism is often realized through local public goods, thus also benefiting local minorities (Ejdemyr, Kramon and Robinson 2018; Harris and Posner 2019). In turn, governments' coethnics benefit *individually* from favoritism where they are a local minority (Beiser-McGrath, Müller-Crepon and Pengl 2021), thus decreasing their incentives to assimilate to the local majority in control units. This decreases estimated treatment effects. Third, where larger local budget increase absolute levels of ethnic favoritism, the *relative* gap between the local plurality group and minorities my decrease if the marginal value of additional ethnic favoritism decreases with larger budgets. To the extent that individuals are incentivized by their relative status, this may again dampen a positive relationship between larger local budgets and greater incentives for minorities to assimilate or migrate.

Table A4: Heterogeneity by inclusion into national exec. power and ethnic conflict

	Outcome: Plurality group share (0-1)					
	Regions			Districts		
	Base	W/in grps.	Frac. dim.	Base	W/in grps.	Frac. dim.
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.157***	0.122***	0.140***	0.089***	0.070***	0.079***
	(0.024)	(0.020)	(0.026)	(0.016)	(0.012)	(0.016)
Treated × Avg. Inclusion (0-1)	−0.044		−0.031	−0.023		−0.016
	(0.038)		(0.037)	(0.023)		(0.023)
Treated × Any Conflict (0/1)		0.099**	0.091*		0.069*	0.068*
		(0.050)	(0.051)		(0.038)	(0.038)
Cutoff	20km	20km	20km	20km	20km	20km
Max fractal dimension	2	2	2	2	2	2
Running var linear	yes	yes	yes	yes	yes	yes
Survey FE:	yes	yes	yes	yes	yes	yes
Border FE:	yes	yes	yes	yes	yes	yes
Mean DV:	0.33	0.33	0.33	0.39	0.39	0.39
Observations	13,549	13,549	13,549	21,213	21,213	21,213
Adjusted R^2	0.608	0.610	0.610	0.650	0.651	0.651

Notes: OLS linear models. Constitutive terms and interactions with distance to border dropped from the table. Standard errors clustered on the point and administrative unit × treatment levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01

In turn, the occurrence of ethnic conflict might increase treatment effects as rebel groups and the government may (violently) discriminate against the co-ethnics of their opponents in the territories they hold, leading to flight and ethnic passing and assimilation where this is possible. If patterns of territorial control align with administrative borders, this would increase observed treatment effects in units with a plurality group that has been mobilized for ethnic civil war in the past.

In order to probe these arguments, I draw on the Ethnic Power Relations data and merge it with administrative units' plurality groups via data from Müller-Crepon, Pengl and Bormann (2022). I then code (1) the share of years since independence plurality groups were included in the national executive and (2) create a dummy for whether groups were part of an ethnic constituency mobilized in a post-colonial ethnic conflict. The results of estimating simple interaction effects (Table A4) show no heterogeneous treatment effects among nationally powerful plurality groups. In turn, units with a plurality group with a history of ethnic civil war do show larger treatment effects ($p < .1$). Yet, since that group is relatively small (11% of the sample), treatment effects are very similar to the baseline estimates for groups without a history of ethnic conflict.

E Regression discontinuity analysis: Effect timing

To test whether borders of colonial or post-colonial origin drive the effects, I systematically compare the alignment between both sets of borders and analyse treatment effects in the resulting (non-)aligning subsets. While the post-colonial border data from 1990 remains the same, the colonial-era administrative borders comprise regional borders from countries' independence (Müller-Crepon 2021) and district borders from Huillery (2009) and Müller-Crepon (2020) from French and British colonies. I measure the alignment between post-colonial region and district borders with their colonial counterparts and vice-versa. I classify a border as "aligned" if more than 90% of points sampled along it (every 500m) lie within 5km of one of the target borders. Note that such alignments can go along with unit splits and mergers that left a border unchanged. A border is not "not aligned" if less than 50% of its points lie within 5km of the target borders. Among colonial regional (district) borders, 92% (83%) align with 1990 district borders and 4% (6%) do not, i.e. they disappeared. Among 1990 regional (district) borders, 62% (57%) align with colonial district borders and 26% (36%) do not, i.e. they were created post-colonially.

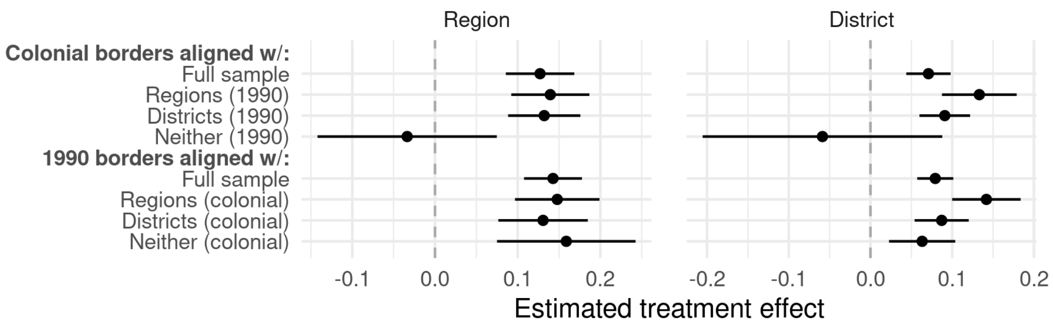


Figure A8: Main treatment effect by intertemporal border alignment.

I first test whether my results are robust to the use of colonial border data. I distinguish between the full set of colonial borders, those that continued their existence as regional or district borders in 1990, and those that disappeared. The results in Figure A8 show that effects are smaller at the level of colonial regions (but not districts) than at baseline. Beyond attenuation bias from noise in the manually digitized border data, this difference is driven by regional borders that disappeared, which entail no substantive treatment effects. Lastly, colonial regions tended to be larger and more diverse than those in 1990, further reducing treatment effects. I find broadly similar effects of colonial regional borders that survived as regional or district borders until 1990. Consistent with the main results, colonial district borders have slightly larger effects if they align with 1990 regional borders.

I next test whether newly created post-colonial borders drive the main treatment effects. To that intent, I divide administrative borders observed in 1990 into those that align with borders of colonial regions and districts, and newly created ones that align with neither.[27] Evidencing consistent effects throughout the colonial

[27]Note that I have no data on colonial units below the district level.

and post-colonial periods, Figure A8 shows similar effects among all three border types, with substantially larger effects seen along district borders that align with colonial regions. The analysis also suggests that endogenous border changes do not drive the results, since we would otherwise expect newly created post-colonial borders to dominate them.

