Building Tribes: How Administrative Units Shaped Ethnic Groups in Africa

Carl Müller-Crepon*

University of Oxford

February 3, 2022

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.

*email: carl.muller-crepon@politics.ox.ac.uk. I thank Jeremy Bowles, Lars-Erik Cederman, Jessica Gottlieb, Lukas Haffert, Horacio Larreguy, Emma Lindstam, 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 neatly bounded administrative divisions, shaped ethnic groups in Sub-Saharan Africa. In doing so, I build on seminal works on the relation between the colonial imposition of territorial governance and ethnicity by Mamdani (2001, 2020) and Posner (2005). I argue that local and regional governments tend to favor the largest ethnic group in their population, in particular where (neo-)traditional institutions prevail. 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 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 contribute to this literature a focus on the effects of territorial rule imposed by colonialism, which revolutionized local governance, fostered the "invention of tradition" (Ranger 1997), and thus powered the transformation of ethnic identities (Lentz 1995; Southall 1970; Young 1985). This focus also highlights an analogy between

[1] A related literature assesses individuals' ethnic versus national identification (e.g., Eifert, Miguel and Posner 2010; Green 2020; Robinson 2014).

subnational 'tribe-building' in Africa and 'nation-building' elsewhere, both powered by material and ideological forces that aimed at increasing the coincidence between ethnic groups and political units (Argyle 1969). Studying the effect of often haphazardly drawn African borders on ethnic geography is not only intrinsically important but also circumvents inference problems from endogenous borders elsewhere.

I test my argument that administrative units shaped ethnic groups by estimating the change in ethnic group shares at units' boundaries. Employing a regression discontinuity design, I find that the share of regions' (districts') main ethnic group sharply increases by 14 (8) percentage points at their borders with units dominated by a different group. Ruling out omitted variable bias and reverse causality from endogenous borders and their change, these effects are similar when restricting variation to within colonial-era settlement areas of ethnic groups, as well as to variation at relatively straight, arbitrarily drawn subnational boundaries. Consistent with arguing that ethnically biased governance creates incentives for minority assimilation and emigration, estimated treatment effects are larger where traditional institutions are more powerful and where predominant groups enjoy larger size advantages.

I present direct evidence that assimilation and emigration by local minorities account for these border effects. Using a variant of the main regression discontinuity design, I find that local minority status increases assimilation to the majority through language adoption and intermarriage. In addition, I present strong evidence for ethnic sorting across administrative units in census data on 33 million individuals: local minorities exhibit higher rates of emigration and lower levels of immigration than predominant groups.

Providing evidence on how administrative geographies transformed ethnic identities 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 state 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 governments tend to ethnically specialize their governance towards large ethnic groups and discriminate against local minorities. Disenfranchised minorities can improve their lot through assimilation or emigration, thereby selecting into a local majority. Because ethnically biased governance and minority responses are delimited by administrative borders, the resulting transformation of ethnic groups aligns ethnic with administrative boundaries. Minorities can alternatively demand secession, creating new borders that more closely mirror ethnic geography. While of great importance, I here focus on ethnic change, leaving the change of borders 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 heterogeneity makes it impossible to devise homogeneous yet non-overlapping and contiguous divisions. Governments of multiethnic populations frequently favor large, powerful groups against minorities. Their ethnic favoritism stems from leader's ethnic identity or the 'ethnic specialization' of governance.

The most extensively analyzed cause of ethnic favoritism is that governments cater goods and services to 'their' ethnic constituencies (De Luca et al. 2018; Franck and Rainer 2012). Leaders may intrinsically prefer to see their ethnic group prosper (Chandra 2007) and more profitably exchange political support for material benefits with their co-ethnics (Fearon 1999). Because large ethnic groups hold, on average, most executive power (Bormann 2019), this form of ethnic favoritism disadvantages ethnic minorities.

But ethnicized governance also emerges where governments 'specialize' in large ethnic groups by using specific languages or legitimizing their power with references to ethnic traditions. Both increases co-ethnics' "quasi-voluntary compliance" (Levi 1988) and can reduce the costs of governance. 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 governing large ethnic groups leaving minorities worse off.

Similar incentives can lead governments to ethnically homogenize minority populations (Alesina and Spolaore 2005). In particular through education policies, governments 're-educate' ethnic minorities to learn the majority language and customs, become part of the majority population, and increase its 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. Ethnic minorities in turn end up disenfranchised and under-serviced and may become targets of homogenization policies. The strength of these dynamics increases with the relative size of the largest ethnic group.

Minority responses to ethnicized governance

Ethnically biased governance often leaves minorities with less access to state-provided goods. However, minority members may improve their lot through assimilation, emigration, or mobilization for creating their own governance unit.

Ethnic assimilation aims at changing one's language, religion, appearance, and taste to credibly claim membership and benefits of the majority group (Laitin 1995). Assimilation thus consists of a diverse set of actions that increase one's ability to 'pass' as a member of the majority group. Because some hard-to-change group characteristics are innate or learned during early childhood, assimilation often proceeds intergenerationally through ethnic intermarriage (Kalmijn 1998).

Minority members can also exit their governance unit through migration. Facing discriminatory governance at home, individuals may either ethnically sort into governance units where they belong to the ethnic majority or migrate to prosperous areas where government 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

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

to assimilate or emigrate. Note, however, that parallel *heterogenizing* processes such as migration unrelated to ethnic sorting likely prevent a stable equilibrium.

Where political boundaries can be contested, spatially concentrated minorities may also mobilize against their discrimination and demand their own governance unit, achieved often by 'upgrading' one or multiple existing subunits (e.g., districts or counties; Doyle 2009; Green 2008; Grossman and Lewis 2014). While border change can increase the overall congruence between administrative and ethnic geography, it also tends to create new minorities as local ethnic diversity often prohibits upgraded or new borders' perfect alignment with the prevailing ethnic geography (Sambanis and Schulhofer-Wohl 2009). While certainly an important process, I here focus on *ethnic* change through assimilation and migration and address endogenous border change as an empirical challenge below.

In sum, I argue that the initial geography of governance units and ethnic groups determine the status of ethnic groups 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. This leads to a sharp increase of the local population share that belongs to a unit's predominant ethnic group at borders to units with differing predominant groups.

Administrative units and ethnicity in Africa

Sub-Saharan Africa provides a suitable testing ground for this argument. As outlined in the following, European colonialists create multi-ethnic administrative units with borders that often disregarded local ethnic geographies. Because governance through 'traditional' authorities was ethnicized, local minorities had incentives to assimilate or emigrate. Spatially structuring this process, the administrative partitioning of the (post-)colonial state molded ethnic groups.

While this empirical focus is valuable in its own right, I expect territorial governance to affect ethnic identities beyond the subnational level in Sub-Saharan Africa. Yet, my empirical focus comes with two benefits. First, African states are among the most ethnically diverse, and most feature no clear majority group or homogenization into single ethno-national identities (Laitin 1992, 1994). Focusing on how regional borders 'cement' subnational ethnic identities contributes to explaining this pattern. Second, African borders were often haphazardly drawn (e.g., Herbst 2000). Studying their effects on ethnic groups thus minimizes the risk of reverse causality incurred 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, polities were unbounded and non-contiguous, their power radiating outwards from the center (Herbst 2000; Wilks 1975). Political borders were conceptually even more foreign to acephalous societies where the lack of centralized power made separation lines superfluous.

European colonization radically changed these political topographies. The conquerors carved up the continent into colonies partitioned into administrative units to roll out the territorial governance that established purported effective control. The creation of administrative partitionings – 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.

The design of governance units determined their initial ethno-demographic composition and may have been influenced by ethnic geography. Such influence should have been strongest in indirectly ruled areas, in particular the British colonies, where power often remained with precolonial authorities (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). 'Tribes' and their 'homelands' were therefore seen as 'natural' governance units (Asiwaju 1970; Crowder 1968), an idea that clashed with a reality of interspersed ethnic settlement areas (Cohen and Middleton 1970) and flexible political loyalties that all too often cut across ethno-spatial lines (Lentz 1995, Southall 1970). As a result, pragmatism coupled with administrative and geographic exigencies determined the precise location of borders (Lentz 2006, p. 53) and "forced 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 and were directly tied to customary law that increased the status of locally predominant groups (Mamdani 2001). This ethnicization of local governance discriminated minorities and set the stage for the shaping of ethnic groups along administrative borders.

Local ruling elites fostered the homogenization of local ethnic identity for their own benefit. Because rulers' credentials were often challenged in absence of their 'customary' legitimacy as perceived by the population, 'customs' and 'traditions' were important factors in struggles over political power. Thus, the "invention" (Ranger 1997) of traditions and history became a tool for local elites' political survival still used today (Iliffe 1979; Robinson 2017). Political battles over the customary played out, for example, when Councils of Elders in the Kenyan Taita Hills synthesized lineage practices to control the chiefs (Bravman 1998) or when obas, chiefs, and educated elites reconstructed contending versions of traditional authority in Nigeria's Oyo Province (Vaughan 2003). These processes were often exacerbated where economic stakes increased the prize of offering the prevailing definition of the majority (La Fontaine 1969; Pengl, Roessler and Rueda 2021).

