# Building Tribes: How Administrative Units Shaped Ethnic Groups in Africa

Carl Müller-Crepon*

University of Oxford

January 14, 2021

**Abstract**

Ethnic identities around the world are deeply linked to the modern territorial state, yet the extent to which states have shapes and still shape ethnic groups is empirically unclear. I argue that governments at the national and subnational level have incentives to bias governance in favor of the largest ethnic groups in their territory. The resulting disadvantages for ethnic minorities can motivate minority assimilation and emigration. Both gradually align ethnic with administrative boundaries. I examine this process at the subnational level across Sub-Saharan Africa. Exploiting credibly exogenous, straight borders allows for causal identification. I find substantive increases in local population shares of administrative units' predominant ethnic groups at the border, showing that administrative geographies shaped ethnic groups. Additional analyses demonstrate that ethnic assimilation and emigration of local minorities drive the phenomenon. 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.

## Introduction

Ethnicity constitutes one of the globally most salient political cleavages. It leaves its imprint on important political processes in multiethnic countries, affecting public goods provision (Alesina, Baqir and Easterly 1999; Alesina and Ferrara 2005), redistribution (Franck and Rainer 2012; De Luca et al. 2018), and violent conflict (Horowitz 1985; Montalvo and Reynal-Querol 2005; Cederman, Gleditsch and Buhaug 2013). While it is well recognized that ethnicity and ethnic boundaries are socially constructed (Barth 1969; Wimmer 2013), less is known about their relation to modern state governance. While substantive qualitative evidence shows how state building shaped ethnic identities not only in Europe (Weber 1977) but also in Africa (Southall 1970; Iliffe 1979; Young 1985), there is only sparse systematic evidence on the transformative effect of modern statehood on ethnic identities.

This paper addresses this gap and examines how the introduction of territorial governance through neatly bounded administrative divisions shaped ethnic identities in Sub-Saharan Africa. I argue that local and regional governments tend to bias governance in favor of the largest ethnic groups in their territory. This incentivizes disadvantaged local ethnic minorities to respond by assimilating to the majority identity or emigrating to co-ethnic governance units. The resulting change in ethnic demography crystallizes ethnic boundaries along administrative borders and increases the congruence between ethnic and administrative geographies.

I study this process in Sub-Saharan Africa where colonial rule imposed territorial governance and administrative borders without much respect for preexisting ethnic geographies. Local governance was nevertheless rapidly ethnicized, in particular by drawing on partly invented 'customary' practices and institutions. I argue that the resulting disenfranchisement of local minorities triggered their assimilation and/or emigration, reactions that over time crystallized ethnic boundaries along administrative borders. I find evidence for my argument in discontinuities of local ethnic demographics along credibly exogenous administrative borders, as well as in data on ethnic assimilation and migration of ethnic minorities. In effect, my findings describe one important mechanism behind Iliffe's (1979, p. 324) statement that "Europeans believed Africans belonged to tribes; Africans built tribes to belong to."

Analyzing the transformation of ethnic identities in the context of colonial and post-colonial state building. I particularly stress the instrumental role of ethnicized governance as a driver and administrative boundaries as delimiters of processes of ethnic identity transformation. Contributing to our understanding of the foundations of (political) ethnicity in Africa (Green 2020, 2018; Michalopoulos 2012; Pengl, Roessler and Rueda 2020; Posner 2004; Robinson 2017), this study offers evidence on how territorial rule imposed by colonialism fostered the "invention of tradition" (Ranger 1997) and transformed ethnic identities (Lentz 1995; Southall 1970; Young 1985).

The conceptual parallels between 'nationalism' and 'tribalism' highlight the analogy between subnational 'tribe-building' in Africa and nation-building around the globe (Argyle 1969; Berman 1998). In Europe, the rise of modern states was inextricably linked to the nationalist striving for territorial self-governance of ethnic groups. This involved the 'right-shaping' and 'right-peopling' of states, thus ethnically homogenizing polities (O'Leary 2001, also Brass 1991; Mann 2005; Wimmer 2018). I characterize a very similar process at the subnational level in the multiethnic states of Sub-Saharan Africa, focusing particularly on the role of administrative borders and their structuring effects on ethnic geography. Studying the effects of often haphazardly drawn African borders is important in its own right and circumvents inference problems associated with endogenous borders elsewhere.

I test my argument that administrative units shaped ethnic groups by estimating the change in ethnic geography at units' boundaries. Employing a regression discontinuity design, I find that the share of regions' (districts') main ethnic group sharply increases by 10 (6) percentage points at their borders with units dominated by a different group. These effects are robust to estimating changes at borders that run within the settlement area of the same ethnic group as mapped at the end of the colonial period, as well as at relatively straight, arbitrarily drawn subnational boundaries.

I provide evidence that assimilation and emigration patterns among local ethnic minorities account for the effects of administrative borders on local ethnic demography. I find that local minorities assimilate through language adoption and intermarriage with members of locally predominant ethnic groups. In addition, an analysis of migration patterns of 33 million individuals recorded in census data evidence ethnic sorting across administrative units: local minorities exhibit higher rates of emigration and lower levels of immigration than local majority groups.

Providing evidence on how territorial governance shaped the transformation of ethnic identities in Africa, the paper highlights the endogeneity of ethnic identities as a larger issue for the study of identity politics and the effect of ethnicity. I demonstrate that ethnic identities are, at least in the long run and within (unknown) limits, a result of state governance. This root of ethnic identities raises the crucial question of when, where, and how else political elites and citizens foster ethnic change to warp the political playing field to their own favor (e.g. Brass 1991; Posner 2004).