F Regression discontinuity analysis: Robustness checks

This section presents the full results of the robustness checks to the regression discontinuity design discussed in the main body of the paper. By way of summary, Figure A9 combines the coefficients from most analyses discussed below.

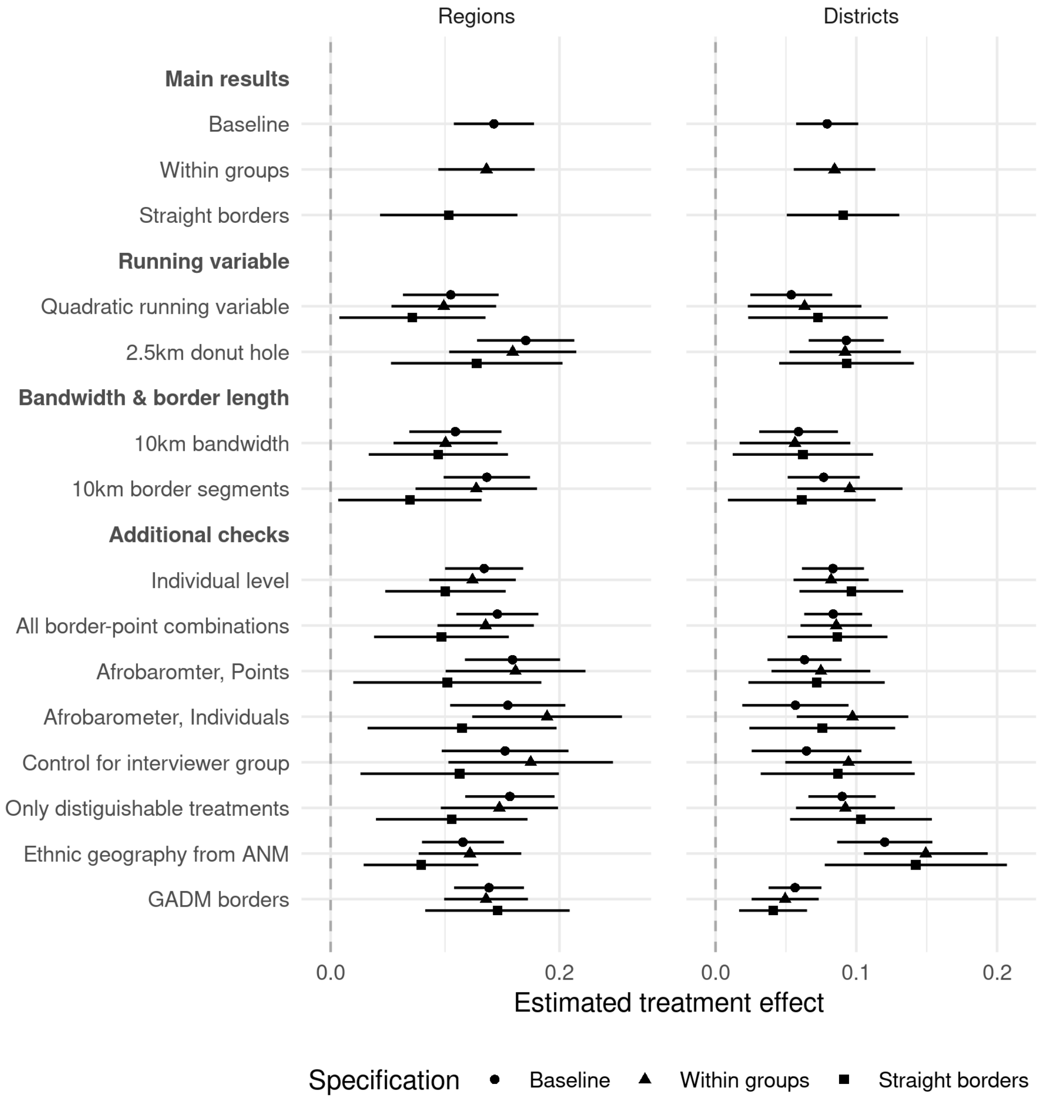


Figure A9: Summary of results from varying robustness checks.

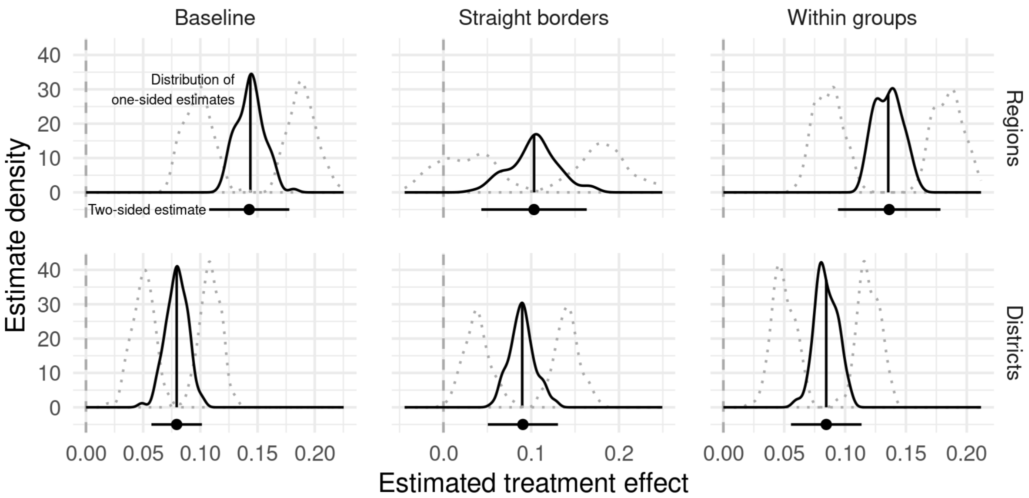


Figure A10: Estimate distributions from one-sided design

Note: Treatment and control status are randomly assigned to the two sides of each border. Distributions (with mean) of estimates (with upper/lower 95% CI bounds in dotted grey) result from re-estimating the main specifications for 100 sets of random assignment.

F.1 Comparison with one-sided RD Design

In the two-sided RD Design each EA enters the data in the control and treatment group (with its two corresponding outcomes). I here test whether this strategy leads to estimates that are equivalent but more precise than those resulting from a one-sided design, where I randomly assign treatment/control status either side of a border. Because there are many different assignment constellations across the many borders in the data, I repeat this random assignment 100 times re-estimating the main specifications each time. This produces the estimate distributions in Figure A10. The results show that the two-sided treatment effect estimates correspond exactly to the average one-sided estimate but is estimated more precisely due to the increased statistical power. However, virtually all one-sided estimates are, with the exception of the one along the comparatively few straight regional borders, highly statistically significant as well.