In sum, qualitative evidence describes the ethnicization of local governance and the distribution of the "goods of modernity" (Bates 1974): enforcement of property rights, access to markets, and social services. Based on ethnically exclusive definitions of the 'customary,' these goods could be distributed to local elites and their constituents (e.g., Bates 1974; Posner 2005; Vaughan 2003). Ethnic favoritism is thus not only a feature of African national politics (Franck and Rainer 2012), but also local politics. In Afrobarometer (2018) surveys, local minorities perceive local authorities as unresponsive, tend to mistrust them, and approve of them less. These patterns are stronger vis-à-vis traditional than state authorities (see Appendix A).[3]

Delimiting the territorial scope of ethnicized governance, administrative borders shaped identities in two ways. First, they defined the space within which identities were invented or negotiated. Returning to the Taita Hills, where political power was precolonially decentralized, Bravman (1998, p. 111) notes that "institutions, practices, and discourses of authority that defined Taita as single administrative entity promoted new social categories".[4] Second, where identities were more clearly delineated, borders defined units' ethnic make-up and assigned minority or majority status to individuals. I therefore predict that minority members had incentives to respond to their discrimination by becoming part of a locally powerful group through assimilation or emigration (see also Posner 2005).

Ethnic assimilation is a historical constant across Sub-Saharan Africa, in particular among ethnic 'strangers' (e.g., Cohen and Middleton 1970). For example, Gravesen and Kioko (2019) describe how Kenyan Kikuyu settlers in a former Maasai reserve assimilated by adopting language and traditions, as well as intermarriage to secure land rights.[5] Relatedly, Green (2021) suggests that some African citizens instrumentally adapt their identity to become co-ethnics of their president. Drawing on migration as a vehicle of 'revolt,' Asiwaju (1976) describes how ethnic minorities emigrated due to local discrimination.

In sum then, prior literature suggests that ethnicized territorial governance spurred ethnic change within the boundaries of administrative units. The following quantitative analysis systematically examines the resulting alignment of ethnic groups with administrative borders and assesses individual-level assimilation and emigration as underlying mechanisms.

[3]Similarly, Ejdemyr, Kramon and Robinson (2018) and Harris and Posner (2019) find MPs to engage in ethnic favoritism, in particular in ethnically segregated constituencies.

[4]This effect was also driven by non-state organizations, in particular missions, that aligned their activities (education and language standardization) with administrative borders (e.g., Chimhundu 1992).

[5]See Stahl (1991) for similar evidence from Ghana and Schultz (1984) from Northern Cameroon.

Research design

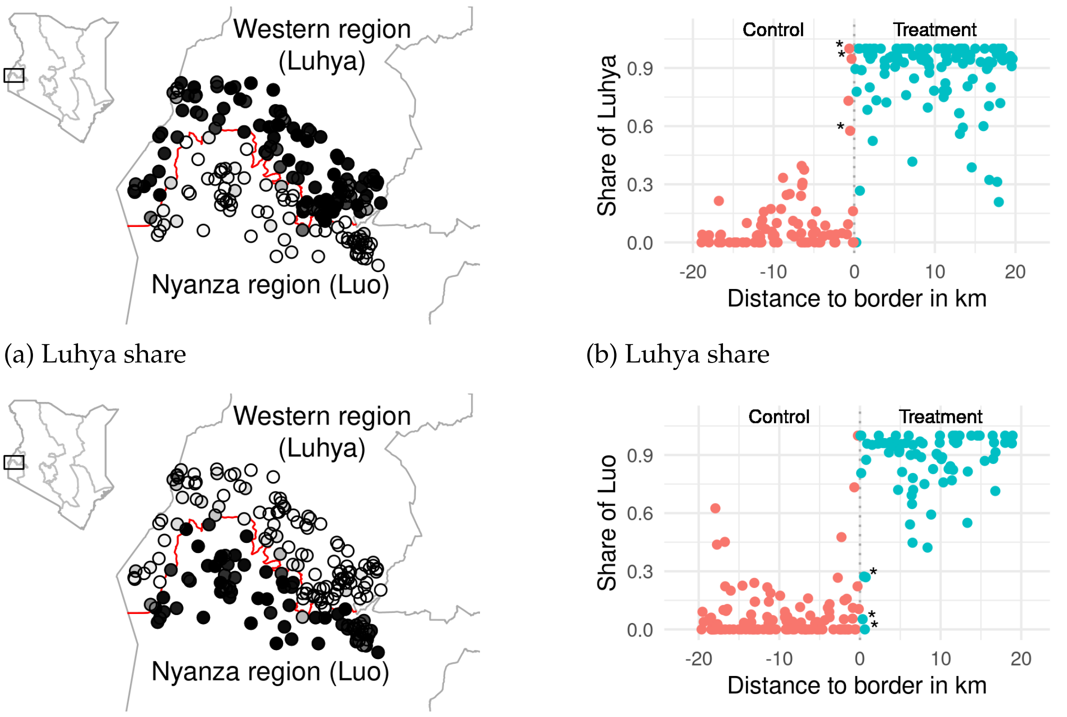
This section introduces the research design used to investigate whether administrative boundaries have shaped ethnic groups in Sub-Saharan Africa. I do so by examining individuals' ethnic identity across DHS (2018) surveys from 25 countries. Building on studies of African border effects (McCauley and 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 (Keele and Titiunik 2015; Henn 2021).[6] Focusing on credibly exogenous intra-ethnic and straight borders addresses concerns about omitted variable bias and reverse causality. Additional analyses shed light on ethnic assimilation and migration of local minorities as drivers of this process.

The curious (and extreme) case of the Kenyan Luhya

The intuition behind the regression discontinuity design is well captured by the case of the Luhya in western Kenya. At the outset of the 20th century, North Kavirondo district was mainly inhabited by Bantu-speakers, speaking between 15 and 26 mutually intelligible dialects (MacArthur 2012; Wagner 1949). Elites of one of the local dialects, Wanga, held power in the district, which became part of Kenya's postcolonial, first-level administrative Western province. Gold, discovered in the early 1930s, was the common good to secure, leading local elites to foster a collective identity against white settlers. Thus, the ethnic label Luhya (or Luyia, 'kinship') was invented, quickly becoming 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, fostered among others by the Luo Language Committee (Peterson 2018). How did this transformation of ethnic identities affect the ethnic demography in the Western and Nyanza provinces? Evidencing the extreme success of the Luhya identity as regionally bounded con-

[6] An alternative research design would compare individuals' (localities') ethnic identity (composition) before and after the introduction of administrative borders. This is currently not feasible as no precise local data reaches back into the colonial period. Virtually all post-colonial data on ethnicity is either 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).

struct, Figure 1 shows a sharp change at the border between today's Western and Nyanza provinces (i.e., between the former North and South Kavirondo districts). The share of the Luhya population 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. Beyond being driven by the vertical integration of Luo and Luhya (sub)identities, this sharp jump in the shares of Luo and Luhya 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. This is only true, of course, if the border was locally drawn in a manner unrelated to ethnic identities. I will pay particular attention to this requirement.

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. This implies that each point p enters the analysis twice with two outcomes. It is part of the treated population with the local share of the plurality group of its own administrative unit as the outcome. And it is part of the control group with the share of the plurality group of the neighboring unit as the outcome.

This research design avoids the arbitrary assignment of treatment and control groups. The two-sided design therefore leads in expectation to the same but more precise point estimates than a one-sided design (see Appendix D.1). In addition, it balances treatment and control groups. Because each point is part of both groups, there are no differences in geographic covariates between them and the density function of the running variable is perfectly continuous at the border (Appendix C).

With this, 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 survey cluster 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 point. Following standard spatial regression discontinuity specifications (Keele and Titiunik 2015), I add a border $\times$ treatment-side fixed effect $\alpha_{b,t}$ as well as a survey-round fixed effect γ_s, the distance $D_{p,b}$ of p to the respective border segment b, and the interaction of that distance with the treatment dummy.

I cluster standard errors on the point level to account for the correlation of the outcomes within points, and the administrative unit $u \times$ treatment t level to capture the spatial clustering of treatment assignments. in the baseline specification, I only analyze points within 20km to the closest border.[7]

To estimate Eq. 1, I combine spatial data on administrative units and ethnic settlement areas with georeferenced Demographic and Health Survey (DHS 2018) data collected since the 1990s in 25 African countries (see Appendix B.1). I draw on the geographic data on districts and regions in 1990 from the FAO's (2014) GAUL database[8] and derive the local plurality group of each administrative unit by spatially intersecting it with Murdock's (1959) map of ethnic groups. 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).

In the next step, I delineate all borders between administrative units with differing plurality groups. I assign each survey cluster p from the DHS to its administrative district and region. I then only keep border segments between administrative units with at least one survey cluster closer than 20km to either side of the border. If a cluster p 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 cluster p that identifies with the plurality groups in its own and neighboring unit. Ethnic identities are enumerated by the DHS mostly as respondents' ethnic group, tribe, and language. I use the ethnic linkage data from Müller-Crepon, Pengl and Bormann (2021) to match ethnic labels from the DHS data with those on Murdock's map.[10]

The data used for estimating the discontinuity of ethnic demography at administrative borders comes with two main caveats. First, the precise locations of administrative boundaries are uncertain.[11] This uncertainty affects observed treatment assignment and biases estimates of β_1 towards zero. Second, DHS cluster coordinates are disturbed by random noise to preserve respondent's privacy, dis-

[7]I show robustness to varying thresholds below. Because marginal increases in the bandwidth add points in the treatment and control groups, optimal bandwidth estimators (Imbens and Kalyanaraman 2012) lead to inconsistent estimates (see also Henn 2021).

[8]Appendix D.10 shows robustness to colonial and alternative contemporary border data. Current border data, e.g. GADM, come with a higher risk of reverse causality.

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

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

[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.

placing 99% (1%) of all rural clusters by up to 5km (10km) (Burgert et al. 2013).[12] As noted in Figure 1, survey clusters close to administrative borders can therefore be spatially associated with the wrong administrative unit and treatment status. A number of robustness checks address these issues.

The main identifying assumption for an unbiased estimate of β_1 in Eq. 1 holds that borders are drawn as-if-randomly within their local spatial neighborhood. 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. Other borders were drawn more haphazardly, as straight lines cutting across geographical features and ethnic groups. I will use such borders below to causally identify β_1.