## Theoretical argument

Governance through territorially defined administrative units transforms ethnic groups because local governments tend to be biased against local ethnic minorities. Such bias arises where governments ethnically specialize their governance to cater goods to, reduce communication barriers with, and foster popular loyalty among a subset of their subjects, often the largest ethnic group. In response to their diminished political status, local minorities can assimilate to the majority identity or emigrate. Because local governance and its ethnic biases are spatially restricted by the borders of administrative units, the resulting transformation of ethnic groups aligns ethnic boundaries with administrative borders. The following sections sketches the details of this argument.

### Ethnicized territorial governance over heterogenous populations

The introduction of territorial governance, that is governance within contiguous, non-overlapping geographic units, typically creates multiethnic polities. This is because ethnic heterogeneity at the smallest perceivable geographic level (say the household or village) makes it impossible to devise monoethnic polities. Where polities are instead defined along ethnic, religious, or clan lines, they can be monoethnic.[1] Governments of multiethnic societies frequently favor large, powerful ethnic groups and are biased against smaller minorities. Their ethnic favoritism can have two sources, one rooted in the ethnic identities of executive leaders and the other in an endogenous 'ethnic specialization' of governments.

The first and most extensively analyzed cause of ethnic favoritism is that governments cater goods and services to 'their' ethnic constituencies (e.g. De Luca et al. 2018; Franck and Rainer 2012). For one, leaders may simply intrinsically prefer to see their ethnic group prosper (Chandra 2007). For another, they can more profitably exchange political support for material benefits with their co-ethnics, with whom they have lower transaction costs due to a common culture and dense social ties that facilitate enforcement (Eubank 2019; Fearon 1999; Larson and Lewis 2017). Because large ethnic groups hold, on average, most executive power (Bormann 2019; Francois, Rainer and Trebbi 2015), this form of ethnic favoritism disadvantages small ethnic groups.

A second source of ethnicized governance emerges from governments' incentive to lower the transaction costs of governance by 'specializing' in the largest ethnic group among their subjects. High transaction costs, caused for example by language barriers (Laitin 1992), increase the costs of tax collection and government services provided in return. Facing a multiethnic population, governments can choose the degree to which they tailor their activities to each ethnic group. If there are economies of scale to ethnic specialization, the costs of specialization increase less in the size of ethnic groups than its benefits. Thus, specializing in small groups yields smaller (and possibly negative) net benefits than specializing in the large group(s) among their citizens. Governments therefore specialize most in governing the largest ethnic group.

Tailoring governance towards the largest ethnic group may also have an ideological, 'ethno-nationalist' or 'tribalist,' component that can foster governments ability to rule. In particular, rulers can leverage references to ethnic traditions that legitimize their power and increase co-ethnics' "quasi-voluntary compliance" (Levi 1988) through a shared identity and informal norms. The effect is, again, a facilita-

[1] Obviously, such governance confronts much different problems, in particular the production of geographically located public goods and services in ethnically diverse areas.

tion of governance.

Similar incentives to those that cause governments' bias towards the largest ethnic group also lead governments to ethnically homogenize minority populations (Alesina and Spolaore 2005). Using education policies in particular, governments can "re-educate" members of 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). Such ethnic engineering is, however, far from uncontroversial and often causes backlash among ethnic minorities (Hechter 1975; Fouka 2019). At the extreme, some governments violently 'right-people' their population through displacement and genocide (O'Leary 2001; Mann 2005).

In sum, governments are ethnically biased either because of their own ethnic identity or to lower transaction costs with the largest groups among their populace. Ethnic minorities in turn end up disenfranchised and under-serviced and may be targeted by government interventions that aim to ethnically homogenize their population.

### Individual responses to ethnicized governance

Ethnically biased governance can come at a steep cost for ethnic minorities. Disregarded by their government, minorities often end up with less access to state-provided goods, ranging from security and justice to public services such as education and health care. But members of minority groups may be able to change their fate by engaging in actions reminiscent of Hirschman's (1970) triad of voice, exit, or loyalty. Minorities ca protest and demands political representation or even the secessionist establishment of a new political unit, but they may also choose to emigrate to a polity where they are part of the ethnic majority or remain loyal to their government and assimilate to become part of the local majority.

Ethnic assimilation aims at changing one's language, religion, appearance, dress, and taste such that they can credibly claim benefits of membership in the majority group (e.g., Laitin 1995). Assimilation thus consists of a diverse set of choices that increase the degree to which one can 'pass' as a member of the majority group.

Because some salient group characteristics are learned during early childhood or are even innate as in the case of phenotypical ethnic markers, assimilation often proceeds as a intergenerational process.

Beyond assimilation, minority members can also exit their governance unit through migration (Docquier and Rapoport 2003; Hirschman 1970). Facing discriminatory governance at home, individuals may either ethnically sort into governance units where they belong to the ethnic majority or migrate towards economically more prosperous areas where government discrimination is offset by economic opportunity.

Assimilation and emigration by ethnic minorities increase the relative size of the largest ethnic group in a governance unit. In a manner reminiscent of Schelling's (1971) tipping point model, this will, ceteris paribus,[2] reinforce governments' ethnic biases, creating ever more incentives for minorities to assimilate or emigrate.