F.2 Running variable

Because a visual inspection of the data indicates a non-linear trend very close to the examined borders (below 5km), Figure A9 above shows robustness of the results with respect to the specification of the running variable, the distance to the border. As noted in the main discussion, this non-linear geographic trend most likely stems from noise in the spatial attribution of treatment and control units to the survey clusters. But it might, of course, also reflect real patterns. In this case, the main results would be biased. To assess the conservative scenario in which these non-linear dynamics are not caused by noise, I control for a linear and quadratic trend towards the border in treatment and control groups. While the results show a smaller effect of regional (district) border than estimated in the main specification, the estimates remain statistically significant.

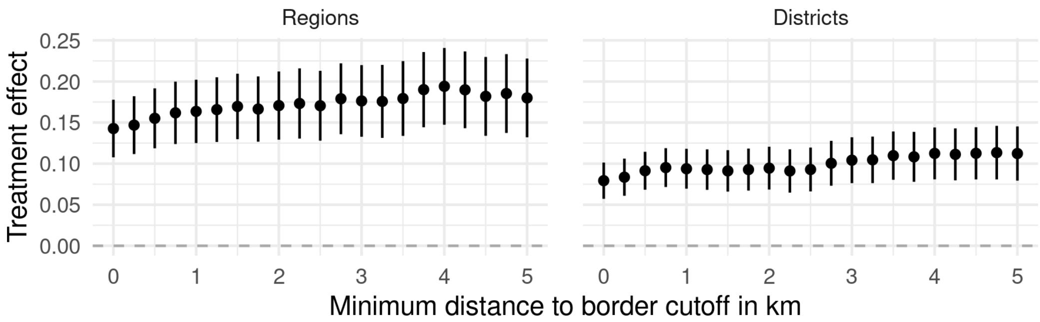


Figure A11: Estimate of border effect on local ethnic identities by size of 'donut'-hole around the border.

Note: Main specification with varying minimum border distance of observations (x-axis).

We can also examine the opposite, less conservative assumption that deviations from the linear distance trend close to the border reflect pure noise in the match of administrative units with DHS clusters. Doing so results in estimating a 'donut'-RDD in which we drop all EAs closer than some threshold x km, with $x \in [0, 5]$km, to the border. Varying this threshold x between 0 and 5 km, Figure A11 shows that the estimated effect size increases as the "donut-hole" around the border becomes larger, i.e., as we assume larger shares of the data around the border to be geographical misattributed to treatment/control units.

F.3 Variation in bandwidth and border segment length

The main analysis relies on variation among survey clusters within 20km to border segments defined along the full length of a border between two administrative units. The following analyses test the robustness of the results along two dimensions of this choice, the bandwidth (maximum distance of clusters to the border) and the length of border segments.

As plotted in Figure A12, the estimated effects slightly grow as we increase the bandwidth above 20km. As one shrinks the bandwidth, estimates become smaller but remain significant and substantively meaningful up to a small bandwidth of

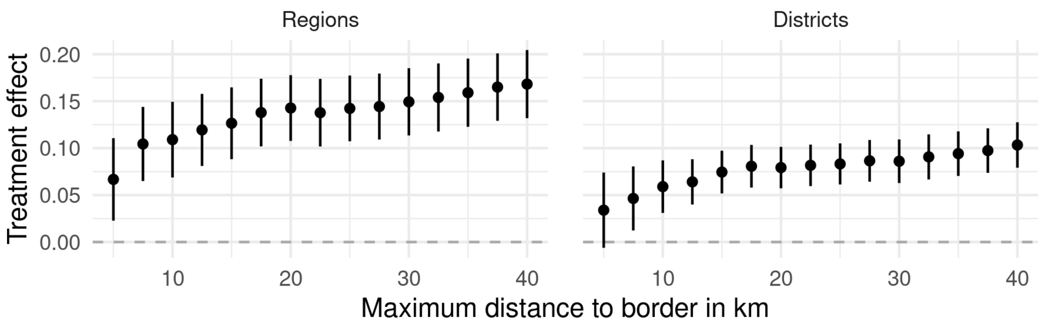


Figure A12: Estimate of border effect on local ethnic identities by bandwidth.

Note: For other specification details, see also baseline Models 1 and 4 in Table 1 in the main text.

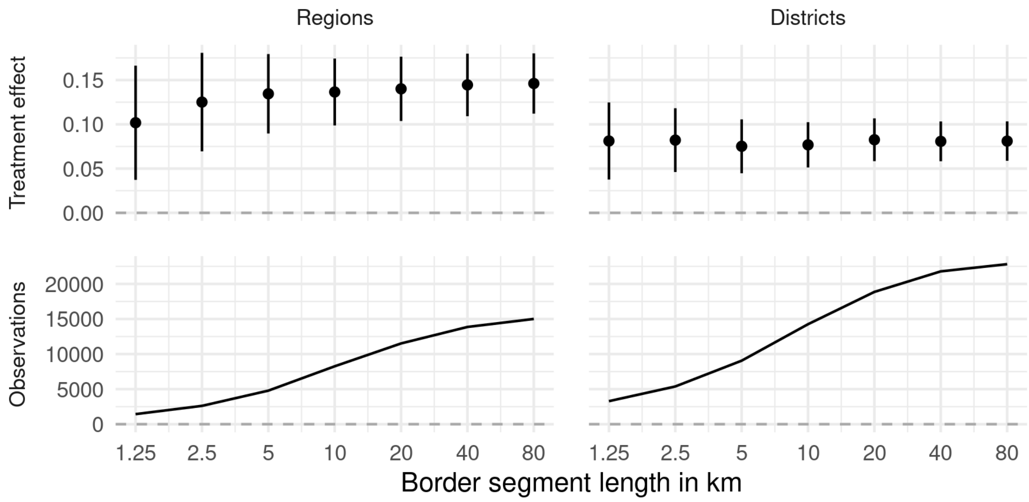


Figure A13: Estimate of border effect on local ethnic identities by maximum length of border segments.

Note: Stable bandwidth of 20km. See also baseline Models 1 and 4 in Table 1 in the main text.

5km. At that bandwidth, a large proportion of the data will be affected be geographical misattribution, particularly around district borders which are not respected at all by the random displacement of DHS survey cluster coordinates.