If the main identifying assumption holds, the RDD identifies the local 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, which is more rural, older, less educated, and materially poorer than other respondents in the same unit. Historical, ethnic, or environmental characteristic of these survey clusters do not differ systematically (Appendix C).

Empirical analysis

Figure 2 plots the distribution of the share of local plurality groups in the treatment and control groups along the 323 regional and 1019 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. The respective increase along district borders amounts to 8 percentage points. These average treatment effects are of substantive size if compared to the average plurality group share of 33 (39) percent in regions (districts). They are precisely estimated and sta-

[12]The DHS corrects this displacement to assign points to the 'right' region (but not district). However, it relies on the GADM data with contemporary borders potentially affected by reverse causality.

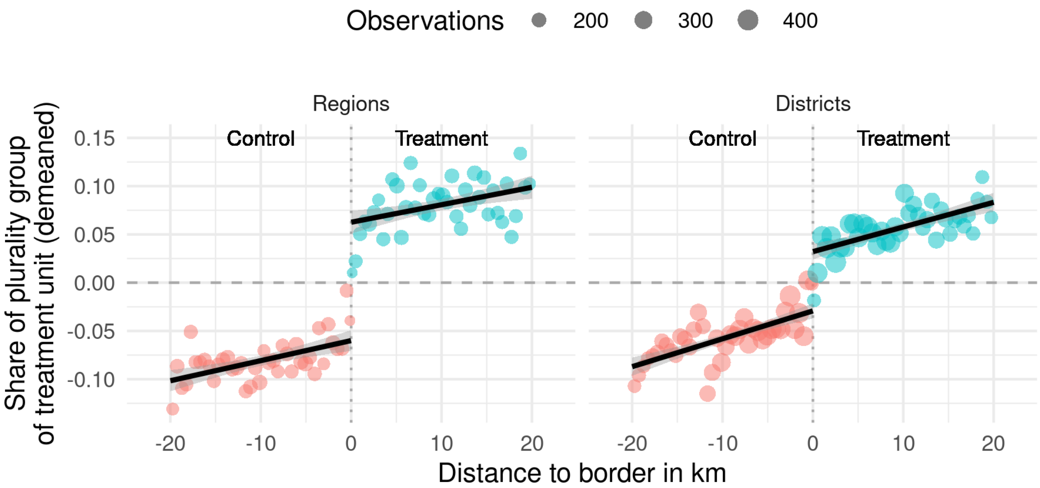


Figure 2: Local population shares of units' plurality groups increase at their borders.

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.

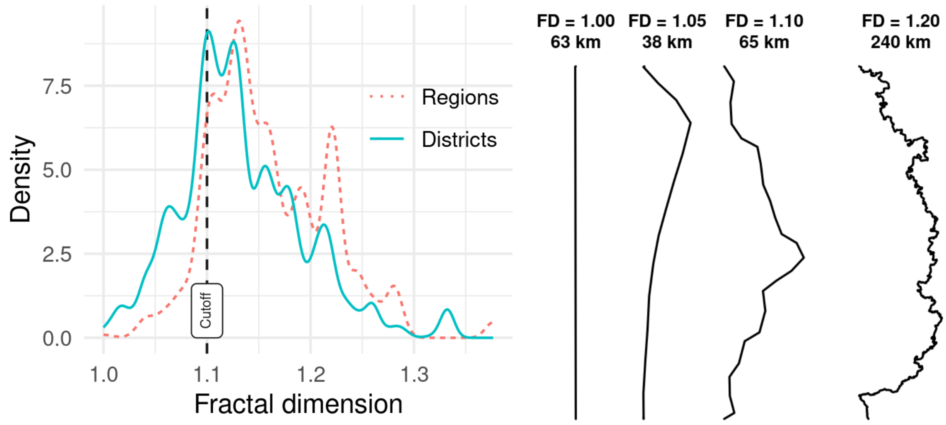
tistically highly significant.

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)
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
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 point and administrative unit × treatment levels. Significance codes: *p<0.1; **p<0.05; ***p<0.01

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



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

Figure 3: Fractal dimension of borders.

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

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 with the baseline results.

Second, a set of relatively straight borders 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 (Herbst 2000; Touval 1966), I measure borders' straightness as their fractal dimension (Alesina, Easterly and Matuszeski 2011), the degree to which they fill a two-dimensional plane.[13] 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.

To assess whether the baseline results hold along relatively arbitrary, straight borders, I limit the sample to points along borders with a 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 consist of comparatively few, straight line segments. The effects of straight borders in Models

[13]See Appendix B.2.

3 and 6 are consistent with the baseline results. The estimated effect of regional borders slightly decreases to 10.3 and the effect of district borders slightly increases to 9.1 percentage points. This stability is also apparent with varying cutoff-values for borders' fractal dimension and when splitting sample along quartiles of borders' alignment with rivers or watersheds as alternative proxies of the arbitrariness of borders (Appendix D.2). Taken together, the tests suggests that the results are not substantively affected by omitted variable bias or reverse causality.

Why are effects of district borders smaller than those of regional ones? Many district borders run within regions, while regional borders often entail the combined effect of changes in districts' and regions' plurality group. Supporting this interpretation, Appendix D.3 shows that 'pure' district borders that run within regions entail an effect of about 4 percentage points, while the effects of district borders aligned with regional ones are equivalent to regional border effects in Model 1. This suggests that border effects increase in the scale of administrative units, with national borders presumably associated with the largest effects.

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

Treatment effect heterogeneity

I propose above that anti-minority biases in local governance drive the minority assimilation and emigration that shape ethnic groups along adminsitrative borders. Minority disadvantages likely grow with the power of traditional institutions that are ethno-culturally legitimized.[14] They should also increase in the population share and margin of the plurality group as both incentivize greater ethnic specialization of governance.

[14]See also Appendix A.

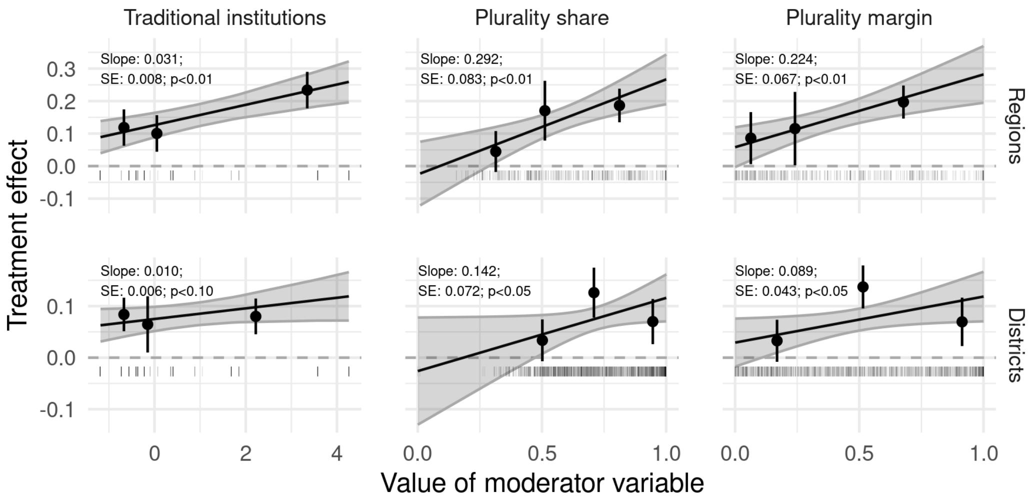


Figure 4: Treatment effect heterogeneity

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

I test these arguments by linearly interacting the treatment indicator and all other RDD-terms in Eq. 1 with three variables that operationalize these moderators: (1) a constitutionalization index of traditional institutions that proxies for their influence;[15] (2) the share of a unit's territory[16] settled by the plurality 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. 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, however, this analysis shows that administrative units with powerful traditional institutions and large plurality groups are associated with effects that are 1.5 to 2 times larger than average treatment effects.

[15]Constructed as the first principal component from Holzinger et al.'s (2019) data on the constitutional acknowledgment, regulation, and positive integration of traditional institutions, which explains 88% of the variables' variation.

[16]This is consistent with the territorial derivation of groups' plurality status.

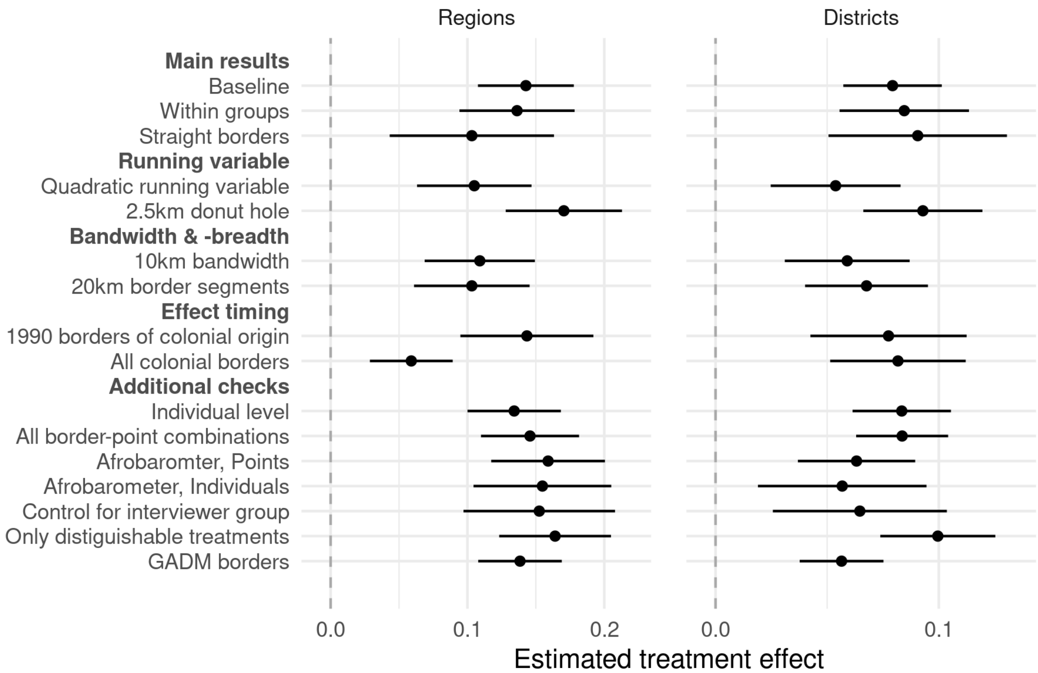


Figure 5: Summary of results from varying specifications.
Note: All results are available in detail in Appendix D.