However, in a world where political boundaries can be contested, sufficiently large and spatially concentrated minorities may choose to mobilize and demand their own governance unit (Alesina and Spolaore 2005; Grossman and Lewis 2014; Weidmann 2009). However, because ethnic groups seldom live perfectly segregated, secession tends to create new minorities (Sambanis and Schulhofer-Wohl 2009). While certainly an important mechanism that contributes to an increasing congruence of political and ethnic boundaries (Grossman and Lewis 2014; Müller-Crepon, Schvitz and Cederman 2020), my focus will here be on *ethnic* change through assimilation and migration, leaving aside the determinants of border change.

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, assimilation and emigration processes among minority groups 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 different predominant groups.

[2]However, other ethnically *heterogenizing* processes happen in the meantime, in particular immigration driven by processes other than ethnic sorting.

## Administrative units and ethnicity in Africa

I test this argument using data on ethnicity and administrative units in Sub-Saharan Africa. The following section contextualizes these data and the cases they contain. I argue that European colonialists have drawn subnational borders with limited information and in disregard of local ethnic geographies, thus creating ethnically heterogeneous political units. Because local governance was soon ethnicized through the heavy reliance on partially invented 'traditional' authorities, local minorities had considerable incentives to assimilate or emigrate. By spatially structuring this process, the administrative partitioning of colonies molded ethnic identities.

I do not expect that the effect of territorial governance and administrative borders on ethnic identities is strictly limited to Sub-Saharan Africa or the subnational level. However, examining the subnational level of multiethnic countries in Africa is valuable not only in its own right, but for two additional reasons. First, African states are among the most ethnically diverse, most do not feature a clear majority group, and ethnic homogenization does not occur at the national level (Laitin 1992, 1994). Focusing on how regional borders 'cement' subnational majority identities contributes to explaining this pattern.[3] Second, the setting carries a significant empirical advantage. Because African borders were often haphazardly drawn (Herbst 2000; Michalopoulos and Papaioannou 2016), studying their effects minimizes the risk of reverse causality by which ethnic geographies have affected the spatial design of governance units in other world regions.

### The colonial introduction of administrative borders

Defining the state via its territory demarcated by borders is integral to the idea of modern statehood (e.g. Mann 1984; Weber 1919), but "was virtually unknown in most places in Africa during the period before the European partition" (Asiwaju 1983, p. 45). Similar to medieval Europe (Ruggie 1983), jurisdictions were non-contiguous, sometimes overlapping and sometimes separated by frontier zones (Herbst 2000; Wilks 1975). Political control was strong at states' core and radiated

[3] Relatedly, Ichino and Nathan (2013) observe how members of local ethnic minorities support the same candidates as their ethnic majority neighbors in Ghanaian national elections.

outwards. The virtual inexistence of precolonial maps only underscores the absence of demarcated boundaries (Herbst 2000). Evidently, the concept of 'borders' between political entities was even more foreign to acephalous societies where political power was not centralized to begin with. After all, the political boundary comes only to life as a separating line between political entities (Kristof 1959).

Colonization after the Berlin conference in 1885 radically changed this topography of state power. Not only did the European conquerors carve up the continent into colonial territories with fixed and inflexible borders. They also partitioned their colonies into administrative units, thereby rolling out the territorial governance that established 'effective control,' even if it was effective in name only. The creation of non-overlapping and neatly bounded administrative divisions – regions, districts, and subdivisions – was as much of a governance revolution as the drawing of international borders. Both introduced sharp lines that delimited the territorial scope of states and their administrative divisions (see Asiwaju 1983).

### Ethnic foundations of administrative geographies

The geographic design of governance units was crucial in defining their initial ethno-demographic composition. For British colonies, historians stress that the predominant administrative mindset expected every individual to belong to one tribe, defined as "discrete, bounded groups, whose distribution could be captured on an ethnic map" (Young 1985, p. 74).[4] With the British policy of declaring 'tribes' as 'natural' social units of local governance, tribal homelands were destined to become administrative units (Asiwaju 1970; Crowder 1968; Miles 1994; Spear 2003).

This idea clashed with a reality in which multiple groups often settled in the same environmental niche (Cohen and Middleton 1970, p. 11) and political loyalties were "complex, flexible and amorphous, sometimes overlapping, sometimes complementary, and did not add up to clearly demarcated tribes" (Lentz 1995, p. 317, Southall 1970). Unable to deal with such ambiguity, pragmatism coupled with administrative and geographic exigencies determined the precise location of borders (Lentz 2006, p. 53). Even though borders cut through ethnic geographies, eth-

[4]Notice the interesting parallel to the European ethno-nationalist discourse dominating the turn of the 19$^{th}$ century and, so Berman (1998), the minds of colonial officials.

nic labels were attached to the units they enclosed. On their basis, administrations then specialized in their population – for example, training staff in the dominant language, establishing ethnically differentiated policies,[5] or encoding customary law.

The historical case for French colonies is different for their more centralized governance. Although French administrations relied on local intermediaries as well, they tended to crush pre-existing institutions, replace them with more uniform institutions of their own making, and hand less power to local rulers (Cohen 1971; Conklin 1997; Crowder 1964; Müller-Crepon 2020). With colonial governance less ethnicized under French than British rule (Ali et al. 2018), French governments created administrative units that potentially aligned with precolonial ethnic geographies even less than the British did.

### Administrative geographies and the transformation of ethnicity

The creation of territorial governance units entailed a seminal change in the relationship between rulers and their people: the switch from governance based on 'ethnic' allegiances between people and their rulers to governance over a spatially demarcated population (Herbst 2000). But because ethnic identities and ties did not disappear overnight, local rulers oftentimes ethnicized governance over the population of 'their' governance unit. Members of local minorities thus confronted misaligned and discriminatory local governments. This constellation bore the seeds of the transformative effect of administrative divisions on ethnic groups.