Even with a smaller bandwidth the model relies on variation between potentially distant survey clusters located at opposite ends of a long border. To assess whether the results hold along short border segments, I define sub-border segments of lengths varying exponentially between 1.25 and 80 km. To do so, I first record for each EA p the closest point π_p on a border. For each target segment length, I then identify clusters of points π along a border that have a maximum diameter of the defined segment length. This ensures that wiggly and straight segments are treated similarly.[28] The clustering is implemented via a hierarchical clustering algorithm that maximizes the size of the sample kept in the analysis.[29] Estimating the main Eq. 1 with border-segment fixed effects so defined ensures that we exploit only variation from within a small geographic neighborhood. As visualized in Figure A13 and despite the small sample size around short segments, doing so does not change the results in a substantive way. Regional and district border estimates are stable for the whole range of segment lengths between 2.5 and 80km and only slightly decrease to 10 percentage points with 1.25km short segments at the region level.

F.4 Units of analysis

The main analysis is interested in the effect of administrative borders on local ethnic demographics and therefore uses the survey clusters as the main unit of analysis

[28]Note that this is not the case if the length of segments is measured along the path of a border. Then, straight segments would cover a greater distance than wiggly ones.

[29]This is necessary as an arbitrary segmentation of borders would create unnecessary many segments with few EAs or EAs on only one side of the border, segments which are dropped because they add no information.

and its share of respondents belonging to an administrative unit's plurality group as the outcome of interest. However, because ethnic identities are rooted in individuals, the individual respondent constitutes an alternative unit of analysis. To assess whether this choice changes the main results, I use a simple dummy variable for whether a respondent identifies with the plurality group of the treatment administrative unit as the main outcome and follow the main specification from Eq. 1 in all other regards. Figure A9 plots the resulting treatment effect estimates which are next to identical to the main estimates.

In addition, I revisit the decision to attribute each survey cluster only to one border segment rather than all border segments in its neighborhood. Because some EAs are closer than 20km to more than 1 border segments, doing so ads some (marginal) information to the analysis, but makes it necessary to down-weight observations so that each survey cluster receives the same weight in the analysis. Results in Figure A9 show that doing so does not change the results.

F.5 Alternative survey data from the Afrobarometer

I also draw on the Afrobarometer (2018) Surveys rounds 1-6 to vary the source of survey data used to measure the proportion of local populations speaking the language of administrative units' plurality group. With the data, I proceed in the very same way as with the DHS data, matching each survey cluster – geocoded by Ben Yishay, Ariel Rotberg et al. (2017) – to its administrative districts and regions and nearby borders. I then assess whether respondents speak the language of units' plurality group as identified by Murdock (1959), again drawing on the ethnic group mapping from Müller-Crepon, Pengl and Bormann (2022). Finally, I re-estimate the main models using both survey clusters and respondents as the unit of analysis. Despite differences in the countries covered and ethnic groups enumerated by the Afrobarometer, the results plotted in Figure A9 above closely coincide with those obtained from the DHS data.

F.6 Testing for survey-related biases

Two biases that may be inherent to the DHS data may affect the results. The first bias relates to the potential desire of respondents to appear as coethnics of interviewers.[30] If survey teams are organized based on administrative regions and staffed to correspond to regions' (or districts') plurality ethnic groups, such social desirability bias would lead to sharp changes in the reported ethnic identity of respondents at regional and district borders. While social desirability bias of the magnitude of the baseline results would be surprising, such a dynamic may nevertheless lead to an overestimation of the effect of administrative borders.

To avoid this bias from affecting the results, I return to the individual-level analysis of Afrobarometer surveys from above (Subsection F.5), and use information on interviewers' home language to code whether s/he belongs to the plurality ethnic group of the treatment unit.[31] I then re-estimate the main model from Eq. 1

[30] For the effects of coethnicity between interviewers and respondents see Adida et al. (2016).

[31] Thus, this coding mirrors the construction of the main outcome variable.

adding this dummy variable as a control variable. While I find that the interviewers' plurality group membership is positively related to that of respondents[32], this correlation does not affect the estimated discontinuous jump in ethnic identities at district and regional borders (see Figure A9).

A second caveat inherent in the data is that the language-based matching of ethnic groups mapped by Murdock (1959) and groups enumerated in the DHS is "many-to-many," with some groups from Murdock corresponding to multiple groups in the DHS surveys and vice-versa. Hence, for some borders with two differing plurality groups from Murdock at either side, the same DHS group(s) are attributed to both.[33] In such cases, the main model estimates a causal effect of 0 by default, since the plurality groups as measured in the DHS are the same on both sides of the border. While this may be the result of post-treatment dynamics, which is why the respective observations are regular part of the main analysis, it may also be the result of random coding particularities, leading the main estimates to be downwards biased. To judge the extent of such bias, I drop all observations along borders for which the plurality groups on either side are not distinguishable in the DHS data. The respective results (Figure A9) closely align with the baseline estimates, indicating that such biases are not severe.

F.7 Alternative ethnic settlement data from the Atlas Narodov Mira

I here assess whether the results may be driven by features of Murdock's map of ethnic groups and its comparative imprecision. To do so, I use the second available pan-African data on ethnic settlement patterns, the Atlas Narodov Mira (ANM Bruk and Apenchenko 1964), to identify administrative units' plurality groups. The ANM was compiled by Russian ethnographers and digitized by Weidmann, Rød and Cederman (2010). It has the advantage over Murdock's map that it captures local ethnic heterogeneity through overlapping ethnic settlement patterns[34] and has a higher spatial resolution. Yet, it is dated later than Murdock's map, more obviously affected by national borders, and maps ethnic groups at a higher level of ethno-linguistic aggregation. However, replicating the main estimates using the data from the ANM does not substantively change the results as visible in Figure A9. While the effects at regional borders are of similar magnitude, those estimated at district borders are larger.

F.8 Alternative contemporary border data

An alternative source of contemporary administrative border data consists in the Database of Global Administrative Areas (GADM) which was last updated in 2019. Unfortunately, the data is only one cross-sectional snapshot, leading to increased risk of reverse causality by which today's borders exist due to past ethnic boundaries in space. However, the analyses within ethnic groups and along straight borders again mitigate this caveat. As Figure A9 above shows, the results closely align with the baseline estimates.

[32] This reflects either biased responses or strategic interviewer assignment.

[33] One example consists in the disaggregated mapping of various Akan subgroups by Murdock, all of which are collapsed under the "Akan" label in the Ghanaian DHS data.