Robustness tests

This section presents the main robustness checks summarized in Figure 5 and discussed further in Appendix D. Unless noted otherwise, 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. Most likely, this stems from noise in the spatial attribution of survey clusters to administrative units. To examine however a conservative scenario in which these non-linear dynamics are indeed real, I control for a linear and quadratic trend towards the border in treatment and control groups. As expected, this decreases the estimated effect of regional (district) borders by 4 (2) percentage points.

I also examine a liberal scenario in which deviations from the linear trend are only due to measurement error. I do so by estimating a 'donut'-RDD in which I drop all points closer than 2.5km[17] to the border. Doing so increases the estimated border effects by 2 to 3 percentage points. This difference shrinks with lower minimum distance cutoffs (see Appendix D.4). In sum, the baseline results appear to well-centered between the conservative and liberal estimates.

Bandwidth and -breadth: The spatial setup of the RDD entails two additional and potentially influential parameters, the first being the bandwidth of 20km. As shown in Appendix D.5, results remain mostly stable when subsetting the data to survey clusters closer to the border. As one decreases the bandwidth to 10km, effect estimates shrink to 0.07 percentage points but remain statistically significant. Yet, even with a smaller bandwidth the model relies on variation between potentially distant survey clusters located at opposite ends of a long border. To avoid such cases, I define sub-border segments such that points on its two sides lie within a convex hull of 20km diameter. Estimating Eq. 1 with a fixed effect per border-segment limits variation to small geographic areas but does not change the results substantively.

Effect timing: The use of contemporary data on administrative borders and ethnic identities entails the risk that results are purely driven by post-colonial dynamics particularly affected by border change (Grossman and Lewis 2014), instead of a capturing an effect accumulated since the colonial period. To isolate the latter, I subset the analyzed borders to those that align with colonial borders. This yields very similar effect estimates. Estimating effects across *all* colonial borders, even those that disappeared, leads to positive and statistically significant, yet smaller effects. Lastly, I assess heterogeneity across age cohorts. While assimilation and emigration likely lead to an accumulating and increasing effects over time, we should see no effect among respondents born in the 1950s if the results are driven solely by the (late) post-colonial period. As shown in Appendix D.6, I find that the results are robust even among respondents born in the 1940s.

[17] This is half the displacement radius of 99% of all rural DHS clusters.

Additional robustness checks: Further testing the stability of the results, I find that analyzing individual- rather than cluster-level DHS data does not affect the results. These are also consistent when using geocoded ethnicity data from the Afrobarometer (2018) Surveys rounds 1–6. The Afrobarometer also enumerates the main language of interviewers who may be sent to co-ethnic administrative units, thus creating social desirability bias where respondents want to appear co-ethnic to interviewers from the local plurality group. Such bias appears absent as results are robust to controlling for interviewers' co-ethnicity with units' predominant group. Dropping observations where DHS's ethnic categories do not distinguish between the two plurality groups across a border increases effect estimates.[18] I lastly analyze the DHS data with borders from GADM, an alternative dataset from 2019, and conduct a country-by-country jackknife (Appendix D.11). Both tests show stable results.

In sum, the main results are robust to permutations of the research design. The following analysis of assimilation and ethnic migration patterns further supports the theoretical argument by providing evidence on the micro-foundations of the plurality group's increasing share at administrative borders.

Mechanisms: Assimilation and migration

The regression discontinuity estimates evidence discontinuous changes in the ethnic makeup of local populations 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.

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 fre-

[18] 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.

quently observed in immigrant populations (e.g., Fouka 2019).

Shedding light on the first type of assimilation, the Afrobarometer surveys enumerate 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 an important step in the assimilation process (Cohen and Middleton 1970).

Ethnic assimilation also occurs across generations through marriage between local minority and majority members that increase their childrens' identity choice set (Cohen and Middleton 1970; Goody 1969; 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.[19]

For these three measures of assimilation, Table 2 implements the baseline RDD specification but adds a dummy variable for whether a respondents' ethnic self-identification in the Afrobarometer data and language in the DHS data corresponds to the local plurality (Plur. group member) in interaction with the treatment dummy. The treatment dummy then captures the change of the respective outcomes *among ethnic minorities* at region and district borders.

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. Similarly, I observe an increase of listing the plurality language as minorities' main language of 6.6 (2.8) percent points, though the latter estimate is statistically insignificant. These estimates show that minorities tend to linguistically assimilate to units' predominant groups.[20]

Models 3 and 6 find similar patterns of interethnic marriage. Female minority members are 1.8 (2.6) percentage points more likely to be married to a region's (district's) plurality group member barely inside the unit than barely outside it. Controlling for the share of the plurality group in the local population shows that

[19] Appendix Table A5 shows similar results for men and their first spouses.

[20] Additionally, Appendix Table A3 shows that plurality status increases identification with the group vs. the nation.

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

this effect can be explained by the post-treatment change in the 'supply' of plurality men at the border (Appendix E). Taken together, the analysis evidences that the shaping of ethnic groups by administrative borders partially works through the assimilation of local minorities.

As a caveat, 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 non-migrant minority members. Migrants, a potential 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 minority 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. Disposing only of between-unit migration data, I can only provide evidence on ethnic sorting through migration between administrative units.

To assess ethnically 'biased' migration patterns across administrative units, I rely on census data samples from eleven countries in Sub-Saharan Africa[21] 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 derive their plurality group as the group from Murdock's (1959) map with the largest spatial intersection. The ethnic linkage data from Müller-Crepon, Pengl and Bormann (2021) provide the link between IPUMS' ethnic labels and Murdock's group names, necessary to code groups in the census as local plurality and minority groups.

With the resulting dataset, I conduct three analyses (Table 3). I first estimate the extent to which local plurality status reduces individuals' emigration 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. I then analyze whether local plurality status increases 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 a region (district) dominated by their ethnic kin. Both patterns are identified in the presence of fixed effects that account for ethnic group and administrative unit characteristics.

I end this assessment of spatial ethnic sorting with a full dyadic analysis of migration patterns in Models 3 and 6. Here, the unit of analysis is the ethnic group

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

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. 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: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

nested in directed birth to residence unit dyads. The outcome consists in the share of an ethnic group in a birth unit that has migrated towards a given residence unit. Controlling for dyad and ethnic group fixed effects, the two main predictors assess the degree to which plurality groups differentially migrate between administrative unit. The results coincide with those discussed above. 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 low dyadic baseline migration rates, these effects are substantive and robust to different specifications (Appendix F.1).

Local minority members are thus more likely to exit, and migrants preferentially migrate towards co-ethnic administrative units. In a last step, I test whether these migration patterns indeed relate to the effects of administrative borders on ethnic demographics. To do so, 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 then take these unit-level measures of ethnically differential emigration and immigration and add them in an interaction term to the main RDD

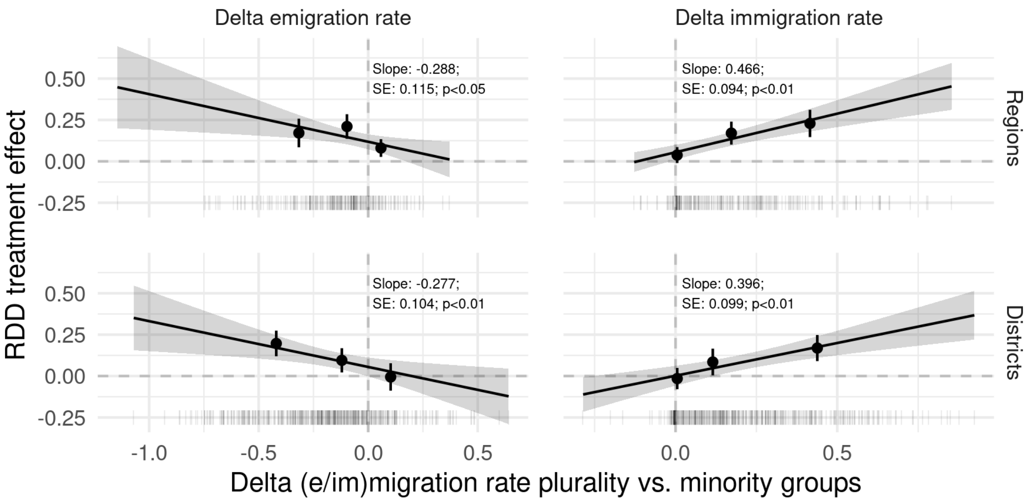


Figure 6: Effect of borders on ethnic identities increase in ethnic differences in e- and immigration.

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

specification (Eq. 1).[22] The resulting interaction effect captures the correlation between the extent of ethnically biased migration and the degree to which the share of a unit's plurality ethnic group increases at its border.