Local ruling elites fostered the homogenization of local ethnic identity for their own good. 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.[6] Thus, the "invention" (Ranger 1997) of traditions and history became a tool for local elites' political survival in the colonial past (Iliffe 1979) and the postcolonial present (Robinson 2017). Political battles over the codification of 'customary' institutions and prac-

[5] E.g. the use of regionally varying forms of taxation.

[6] In addition to challenging ruler directly, challengers could threaten the integrity of local governance units by presenting themselves and their followers as homogeneous group that 'deserved' its own governance unit.

tices played out, for example, when Councils of Elders in the Taita Hills in Southern Kenya tried to control chiefs and "invented Taita traditions" that synthesized lineage practices (Bravman 1998, p. 157). Similarly, "obas, chief and educated elites of Yoruba hometowns reconstructed contending versions of traditional authorities to reinforce and expand their power" in Nigeria's Oyo Province (Vaughan 2003, p. 301).

Such struggles grew fiercer where local resources increased the prize of offering the prevailing definition of the majority. As MacArthur (2016) notes, ethnogenesis among the Luhya in Western Kenya only took off once the discovery of Gold increased the value of an insider group able to ward off immigrants. Among the Kenyan Gikuyu, the arrival of the railroad drastically raised land prices and the stakes of ethnicized politics (Peterson 2004, p. 93-99). Cash crop rents similarly made the Ugandan Gisu aware of their common interests (La Fontaine 1969).

In sum, these examples describe the ethnicization of the local state and its distribution of the "goods of modernity" (Bates 1974): enforcement of property rights, access to markets, and social services. By defining the 'customary,' the allocation of these goods could be tailored to the interests of local elites and their constituents. While such discrimination is most evident in the allocation of land rights, it also affected access to markets and political representation (e.g., Bates 1974; Vaughan 2003). As a feature of African politics, ethnic favoritism continues to affect distributive politics at the local (Ejdemyr, Kramon and Robinson 2018; Harris and Posner 2019) and national levels (e.g., De Luca et al. 2018; Franck and Rainer 2012).

Contemporary Afrobarometer (2018) survey data offer descriptive evidence on the political disenfranchisement of local ethnic minorities. Figure 1 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.

While ethnicized governance as observed until today was not a colonial inven-

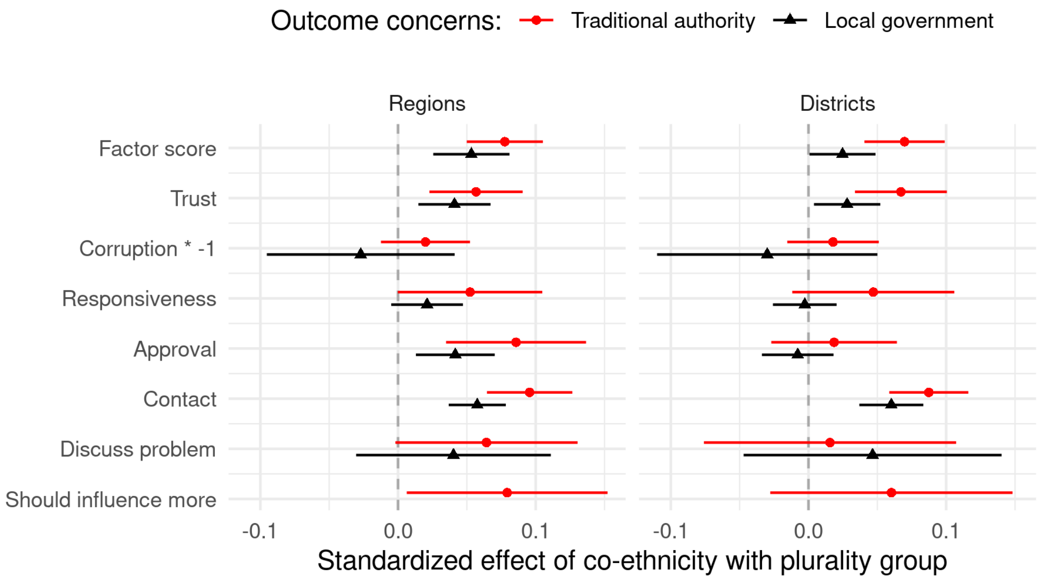



Figure 1: Increased interaction between local plurality members and traditional authorities as well as local governments.

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.

tion, its territoriality is rooted in the administrative borders that defined the local political arena. Administrative units thereby 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".[7]

Second, borders defined the ethnic make-up of administrative units, thereby assigning minority or majority status to individuals. The main theoretical argument predicts that some minority members assimilated and others emigrated in response to the discriminatory policies assigned to them by the borders. Ethnic assimilation

[7]This effect of administrative units was also driven by non-state organizations, in particular missionary societies, that aligned their sphere of influence with administrative borders. According to Chimhundu (1992, p. 87), the missionary fostering of linguistic standardization "helped to create and consolidate new and wider ethnic identities that roughly corresponded with the regions in which the individual church denominations established their spheres of influence."

was a constant historical reality in Africa, in particular among ethnic 'strangers' in acephalous societies (Cohen and Middleton 1970). Relatedly, Green (2020) suggests that African citizens today oftentimes 'switch' their ethnic identity to become co-ethnics to their president to access the spoils of ethnic governance. Drawing on migration as a vehicle of 'revolt,' Asiwaju (1976) shows that some ethnic minorities emigrated in reaction to local discrimination.