[34] I adjust the measurement of ethnic groups' territorial share in administrative units to such overlaps.

F.9 Country-level jackknife

I use a county-level jackknife estimation to gauge the impact every country in the main sample has on the baseline estimates. Figure A14 plots the results of re-estimating the main model, in each iteration dropping observations from the country indicated on the y-axis. The results show that no single country drives the results in a significant manner.

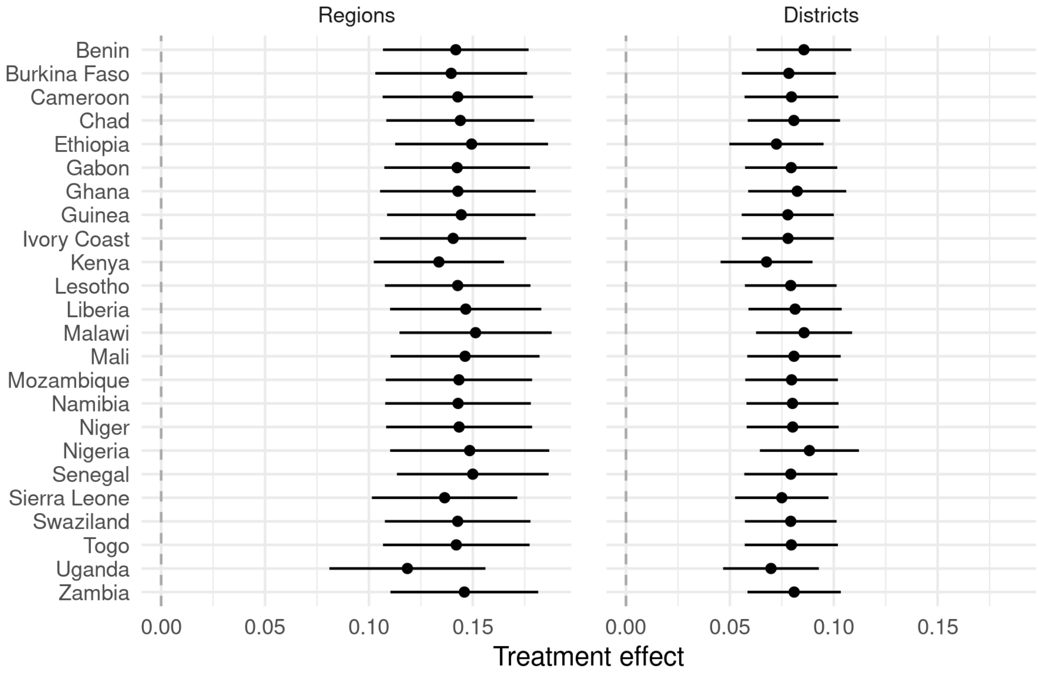


Figure A14: Estimate of border effect on local ethnic identities, dropping one country at the time.

Note: Estimates based on the baseline Models 1 and 4 in Table 1 in the main text.

F.10 Plurality status and ethnic demographics: Generalizability

I here test whether the effect of plurality status on groups' population shares estimated through the RDD – by definition identified as a Local Average Treatment Effect *at the border* of administrative units – generalizes to the inside of administrative units. I do so by taking ethnic groups nested within administrative units (regions and districts) as the main units of analysis. For each group in a unit, I compute the share of the administrative unit their settlement area covers as the result of the spatial intersection between Murdock's (1959) map and the spatial data on administrative units. Consistent with the main analysis, I use the same settlement share to derive groups' plurality status in a unit. I then take the full DHS data and compute, for each group,[35] the respective population share in a unit across all

[35] After matching ethnic group labels in the DHS with Murdock's group names via the LEDA package (Müller-Crepon, Pengl and Bormann 2022).

survey rounds. I then estimate a linear regression of:

$$log(.01+\text{survey share}_{u,g,c}) = \alpha_g+\delta_u+\beta_1\text{plurality}+\beta_2 log(.01+\text{settlement share})+\epsilon_{u,g,c},$$

where the main outcome, groups g share in the survey data of country c, is explained by groups local plurality status and their territorial settlement share of the respective unit u. In addition, I add fixed effects iteratively. I control for group fixed effects α_g to capture Murdock group-level idiosyncrasies and potential omitted variables in the data, as for example the quality of the match between ethnic labels in the DHS and Murdock's map for a particular group. Second, I control for administrative unit fixed effects δ_u to account for potential biasing unit level characteristics, for example the fact that group shares do not always add to 1 within units due to the imperfect matching between the two data sets and the disregard of Murdock groups with a settlement share of 0 (some of which are nevertheless represented in the survey data from a unit).

Table A5: Plurality status and ethnic groups population share in administrative units

	Share among survey respondents (log)					
	Regions			Districts		
	(1)	(2)	(3)	(4)	(5)	(6)
Plurality group	0.586***	0.656***	0.517***	0.427***	0.559***	0.319***
	(0.123)	(0.118)	(0.110)	(0.095)	(0.080)	(0.080)
Settlement share (log)	0.358***	0.335***	0.379***	0.278***	0.244***	0.294***
	(0.036)	(0.034)	(0.035)	(0.033)	(0.026)	(0.030)
Country FE:	–	yes	–	–	yes	–
Admin. Unit FE:	yes	no	yes	yes	no	yes
Group FE:	no	yes	yes	no	yes	yes
Mean DV:	-2.1	-2.1	-2.1	-1.6	-1.6	-1.6
Observations	1,607	1,607	1,607	4,549	4,549	4,549
Adjusted R^2	0.337	0.536	0.566	0.155	0.500	0.443

Notes: OLS linear models. The unit of analysis is the ethnic group nested in administrative units. The outcome is the share of DHS respondents in that unit that identifies with the group across all survey waves. Standard errors clustered on the administrative unit and ethnic group levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01

The results shown in Table A5 show a substantive "plurality group premium". Controlling for groups' territorial share in a unit as well as the full set of fixed effects, the results suggest that plurality status increases groups' population share substantively. For example, at the regional (district) level, Model 3 (6) indicates that a group with a baseline population share of 40% would see an increase of 28 (15) percentage points if made the plurality group in a unit, thus increasing its population share to 68% (55%). These are substantial differences which are larger than those estimated in the RDD setup. While we cannot interpret these correlations causally due to potential omitted variable biases and reverse causality,[36] these results provide further suggestive evidence that the RDD estimates generalize towards the interior of administrative units.