The results in Figure 6 show that the effect of administrative borders increases with (1) the extent to which local plurality members emigrate less from a unit (*Delta emigration rate*) and (2) the differential rate at which migrants immigrate into units in which they are part of the dominant ethnic groups (*Delta immigration rate*). For example, an 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 at the border is not causally identified, this result suggests that ethnically biased migration contributed to the crystallization of ethnic identities along administrative borders.

[22]For consistency, I here couple the DHS data with IPUMS' border data to estimate the RDD.

Conclusion

John Iliffe (1979) argued that "Europeans believed Africans belonged to tribes; Africans built tribes to belong to." This paper has analyzed the effect of territorial rule through administrative units as an important mechanism behind this argument. Colonial governance introduced inflexible administrative borders to the continent that often cut across ethnic geography. Local governance was often 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 locally predominant ethnic group or emigrating.

The empirical analyses provide empirical evidence for this account of the administrative shaping of ethnic groups. I find sharp spatial discontinuities in ethnic demographics at administrative borders. In particular, the share of regions' (districts') predominant ethnic group increases by about 14 (8) percentage points at the border to a unit dominated by a different group. This effect increases in the strength of traditional institutions and the relative size of the dominant group. Additional analyses support the argument that ethnic assimilation and emigration of minorities drive this phenomenon.

Taken together the argument and evidence presented in this paper offers an instrumentalist interpretation of constructivist accounts of the colonial transformation of ethnic identities in Africa. Ethnic identities and geographies are not prehistorically given but politically shaped inside political arenas delimited by administrative borders. Once conceived, ethnic assimilation and migration patterns left ethnic identities crystallized along administrative borders, contributing to the alignment of administrative units and ethnic geographies. These patterns highlight the administrative shaping of ethnicity, one of the most important political cleavages in African and around the globe.

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 Precolonial Africa." *Présence Africaine* 127:43.

Bandyopadhyay, Sanghamitra and Elliott Green. 2021. "Explaining inter-ethnic 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.

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.

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.

Chimhundu, Herbert. 1992. "Early missionaries and the ethnolinguistic factor during the 'invention of tribalism'in Zimbabwe." *The Journal of African History* 33(1):87–109.

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.

Doyle, Shane. 2009. "Immigrants and indigenes: The Lost Counties Dispute and the evolution of ethnic identity in colonial Buganda." *Journal of Eastern African Studies* 3(2):284–302.

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.

Goody, Jack. 1969. Marriage Policy and Incorporation in Northern Ghana. In *From Tribe to Nation in Africa: studies in incorporation processes*, ed. Ronald Cohen and John Middleton. Berkeley: University of California Press pp. 114—–149.

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. 2020. "Ethnicity, national identity and the state: evidence from sub-Saharan Africa." *British Journal of Political Science* 50(2):757–779.

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.

Henn, Sören. 2021. "Complements or Substitutes? How Institutional Arrangements Bind Chiefs and the State in Africa." *Unpublished Working Paper. http://soerenhenn.com/files/Henn_Chiefs.pdf* .

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.

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.

Imbens, Guido and Karthik Kalyanaraman. 2012. "Optimal bandwidth choice for the regression discontinuity estimator." *The Review of economic studies* 79(3):933–959.

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.

La Fontaine, Jean Sybil. 1969. Tribalism among the Gisu. 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 pp. 177–192.

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

Laitin, David D. 1994. "The Tower of Babel as a coordination game: political linguistics in Ghana." *American Political Science Review* 88(3):622–634.

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 I Pengl and Nils-Christian Bormann. 2021. "Linking Ethnic Data from Africa (LEDA)." *Journal of Peace Research, OnlineFirst* .

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

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* .

Sambanis, Nicholas and Jonah Schulhofer-Wohl. 2009. "What's in a line? Is partition a solution to civil war?" *International Security* 34(2):82–118.

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.

Touval, Saadia. 1966. "Treaties, borders, and the partition of Africa." *The Journal of African History* 7(2):279–293.

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

Wagner, Günter. 1949. *The Bantu of North Kavirondo*. London, New York: Published for the International African Institute by the Oxford University Press.

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.

Wilks, Ivor. 1975. *Asante in the ninetheenth century. The structure and evolution of political order.* Cambridge: Cambridge University Press.

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.

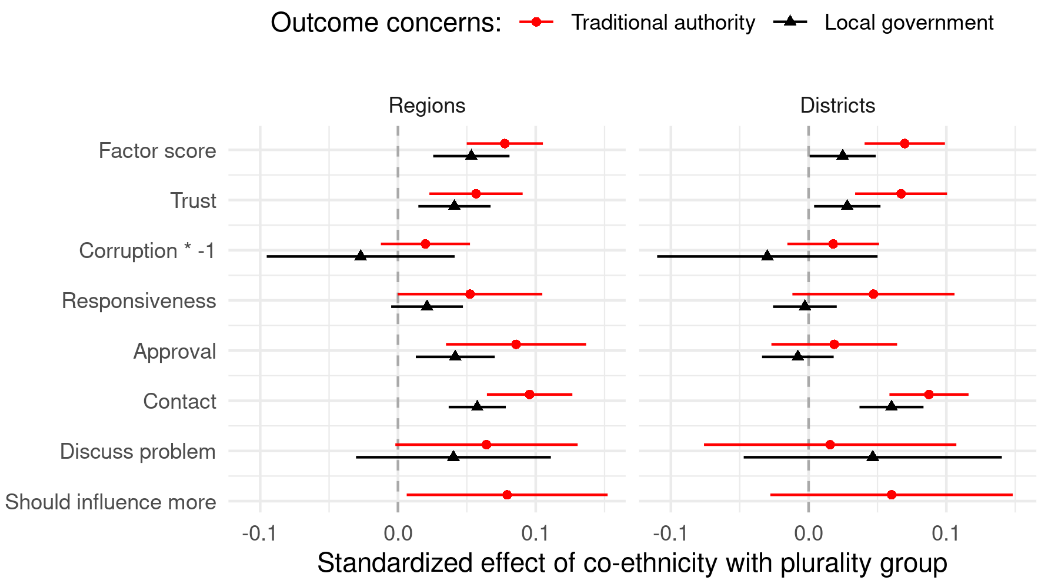


Figure A1: Increased interaction between local plurality members and traditional authorities as well as local governments.

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	yes	2.1, 4.1, 6.1	5, 6	
Chad		7.1		
Côte d'Ivoire	yes	3.1, 3.2, 6.1	5, 6	
Ethiopia		4.1, 5.1, 6.1, 7.1		
Gabon		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		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		5.1, 6.1, 7.1	2, 3, 4, 5, 6	
Namibia		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
Swaziland		5.1	5, 6	
Tanzania	yes		1, 2, 3, 4, 5, 6	
Togo		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.

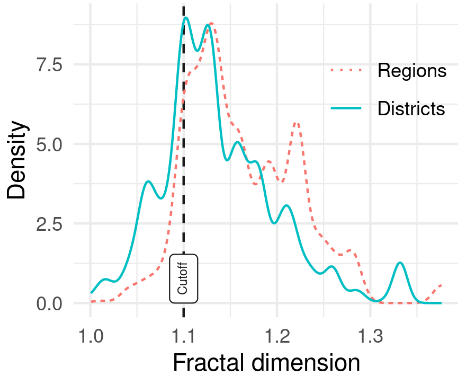


Figure A2: Distribution of fractal dimension values among *borders* in the baseline analysis. Each border receives equal weight.

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 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.[23]

C Regression discontinuity analysis: Descriptive statistics

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

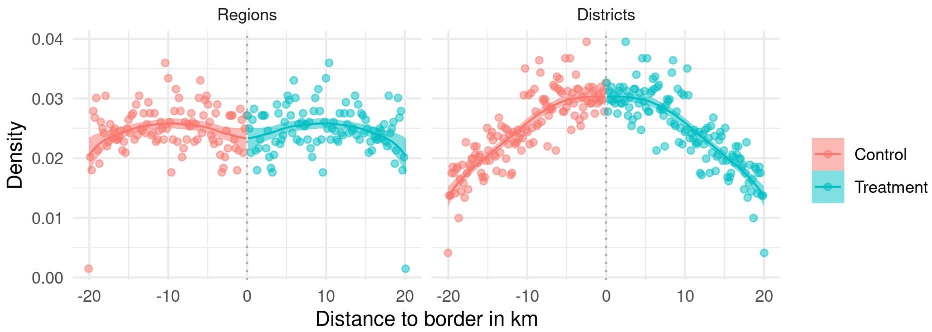


Figure A3: McCrary (2008) test.