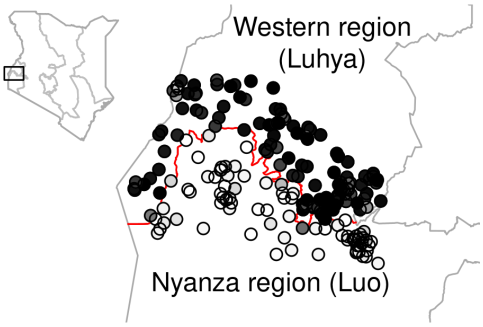
In sum then, the historical evidence suggests that ethnicized territorial governance spurred ethnic change within the boundaries of administrative units. The following quantitative analysis systematically tests whether this process was part of how "the colonial regimes administratively *created* tribes as we think of them today" (Apthorpe 1968, p. 18) and assesses individual-level assimilation and emigration as its driving forces.

## Research design

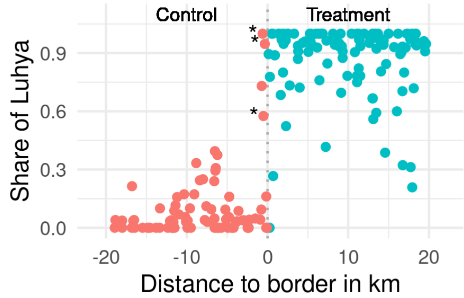
This section introduces the research design used to investigate whether ethnic identities have crystallized along administrative boundaries in Sub-Saharan Africa. I do so by combining data on regional and district boundaries with survey data from the DHS (2018) on individuals' ethnic identity in 25 countries. With these data and building on studies of the effects of African borders (McCauley and Posner 2015), I estimate the effects of regional and district borders on the identification with administrative units' plurality group in a spatial regression discontinuity design (see, e.g., Dell 2010; Keele and Titiunik 2015; Henn 2018). 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

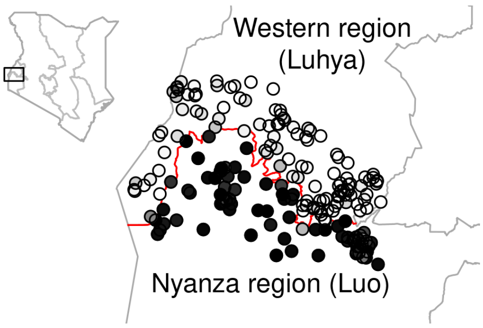
The intuition behind the regression discontinuity design is well captured by the curious and extreme case of the Luhya in western Kenya. At the outset of the $20^{\text{th}}$ century, the district of North Kavirondo was inhabited by a diverse set of Bantu-speakers, comprising anywhere between 15 and 26 distinct yet mutually intelligible dialects, depending on the colonial official and ethnographer consulted (MacArthur 2012; Wagner 1949). Elites of one of the local dialects, Wanga, held



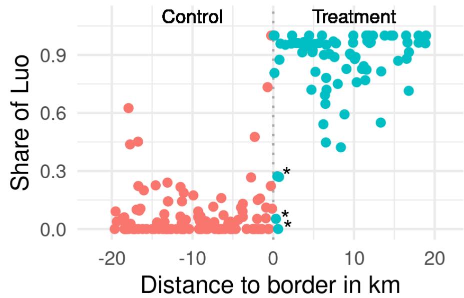

(a) Luhya share




(b) Luhya share




(c) Luo share




(d) Luo share

Figure 2: Luhya and Luo around the Western-Nyanza region 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).

power in the district, which became part of Kenya's postcolonial Western region. The discovery of gold in the early 1930s created a common good to secure, leading local elites to foster a collective identity used to ward off the intrusion by white settlers. Thus, the ethnic label Luhya (or Luyia, 'Kinship') was invented. Promoted by the North Kavirondo Central Association, the Luhya quickly became one of the main Kenyan tribes, counting more than 650'000 members in 1948 (MacArthur 2013).

In neighboring South Kavirondo, the postcolonial Nyanza region, the Luo, while less fragmented, exhibited a similar rise in 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 regions? Evidencing the extreme success of the Luhya identity as regionally bounded construct, Figure 2 shows a sharp change at the border between the Western and Nyanza regions (i.e. between the former North and South Kavirondo districts). We see that the share of the Luhya population drops from an average of about 90% to about 5% as we cross the border from the Western region, where the Luhya are the predominant group, into the Nyanza region. At the same time, the local share of the Luo population increases from about 10% to more than 90% as one enters the Nyanza region, where the Luo constitute the predominant ethnic group. The presence of third ethnic groups makes the two border effects asymmetric.

### Estimation strategy and data

The regression discontinuity design (RDD) generalizes the intuition behind Figure 2. As in the above example, each regional and district border entails two treatments, one for each of its sides. 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 and thereby increases the precision of the treatment effect estimate. In addition, it balances treatment and control groups. Because each point is part of both groups, there are by design no differences in geographic covariates between them and the density function of the running variable is perfectly continuous at the border (Appendix B).

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 $\gamma_s$, 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 account for the spatial clustering of the treatment assignment. For the baseline specification, I only analyze points within 20km to the closest border.[8]

To estimate Eq. 1, I combine spatial data on administrative units and ethnic settlement areas with georeferenced Demographic and Health Survey (DHS 2018) data from 25 African countries (see Appendix A.1). I draw on the geographic data on districts and regions in 1990 from the FAO's (2014) GAUL database[9] 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.[10] Groups' local plurality status so defined acts as a proxy for their politically predominant status – following national-level evidence from Bormann (2019) and Francois, Rainer and Trebbi (2015), I assume that the largest ethnic group on average holds most political power.