[36] As stressed in the main paper, while the precise location of borders is likely as-if-random, the overall design of administrative units at a higher geographic level is likely not.

G Ethnic assimilation: Additional results

This section presents additional results on ethnic assimilation and integration only briefly mentioned in the main paper. First, Table A6 presents the results of a simple cross-sectional analysis of local plurality members' ethnic vs. national self-identification. The results show that speakers of the local plurality language identify more with their ethnic group (as opposed to the nation) than local minority members. This is consistent with the latter overcoming their own ethnic identities in favor of greater assimilation and integration.

Table A7 presents the full results of the analysis of interethnic marriages along regional and district borders. Using the baseline RD-design, Models 1 and 4 in Table A7 show that marriages with men from the local plurality group become 9 (8) percentage points more frequent as one crosses regional (district) borders. This is not all too surprising, since the number of female respondents that identify with that group rises as well. Models 2 and 5 therefore control for whether a respondent speak the local plurality language or not. This yields a smaller, yet statistically significant estimate of an increase of about 2 percentage points along both border types. Lastly, Models 3 and 6 control for the share of the plurality group in the local population (*P.G.S.*) in interaction with the treatment dummy, showing that this effect can be explained by the post-treatment change in the 'supply' of plurality men at the border. Table A8 shows very similar dynamics among the spouses of men interviewed by the DHS.

Table A6: Local plurality status and ordinally scaled strength of ethnic (1) versus national (5) identification

	Regions	Districts
	(1)	(2)
Plurality co-ethnic	−0.057***	−0.047***
	(0.014)	(0.017)
Adm. unit FE	yes	yes
Lang. group FE	yes	yes
Covariates	yes	yes
Mean DV:	3.7	3.7
Observations	103,741	83,082
Adjusted R^2	0.148	0.156

Notes: OLS linear models. Covariates consist of respondents' age and its square and a female dummy. Standard errors clustered on the point and administrative unit levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01

H Migration analysis: Additional results

This section discusses additional analyses of the ethnic migration patterns assessed in the main paper. I first present the main robustness checks mentioned in the article (Subsection H.1) and then discuss the correlation of ethnic migration patterns with the effect of administrative borders on ethnic identities (Subsection H.2).

Table A7: (Minority) women's marriage to plurality group husband

	Married to plurality group husband					
	Regions			Districts		
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.090***	0.018**	−0.003	0.079***	0.026***	−0.003
	(0.020)	(0.008)	(0.005)	(0.014)	(0.007)	(0.004)
Plur. grp. (0/1)		0.745***	0.475***		0.673***	0.409***
		(0.015)	(0.036)		(0.012)	(0.018)
Treated × P.G.		0.014	−0.017		0.004	0.001
		(0.013)	(0.046)		(0.009)	(0.022)
Plur. grp. share (%)			0.537***			0.601***
			(0.037)			(0.020)
Treated × P.G.S.			0.019			0.002
			(0.046)			(0.022)
Cutoff	20km	20km	20km	20km	20km	20km
Running var linear	yes	yes	yes	yes	yes	yes
Survey FE:	yes	yes	yes	yes	yes	yes
Border FE:	yes	yes	yes	yes	yes	yes
Mean DV:	0.35	0.35	0.35	0.4	0.4	0.4
Borders:	246	246	245	743	743	729
Observations	44,680	44,680	44,456	70,056	70,056	68,604
Adjusted R^2	0.479	0.777	0.818	0.534	0.750	0.792

Notes: OLS linear models. Standard errors clustered on the point and administrative unit × treatment levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01

Table A8: (Minority) men's marriage to plurality group wife

	Married to plurality group wife					
	Regions			Districts		
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.095***	0.022**	0.004	0.079***	0.025***	0.0003
	(0.019)	(0.009)	(0.005)	(0.013)	(0.007)	(0.004)
Plur. grp. (0/1)		0.750***	0.508***		0.682***	0.432***
		(0.017)	(0.039)		(0.013)	(0.020)
Treated × P.G.		0.008	−0.017		−0.001	−0.001
		(0.014)	(0.050)		(0.010)	(0.025)
Plur. grp. share (%)			0.474***			0.558***
			(0.041)			(0.021)
Treated × P.G.S.			0.016			−0.004
			(0.050)			(0.025)
Cutoff	20km	20km	20km	20km	20km	20km
Running var linear	yes	yes	yes	yes	yes	yes
Survey FE:	yes	yes	yes	yes	yes	yes
Border FE:	yes	yes	yes	yes	yes	yes
Mean DV:	0.34	0.34	0.34	0.4	0.4	0.4
Borders:	246	246	245	743	743	729
Observations	40,130	40,130	39,920	63,646	63,646	62,400
Adjusted R^2	0.467	0.769	0.800	0.518	0.738	0.773

Notes: OLS linear models. Standard errors clustered on the point and administrative unit × treatment levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01

H.1 Robustness checks

Figure A15 summarizes the results of four robustness check of the dyadic analysis of ethnic migration patterns presented in Models 3 and 6 in Table 3 in the main text. I first distinguish between censuses that base the identification of ethnic groups on citizens' language versus their ethnic self-identification. The results show that the language-based analysis yields larger estimates of ethnically biased migration patterns. I then assess whether the results hold if I base the coding of unit's plurality ethnic group on the modal ethnic identity of people born in a unit before 1960 rather than Murdock's (1959) map. The results indicate larger effects than the baseline estimates. This deviation may either stem from a post-treatment bias by which some ethnic identities relevant for Murdock (1959) lost relevance over time or relate to noise introduced by Murdock's map. Lastly, reducing the sample to the five countries for which I have regional and district-level data does not change the results of the region-level analysis.

An assessment of ethnic migration patterns by birth-decade in Figure A16 shows very persistent patterns as far back as to cohorts born in the 1900s which only decrease for generations born after 1980. Because these are relatively young at the time of census-taking, their life-time migration has not proceeded as much as that of older generation, thus leading to smaller (yet statistically significant) estimates. Lastly, Figure A17 assesses whether the results are due to any single country in the sample by estimating a country-level jackknife. The results indicate that the results remain stable as we successively exclude each country from the sample.