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

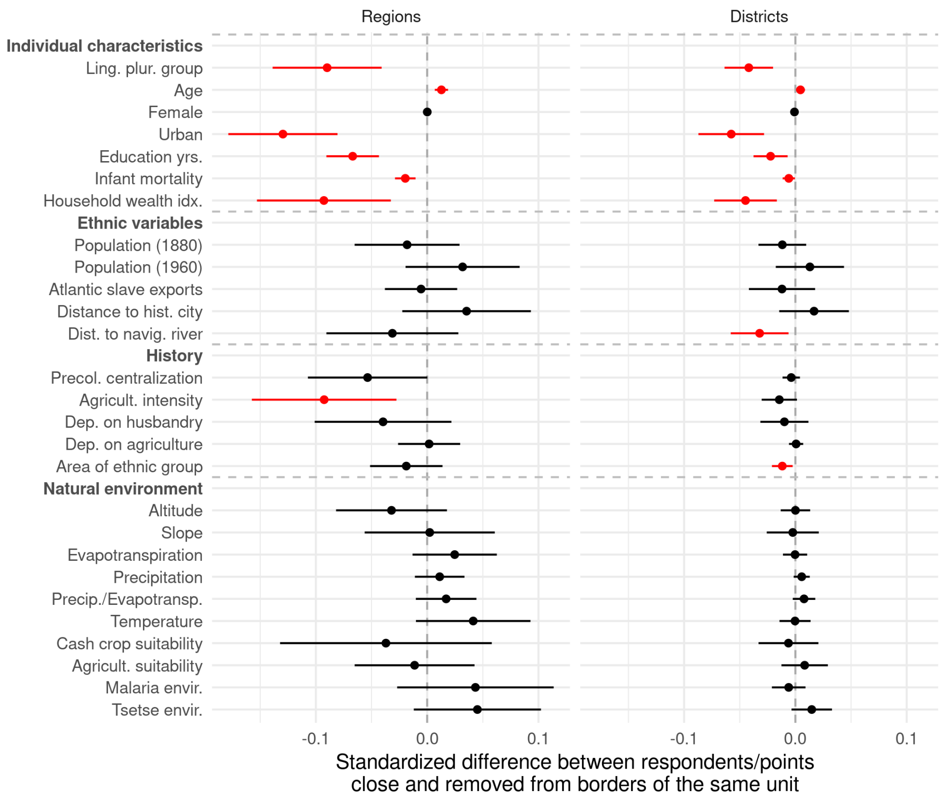


Figure A4: Comparison of DHS respondents and survey clusters close (<20km) to borders that divide units with differing plurality ethnic groups with average observations from the same units.

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 point-level aggregates from the DHS.

D 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 A5 combines the coefficients from most analyses discussed below.

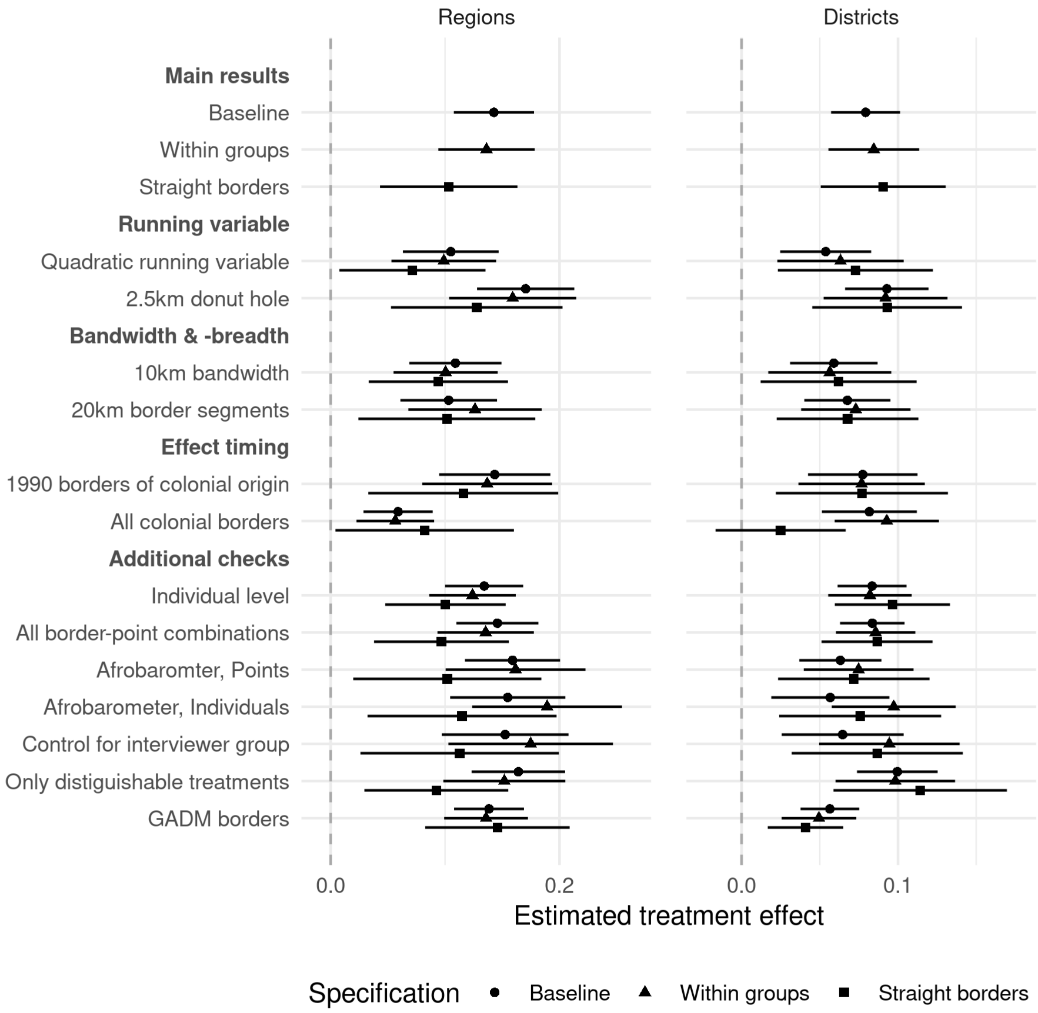


Figure A5: Summary of results from varying robustness checks.

D.1 Comparison with one-sided RD Design

In the two-sided RD Design each point 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 to points on each 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

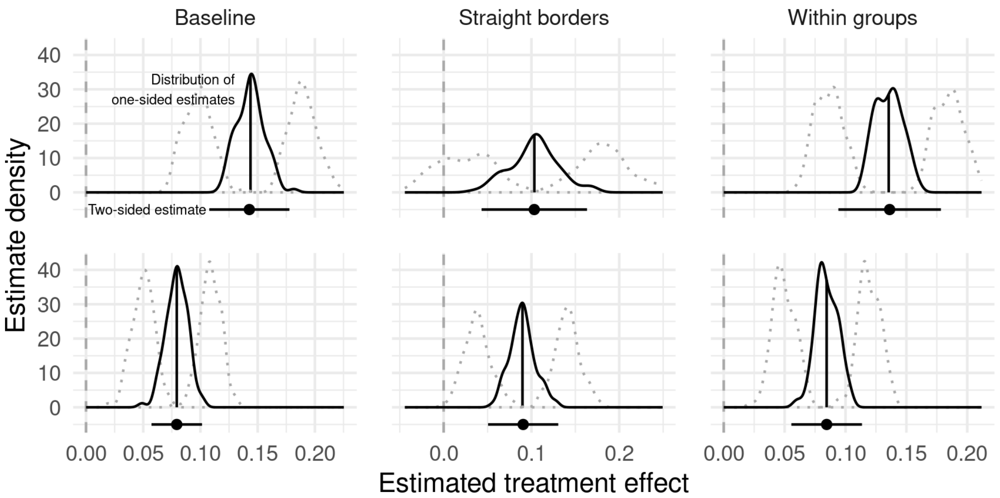


Figure A6: Estimate distributions from one-sided design

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

distributions in Figure A6. 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 colonial district borders, highly statistically significant as well.

D.2 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 A7 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.[24] 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

[24]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.

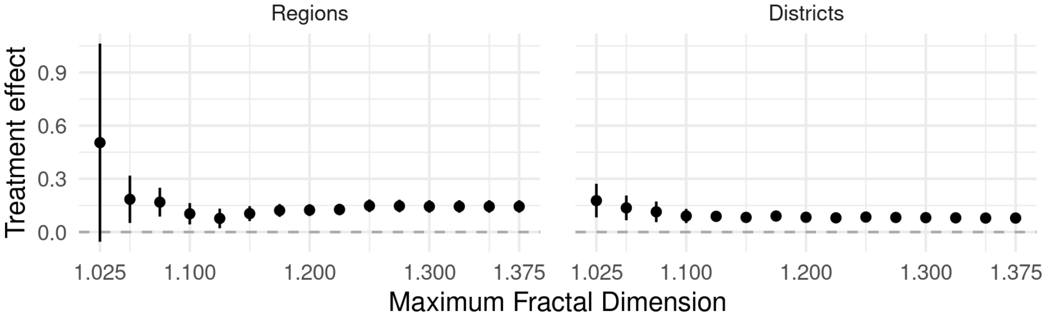


Figure A7: Estimate of border effect on local ethnic identities with different fractal dimension cutoffs.

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).

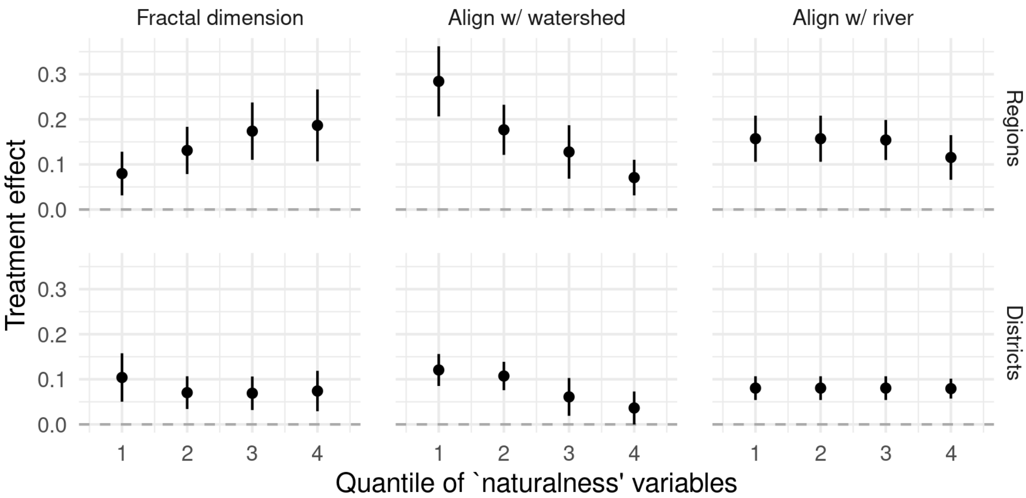


Figure A8: Estimate of border effect on local ethnic identities by degree of naturalness of border.