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.[11] 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 mostly as respondents'

[8] I vary this threshold in a robustness check. 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 2018).

[9] Colonial and alternative contemporary sources of border data are used for a robustness check in Appendix C.8. The GAUL data has the advantage of being 'historical' whereas other datasets, such as GADM, only depict the boundaries of regions and districts of today thus coming with a higher risk of reverse causality.

[10] Taking instead ethnic demographics from the survey data to identify local plurality groups would introduce post-treatment bias.

[11] See maps in Appendix A.1 for the resulting set of boundaries.

ethnic group, tribe, and sometimes as their language.[12] I match ethnic labels from the DHS data with those on Murdock's map, using the ethnic linkage data from Müller-Crepon, Pengl and Bormann (2019).[13]

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. A comparison of the various digital maps of administrative geographies sheds light on the misalignment of the borders. While 50% of border locations in GAUL and GADM[14] differ by no more than 100 meters, 25% diverge by more than 1000 meters (see Appendix A.2). This uncertainty translates into noise in the treatment assignment, which biases estimates of $\beta_1$ towards zero. A second problem comes with random noise in the coordinates of DHS clusters. To preserve the privacy of respondents, the DHS deliberately displaces 99% of all rural clusters by up to 5km and 1% by up to 10km (Burgert et al. 2013). While the DHS claims to ensure the final location falls into the 'right' region (but not district), the DHS mostly relies on the GADM dataset of administrative units, which provides only contemporary data potentially affected by reverse causality. As already visible in Figure 2, survey clusters close to administrative borders can therefore be spatially associated with the wrong administrative unit. A number of robustness checks below will address this issue.

The main identifying assumption for Eq. 1 to yield an unbiased estimate of $\beta_1$ holds that borders are drawn as-if-randomly within their local spatial neighborhood. They must not line up either with sharp precolonial ethnic boundaries (reverse causality) or any other geographical features that cause sharp spatial discontinuities in ethnic geographies (omitted variable bias). As discussed above, some administrative borders were (at least in British colonies) designed to follow ethnic geographies as perceived at the time.[15] Other borders were drawn in a much more haphazard way, as straight lines cutting across geographical features and ethnic

[12]The respective questions vary from survey to survey, ranging from "What is your ethnicity?" to "What is your mother tongue".

[13]Using their linguistic linkage, I link ethnic groups from the DHS to Murdock's labels if they share at least one common dialect.

[14]Available at gadm.org.

[15]See Appendix C.10 for evidence of no strong differences in the patterns found in former French and British colonies.

groups. I will use such borders to causally identify $\beta_1$.

If the main identifying assumption holds, the RDD identifies the effect of the change in the locally predominant ethnic group on the ethnic demography at the border. Because borderlands are almost by definition peripheries, the analysis hinges on variation in a population that has a lower baseline identification with the plurality group of their unit, and is significantly more rural, older, less educated, and materially poorer than other inhabitants of the same administrative unit. Survey clusters do not differ systematically with regard to historical, ethnic, or environmental characteristics (see Appendix B).

## Empirical analysis

Figure 3 plots the distribution of the share of local plurality groups in the treatment and control groups along the regional and district borders in the data. Along the 323 regional borders in the main sample, we see an increase of about 13 percentage points as one crosses from control into treatment units. The respective increase along the main sample of 1019 district borders amounts to 6 percentage points. Table 1 presents the results of estimating Eq. 1. The estimated effects of $\beta_1$ in the baseline Models 1 (regions) and 4 (districts) are consistent with the patterns visualized in Figure 3 and statistically highly significant.

The smaller effect of district borders can be explained by the fact that many of them run within regions. In turn, regional borders combine the effects of changes in districts' and regions' plurality group.[16] Distinguishing between pure district borders and those that run along regional borders, Table A2 shows this intuition, showing that the former's effect amounts to about 4 percentage points, while the latter's is very close to the baseline effect of regional borders. This suggests that border effects increase in the scale of political units, with national borders presumably associated with the largest effects. Note, however, that the status of district, regional, or national borders might be endogenous to the effects they have had on ethnic (and other) identities (Grossman and Lewis 2014).

[16]Note however that with the operationalizaton employed here, not every regional border that entails a change in the regional plurality group comes with a change in districts' plurality group and vice versa.