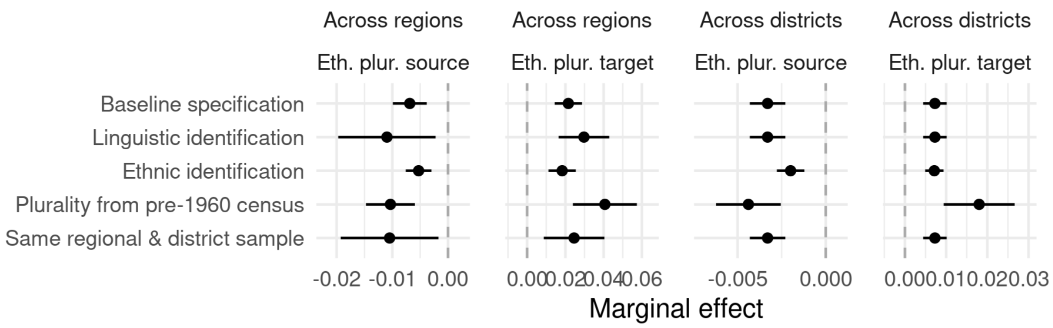


Figure A15: Dyadic migration analysis: Additional results.

Note: Models are based on the dyadic specifications (Models 3 and 6) in Table 3.

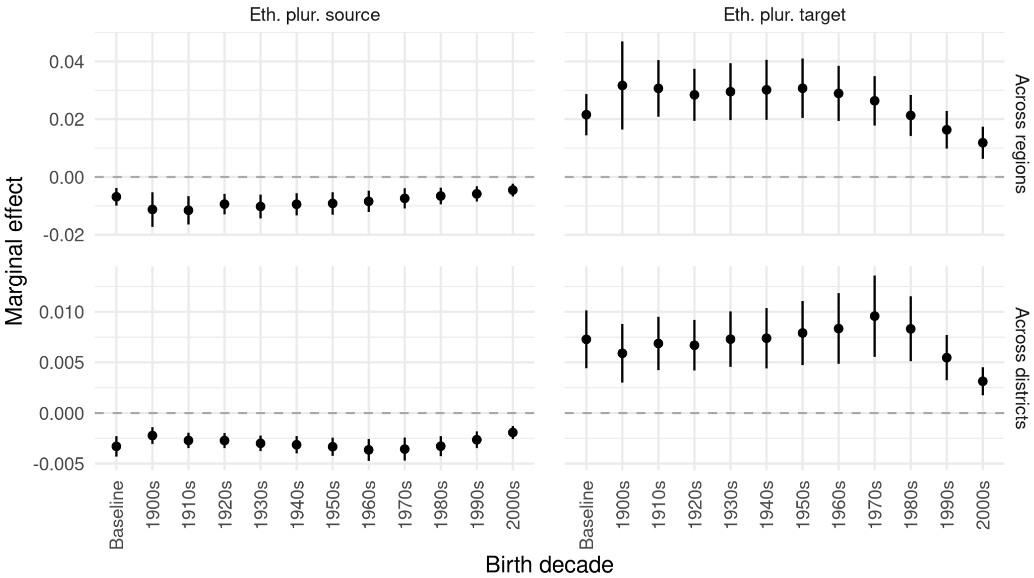


Figure A16: Dyadic migration analysis by birth decade

Note: Models are based on the dyadic specifications (Models 3 and 6) in Table 3.

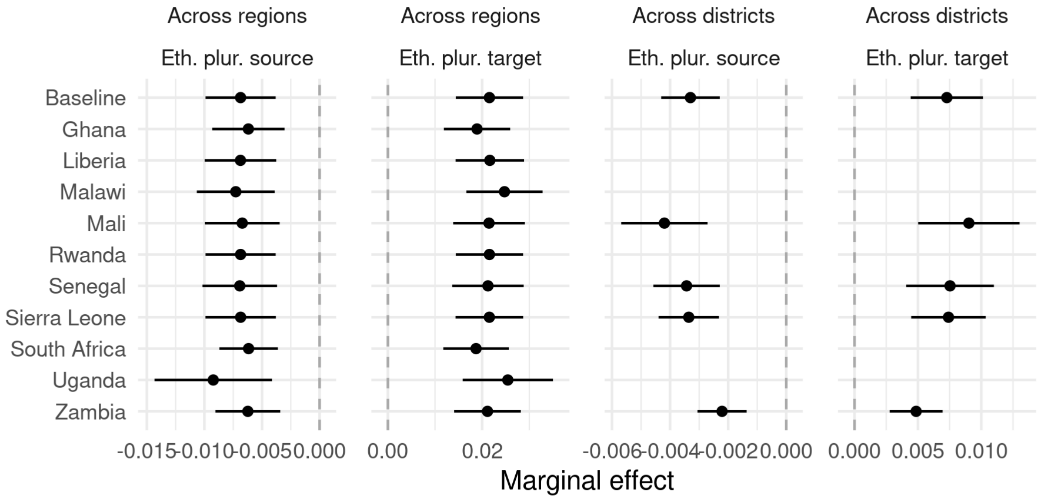


Figure A17: Dyadic migration analysis: Country-level jackknife.

Note: Models are based on the dyadic specifications (Models 3 and 6) in Table 3.

H.2 Correlation of biased migration with RD-estimates

As a last step in the analysis of ethnic migration patterns, I assess whether they credibly contribute to the effects of administrative borders on ethnic demography found in the main empirical section of the paper. To that intent, I proceed in three steps. First, I replicate the main analysis of the effect of regional and district borders on ethnic identities, now using the borders corresponding to the census data to assess discontinuities in DHS respondents' ethnic identification. As shown in Models 1 and 3 in Table A9, the respective results align with the main results at

Table A9: Variation in border effects by extent of ethnically 'biased' migration

	Outcome: Plurality group share (0-1)					
	Regions			Districts		
	Baseline	Emigr.	Immigr.	Baseline	Emigr.	Immigr.
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.153***	0.118***	0.055**	0.091***	0.055*	0.002
	(0.023)	(0.023)	(0.023)	(0.027)	(0.029)	(0.031)
Treated × $\Delta_{u,emigr.}$		−0.288**			−0.277***	
		(0.115)			(0.104)	
Treated × $\Delta_{u,immigr.}$			0.466***			0.396***
			(0.094)			(0.099)
Cutoff	20km	20km	20km	20km	20km	20km
Running var linear	yes	yes	yes	yes	yes	yes
Survey FE:	yes	yes	yes	yes	yes	yes
Border FE:	yes	yes	yes	yes	yes	yes
Mean DV:	0.42	0.42	0.42	0.42	0.42	0.42
Observations	8,328	8,328	8,328	3,713	3,713	3,713
Adjusted R^2	0.599	0.607	0.619	0.620	0.628	0.640