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.

presented in Figure A8. 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.

D.3 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.4 Running variable

Because a visual inspection of the data indicates a non-linear trend very close to the examined borders (below 5km), Figure A5 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.

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 points 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 A9 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.

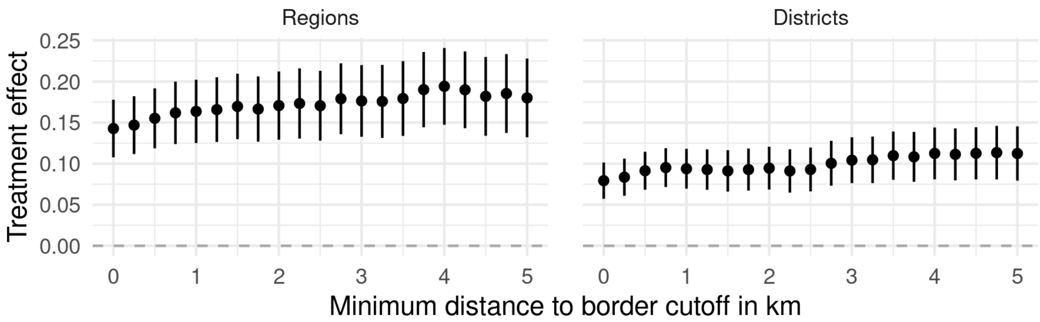


Figure A9: Estimate of border effect on local ethnic identities by size of 'donut'-hole around the border.

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

D.5 Variation in bandwidth and -breadth

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 A10, 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 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 within clusters of points of a defined diameter, I define sub-border segments such that points on both sides of the same border have a maximal distance to each other of between 20 and 100 km with each point being

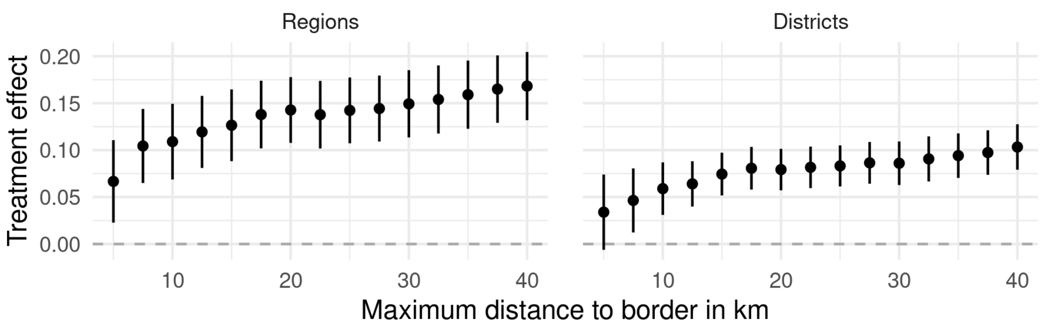


Figure A10: Estimate of border effect on local ethnic identities by bandwidth.

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

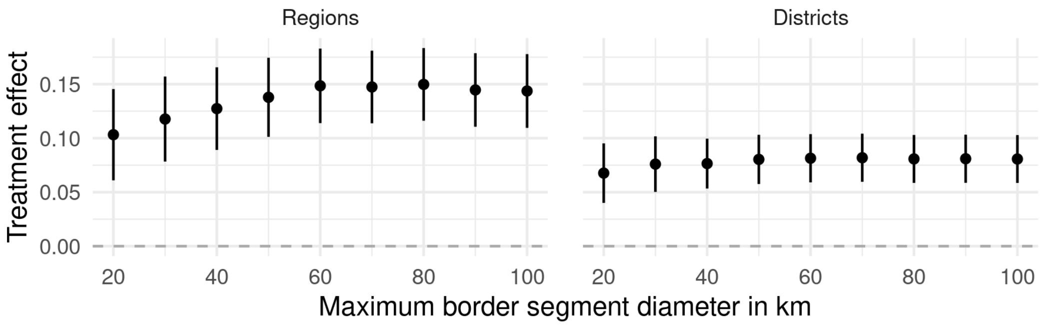


Figure A11: Estimate of border effect on local ethnic identities by maximum diameter of point clusters along border segments.

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

closer than 20km to the border. The identification of these spatial clusters of points is implemented via a hierarchical clustering algorithm that maximizes the size of the sample kept in the analysis.[25] 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 A11, doing so does not change the results in a substantive way with regional border estimates slightly decreasing with a cluster diameter below 60 km and stable beyond that.

D.6 Effect timing

To isolate effects that have accumulated since the colonial period rather than stem from only the (late) postcolonial period, I subset 1990 borders to those that align with colonial borders, analyze the effect of *all* colonial borders, and assess treatment heterogeneity by decade of birth of DHS respondents. The colonial-era data on administrative borders comprises regional borders from the year of countries' independence and has been produced by using information on regional border changes (i.e., mergers and splits) from `statoids.com` to backdate the GAUL database. Such information is not available in a comprehensive manner on districts, which is why I rely on district shapes from French and British colonies in western and eastern Africa published by Huillery (2009) and Müller-Crepon (2020).

I subset colonially-aligned borders from 1990 by measuring the average difference of points along them to the set of colonial borders at the corresponding regional or district level. With the resulting smaller sample and as plotted in Figure A5 above, I estimate very similar effects than in the main specifications.

Because this strategy disregards borders that have (potentially endogenously) disappeared, I also re-estimate treatment effects across *all* colonial borders, even those that disappeared. The main downside of this approach is that the colonial district data in particular cover only part of the continent, thereby limiting the number of borders and survey clusters in the sample. In addition, knowledge on the location of historical borders is relatively scarce, thus increasing noise and uncer-

[25]This is necessary as an arbitrary segmentation of borders creates many segments with points only on one side of the border, segments which are dropped because they add no information.

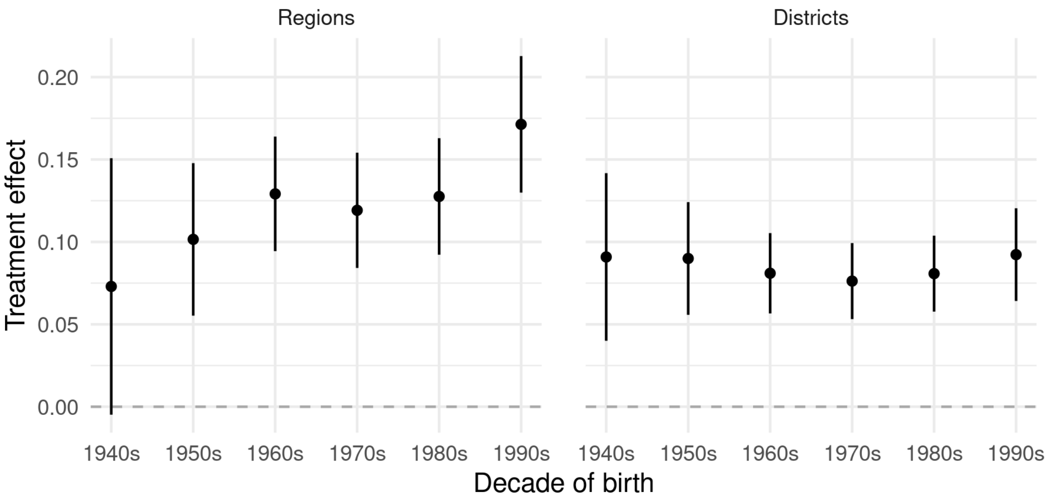


Figure A12: Border effect estimates by birth decade.

Note: Estimation follows the baseline specifications in Models 1 and 4 in Table 1 in the main text.

tainty. However, the results for colonial-era regional and district borders presented in Figure A5 coincide with the main estimate. This is with the exception of straight district (but not regional) borders, where the estimate shrinks and becomes statistically insignificant. This deviation is most likely due to the severely restricted sample, with only 4500 points along 100 borders as compared to 9300 points along 500 borders in the main analysis. The estimates of the effect of regional borders are generally smaller than in the baseline analysis. This is most likely due to the upgrading of ethnically distinct districts to regional status over time.

If all effects are solely driven by (late) post-colonial dynamics in the 1980s and 1990s we should see no effect among cohorts born and raised much earlier in the 1940s and 1950s. Analyzing effect heterogeneity by birth decade, Figure A12 shows that this is not the case. Instead, regional (but not district) border effects slightly increase over time but are substantial even for individuals born in the colonial period.

D.7 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 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 A5 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 points 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 A5 show that doing so does not change the results.

D.8 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 (2021). 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 A5 above closely coincide with those obtained from the DHS data.

D.9 Testing for data-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.[26] 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 D.8), and use information on interviewers' home language to code whether s/he belongs to the plurality ethnic group of the treatment unit.[27] I then re-estimate the main model from Eq. 1 adding this dummy variable as a control variable. While I find that the interviewers' membership in the plurality group is positively related to that of respondents[28], this correlation does not affect the estimated discontinuous jump in ethnic identities at district and regional borders (see Figure A5).

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

[26] For political interviewer effects driven by coethnicity between interviewers and respondents see Adida et al. (2016).

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

[28] This likely reflects practical considerations that lead to geographically non-random assignment of interviewers.

attributed to both.[29] 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 A5) closely align with the baseline estimates, indicating that such biases are not severe.

D.10 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 A5 above shows, the results closely align with the baseline estimates.

D.11 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 A13 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.

[29]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.