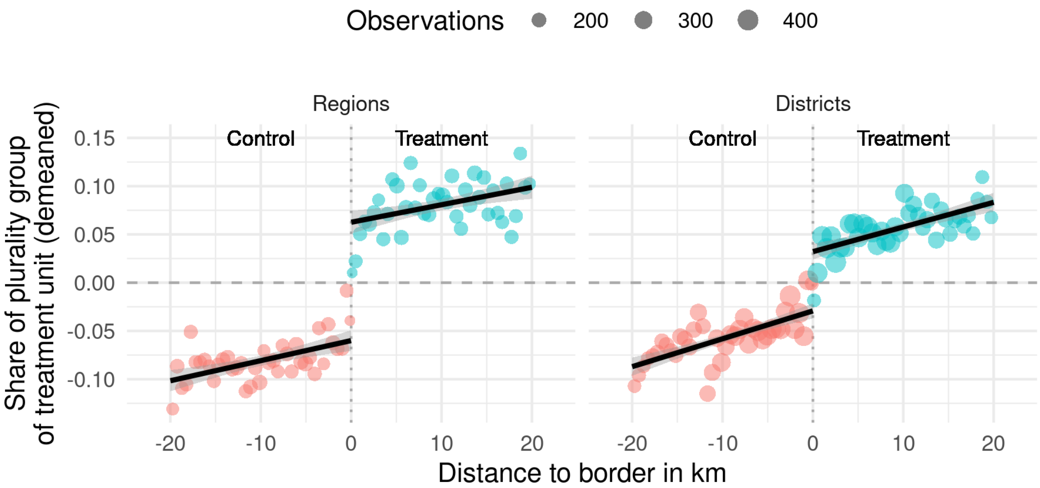


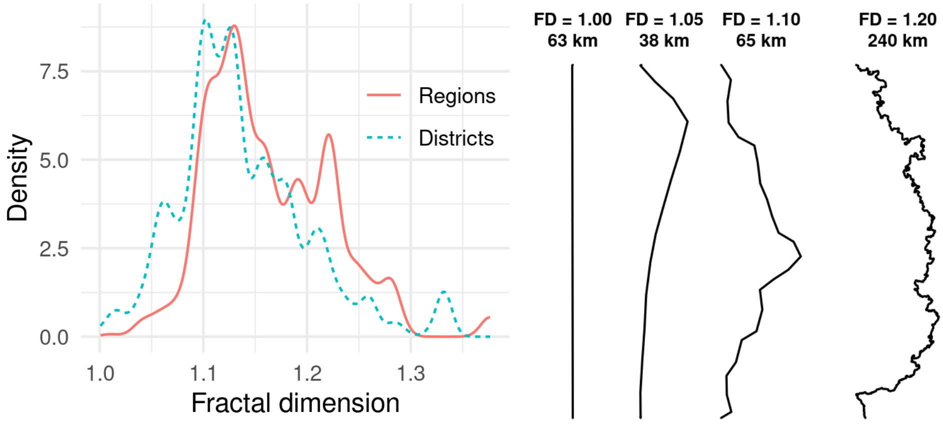

Figure 3: Local population shares of units' plurality groups increase at their borders.

Note: The plots show the demeaned percentage of units' plurality groups within 20km left and right to the borders between treatment and control units. Straight lines and confidence bands result from a linear regression of the trends on each side of the borders.

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




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

Figure 4: Fractal dimension of borders.

Note: Distributions and examples are based on observations in the baseline analysis. For further examples, see Appendix Figure A4.

In order to assess the causal interpretability of these results, I test whether they hold at plausibly exogenous borders. The first set of such borders are those that are not aligned with ethnic geographies but run *within* ethnic groups. Econometrically, this analysis exchanges the previously used border fixed effect with a border-ethnic group intercept.[17] The model then only compares points with each other that are located in the same ethnic group on Murdock's (1959) map. It thereby precludes the results to be driven by an alignment of ethnic borders on that map and administrative unit borders. The results (Models 2 and 5) closely align with the baseline results.

A second caveat might consist in the existence of omitted spatial features that cause administrative borders and sharp discontinuities in the ethnic geography. I test the robustness of the results to this scenario by subsetting the data to points along relatively straight borders, which I assume to be least likely caused by omitted spatial features (Herbst 2000; Touval 1966). I measure the straightness of borders as their fractal dimension (Alesina, Easterly and Matuszeski 2011), informally defined as the degree to which a line fills a two-dimensional plane.[18] Figure 4 plots

[17]This increases the number of analyzed border segments, since a border between two units can run through several ethnic groups. I drop all observations in border-ethnic group buckets that do not feature survey clusters on both sides of the border.

[18]Formally, a line's fractal dimension can be quantified via the boxcount method as $D = \lim_{\epsilon \to \infty} log(N(\epsilon))/log(1/\epsilon)$ where $\epsilon$ is the resolution of a square grid and $N(\epsilon)$ is the number of grid

the distribution of the fractal dimension of borders in the baseline samples, as well as four example lines.

In order to assess whether the baseline results hold along relatively arbitrary, straight borders, I rerun the analysis limiting the sample to points along borders with a fractal dimension of at most 1.1. This corresponds to retaining only (6%) (40%) of the sample along 92 (512) regional (district) borders. As the example in Subfigure 4d shows, borders with a fractal dimension of 1.1 or lower consist of only very few, straight line segments. The effects of straight borders (Models 3 and 6) are consistent with those reported in the baseline Models. The point estimate of the effect along regional borders slightly decreases to .103 and that of the effect along district borders slightly increases to .091. 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 C.1). Taken together, the tests evidence that the results are not driven by omitted variables that cause borders and ethnic discontinuities in space.

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 13 (8) percentage points. This effect is consistent along credibly exogenous borders as identified above. In the next step, I test the robustness of the results to permutations of the baseline specification. I then investigate assimilation and migration as drivers of the crystallization of ethnic groups along administrative borders.

### Robustness tests

This section engages in a series of systematic robustness checks of the regression discontinuity design. Figure 5 summarizes the main estimates and Appendix C presents further details and results. Unless noted otherwise, the same robustness checks for the models that exploit only variation within ethnic groups and across

cells covered by a line. Straight lines have a fractal dimension of 1 and lines that cover the plane approach a value of 2. See Appendix A.3 for details on the computation of $D$.