Notes: OLS linear models. Standard errors clustered on the point and administrative unit × treatment levels. Significance codes: $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

the regional and district level, respectively. Second, I estimate the extent to which ethnic plurality members differentially emigrate from and immigrate separately for each region and district. The respective estimation equations read:

$$Emigration_{u,e} = \alpha_u + \gamma_e + \Delta_{u,emigr.}\,\mathbb{1}(plurality_{e,u}) + \epsilon_{u,e} \tag{4}$$

and

$$Immigration_{u,e} = \alpha_u + \gamma_e + \Delta_{u,immigr.}\,\mathbb{1}(plurality_{e,u}) + \epsilon_{u,e}, \tag{5}$$

where the effect of being a plurality group on ethnic group e' emigration (immigration) rate from (to) administrative unit u is captured by the unit-specific coefficient Δ_u, estimated in the presence of ethic group and unit fixed effects α_u and γ_e.[37] I then merge these unit-level estimates of migration bias ($\Delta_{u,emigr.}$ and $\Delta_{u,immigr.}$) with the DHS data combined with the administrative boundaries from IPUMS, such that each observation from a treatment unit is assigned the $\Delta_{u,emigr.}$ and $\Delta_{u,immigr.}$ of that unit. The resulting dataset then allows me to assess the degree to which treatment effects along administrative borders correlate with the ethnic migration bias observed in the respective regions and districts as

$$\begin{aligned} Y_{p,b,t,s} =& \alpha_{b,t} + \gamma_s + \beta_1 T_{u,t} + \beta_2 T_{u,t}\Delta_u + \beta_3 D_{p,b} + \beta_4 D_{p,b} T_{u,t} + \\ & \beta_5 D_{p,b}\Delta_u + \beta_6 D_{p,b} T_{u,t}\Delta_u + \epsilon_{p,u,b,t}, \end{aligned} \tag{6}$$

which follows the main RD-specification from the main paper augmented with the unit-specific e-/immigration bias from above, where β_2 captures the correlation of the border effect with the migration bias.

[37]This cross-sectional difference-in-difference estimate is possible because ethnic groups e have members in many units u and units harbor are multi-ethnic.

This analysis is carried out in Models 2-3 and 5-6 in Table A9 for regions and districts, respectively. The first models (2 and 5) simple interactions of the treatment dummy with $\Delta_{u,emigr.}$, while the second ones (3 and 6) do the same with $\Delta_{u,immigr.}$. The results across all models show a strong and significant correlation between ethnically biased migration patterns and discontinuities in ethnic demography along administrative borders. In particular, the results show that regions in which plurality group members have a 10 percentage point lower emigration rate than other groups ($\Delta_{u,emigr.} = -.1$) show a 2.9 percentage points larger discontinuity in the size of their plurality ethnic group at their border. The effects of differential immigration are about twice that size and positive, i.e. increased plurality immigration increases the discontinuity in the plurality share at the border. The effect associated with differential immigration rates is about half of that size and a reversed sign. Thus, a higher immigration rate of plurality members is consistently associated with a greater effect of administrative borders on plurality group shares.

I References

Adida, Claire L., Karen E. Ferree, Daniel N. Posner and Amanda Lea Robinson. 2016. "Who's Asking? Interviewer Coethnicity Effects in African Survey Data." *Comparative Political Studies* 49(12):1630–1660.

Afrobarometer. 2018. "Afrobarometer Data." *Available at http://www.afrobarometer.org* .

Beiser-McGrath, Janina, Carl Müller-Crepon and Yannick I Pengl. 2021. "Who Benefits? How Local Ethnic Demography Shapes Political Favoritism in Africa." *British Journal of Political Science* 51(4):1582–1600.

Ben Yishay, Ariel Rotberg, Renee, Jessica Wells, Zhonghui Lv, Seth Goodman, Lidia Kovacevic and Dan Runfola. 2017. *Geocoding Afrobarometer Rounds 1-6: Methodology and Data Quality*. Williamsburg, VA: AidData.

Bruk, Solomon I. and V. S. Apenchenko. 1964. *Atlas Narodov Mira*. Moscow: Glavnoe upravlenie geodezii i kartografii gosudarstvennogo geologicheskogo komiteta SSSR and Institut etnografii im. H. H. Miklukho-Maklaia, Akademiia nauk SSSR.

De Luca, Giacomo, Roland Hodler, Paul A Raschky and Michele Valsecchi. 2018. "Ethnic favoritism: An axiom of politics?" *Journal of Development Economics* 132:115–129.

Ejdemyr, Simon, Eric Kramon and Amanda Lea Robinson. 2018. "Segregation, Ethnic Favoritism, and the Strategic Targeting of Local Public Goods." *Comparative Political Studies* 51(9):1111 –1143.

Franck, Raphaël and Ilia Rainer. 2012. "Does the Leader's Ethnicity Matter? Ethnic Favoritism, Education, and Health in Sub-Saharan Africa." *American Political Science Review* 106(2):294–325.

Harris, J Andrew and Daniel N Posner. 2019. "(Under What Conditions) Do Politicians Reward Their Supporters? Evidence from Kenya's Constituencies Development Fund." *American Political Science Review* 113(1):123–139.

Huillery, Elise. 2009. "History matters: the long-term impact of colonial public investments in French West Africa." *American Economic Journal: Applied Economics* 1(2):176–215.

Lehner, Bernhard, Kristine Verdin and Andy Jarvis. 2008. "New global hydrography derived from spaceborne elevation data." *Eos, Transactions American Geophysical Union* 89(10):93–94.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of econometrics* 142(2):698–714.

Müller-Crepon, Carl. 2020. "Continuity or Change: (In)direct Rule in British and French Colonial Africa." *International Organization* 74(4):707–741.

Müller-Crepon, Carl. 2021. "State reach and development in Africa since the 1960s: new data and analysis." *Political Science Research and Methods* pp. 1–10.

Müller-Crepon, Carl, Yannick Pengl and Nils-Christian Bormann. 2022. "Linking Ethnic Data from Arica." *Journal of Peace Research* 59(3):425–435.

Murdock, George Peter. 1959. *Africa. Its Peoples and Their Culture History.* New York: McGraw-Hill Book Company.

Vinh, Nguyen Xuan, Julien Epps and James Bailey. 2010. "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance." *Journal of Machine Learning Research* 11(Oct):2837–2854.

Weidmann, Nils B., Jan Ketil Rød and Lars-Erik Cederman. 2010. "Representing ethnic groups in space: A new dataset." *Journal of Peace Research* 47(4):491–499.