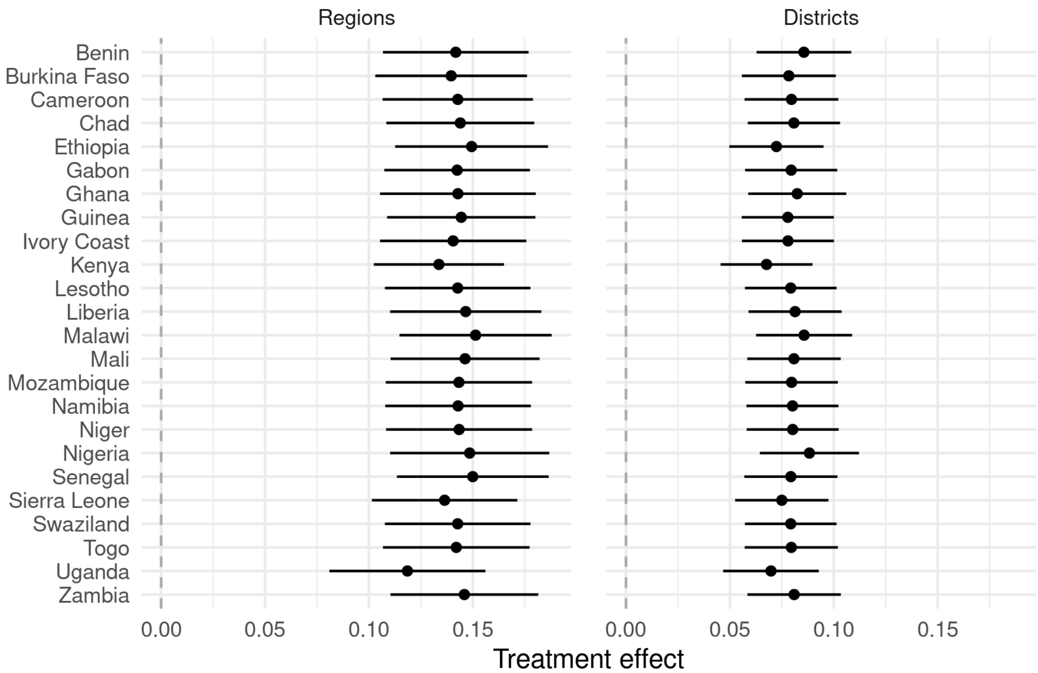


Figure A13: Estimate of border effect on local ethnic identities, dropping one country at the time.

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

E Ethnic assimilation: Additional results

This section presents additional results on ethnic assimilation and integration only briefly mentioned in the main paper. First, Table A3 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 A4 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 A4 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 A5 shows very similar dynamics among the spouses of men interviewed by the DHS.

Table A3: Local plurality status and ordinally scaled strength of ethnic (1) versus national (5) identification

	Regions	Districts
	(1)	(2)
Plurality co-ethnic	-0.061^{***}	-0.041^{**}
	(0.016)	(0.017)
Adm. unit FE	yes	yes
Lang. group FE	yes	yes
Covariates	yes	yes
Mean DV:	3.6	3.7
Observations	99,968	80,406
Adjusted R^2	0.150	0.158

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: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A4: (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 A5: (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

F 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 F.1) and then discuss the correlation of ethnic migration patterns with the effect of administrative borders on ethnic identities (Subsection F.2).

F.1 Robustness checks

Figure A14 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 A15 shows very persistent patterns over time 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 A16 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.

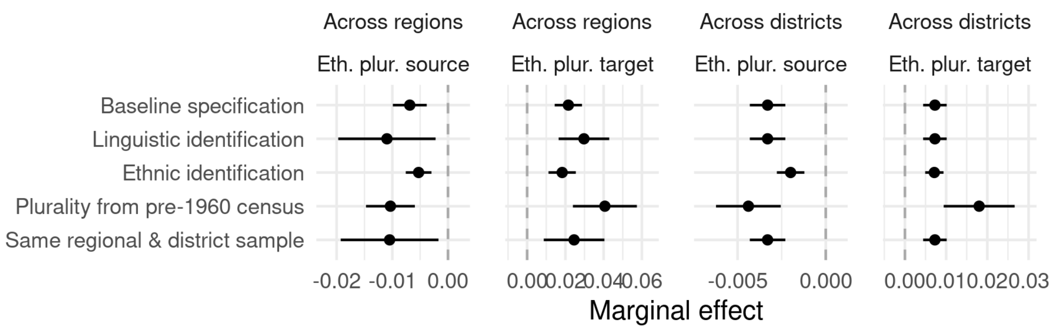


Figure A14: Dyadic migration analysis: Additional results.

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

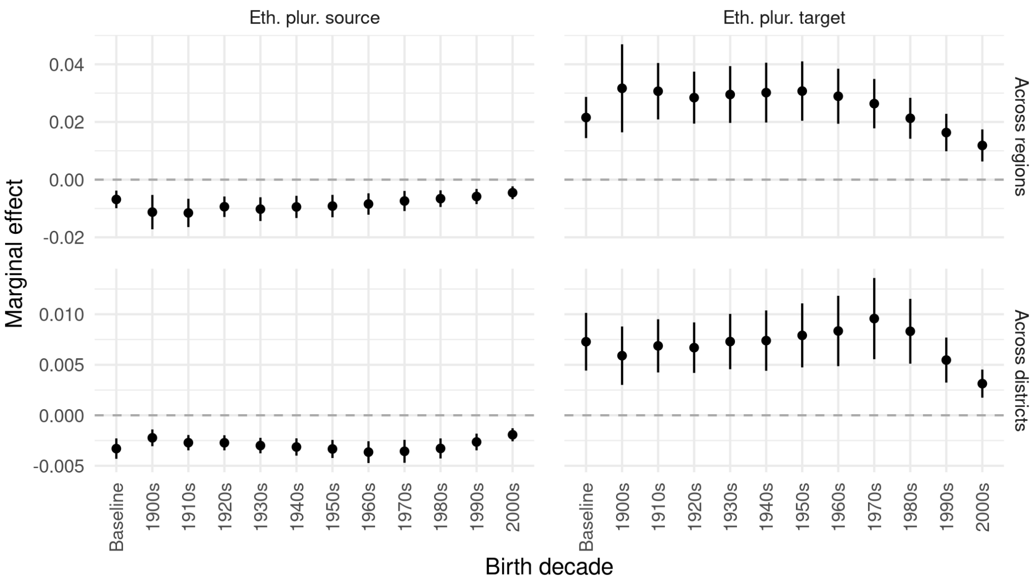


Figure A15: Dyadic migration analysis by birth decade

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

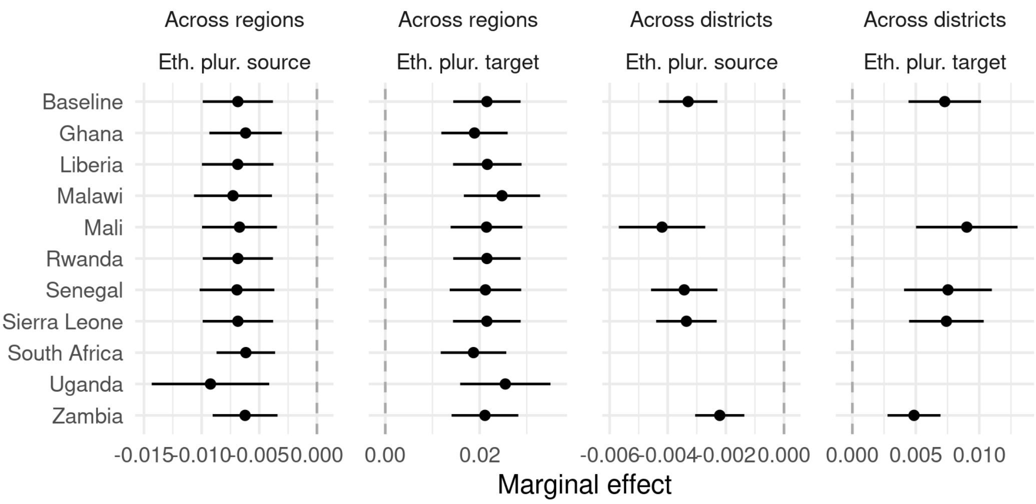


Figure A16: Dyadic migration analysis: Country-level jackknife.

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

F.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

Table A6: 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)
treat	0.153***	0.118***	0.055**	0.091***	0.055*	0.002
	(0.023)	(0.023)	(0.023)	(0.027)	(0.029)	(0.031)
I(treat *emigration)		−0.288**			−0.277***	
		(0.115)			(0.104)	
I(treat *immigration)			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

Models 1 and 3 in Table A6, the respective results align with the main results at 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:

$$Emigration_{u,e} = \alpha_u + \gamma_e + \Delta_{u,emigr.}\, \mathbb{1}(plurality_{e,u}) + \epsilon_{u,e} \tag{2}$$

and

$$Immigration_{u,e} = \alpha_u + \gamma_e + \Delta_{u,immigr.}\, \mathbb{1}(plurality_{e,u}) + \epsilon_{u,e}, \tag{3}$$

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.[30] 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

$$\begin{aligned} Y_{p,b,t,s} =& \alpha_{b,t} + \gamma_s + \beta_1 T_{u,t} + \beta_2 T_{u,t}\Delta_u + \beta_3 D_{p,b} + \beta_4 D_{p,b} T_{u,t} + \\ & \beta_5 D_{p,b}\Delta_u + \beta_6 D_{p,b} T_{u,t}\Delta_u + \epsilon_{p,u,b,t}, \end{aligned} \tag{4}$$

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

[30]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.

of the border effect with the migration bias.

This analysis is carried out in Models 2-3 and 5-6 in Table A6 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.

G 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* .

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.

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, Yannick I Pengl and Nils-Christian Bormann. 2021. "Linking Ethnic Data from Africa." *Journal of Peace Research, OnlineFirst* .

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