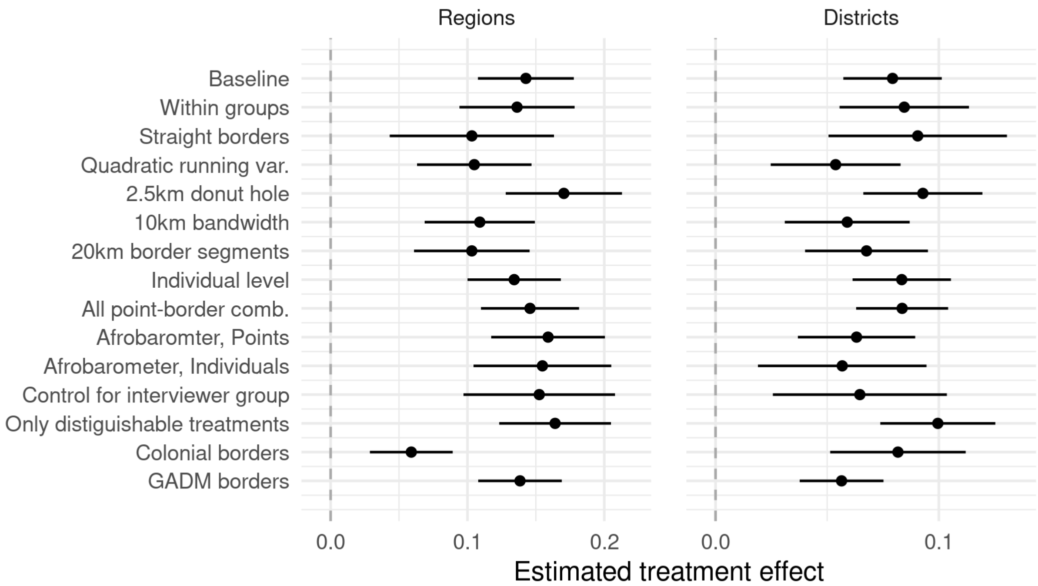



Figure 5: Summary of results from varying specifications.
Note: All results are available in detail in Appendix C.

straight borders coincide with the results presented below.

**Running variable:** I first assess the robustness of the results with respect to the specification of the running variable, the distance to the border. Closely examining Figure 3 we can detect a slight non-linearity in the outcomes very close (<5km) to regional and district borders. As noted above, this does most likely stem from noise in the spatial attribution of survey clusters to administrative units. To delineate the conservative scenario in which the non-linear dynamics are not caused by such noise, I implement the baseline specification controlling 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.

We can also examine the least conservative scenario in which deviations from the linear trend close to the border are only due to noise in the spatial matching of administrative units and DHS clusters. Doing so results in estimating a 'donut'-RDD in which we drop all points closer than 2.5km[19] to the border. Doing so increases the estimated border effects by 2 to 3 percentage points. This difference

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

shrinks with lower minimum distance cutoffs (see Appendix C.3). 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 C.4, results remain mostly stable when subsetting the data to survey clusters closer to the border. As one decreases the bandwidth to 10km, estimates shrink to .07 and 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 assess whether they drive the results, I define sub-border segments such that points on both sides of the same border have a maximal distance to each other of at most 20km with each point being closer than 20km to the border. Estimating Eq. 1 with border-segment fixed effects so defined exploits only variation within a small geographic neighborhood. Doing so does not change the results in a substantive way.

**Additional robustness checks:** I implement several additional analyses to assess the stability of the results under the use of varying data sources and definitions (Figure 5). Using individual-level DHS data rather than cluster-level aggregates does not affect the results. Implementing the same RDD with geocoded ethnicity data from the Afrobarometer (2018) Surveys round 1–6 [20] yields consistent results as well. The Afrobarometer data does also allow for controlling for the main language spoken by interviewers, providing evidence that the results are not due to social desirability bias. Dropping cases in the data for which the ethnic categories enlisted in the DHS survey do not allow for distinguishing the two plurality groups on both sides of a border leads to slightly larger estimates.[21]

Lastly, I reanalyze the DHS data with two different datasets on administra-

[20] I process the Afrobarometer data following the same procedure as applied to the DHS data. Geocodes come from Ben Yishay, Ariel Rotberg et al. (2017).

[21] This is the case in 15 (25) percent of observations in the region (district) samples. In these cases, treatment effects are by design 0 since the same group shares appear as outcomes under the treatment and control conditions. Note, however, that the inability to uniquely link individuals to groups enlisted by Murdock (1959) may be a result of post-treatment ethnic change.

tive borders. The first enlists the historical borders of districts from British and French colonies as well as regions from all countries at their independence. The second dataset comprises contemporary borders retrieved from the GADM database. Colonial regional borders are associated with a effect of about 6 percentage points, while that of colonial districts is largely equivalent to the baseline effects. The differing regional border effect may be related to the postcolonial 'upgrading' of districts settled by regional minorities that became their own region (Grossman and Lewis 2014). The 2019 GADM data leads to estimates that are consistent with the baseline results. Lastly, a country-by-country jackknife shows that the baseline results are not driven by any single country in the data (Appendix C.9).

In sum, these additional analyses suggests that the main results are robust to permutations of the baseline specification.

## Mechanisms: Assimilation and migration

The regression discontinuity estimates show clear evidence of discontinuous changes in the ethnic makeup of local populations at borders between administrative units with differing plurality groups. They do however not allow to assess the drivers of this effect. As argued above, both assimilation and ethnically biased migration patterns could explain the results. I here present evidence for both mechanisms.

### Ethnic assimilation

In the presence of individual-level migration and absent panel data that allows for observing how individuals change their ethnic identities over time, identifying individuals' assimilation is challenging. Assimilation can occur within one's own lifetime and in an intergenerational manner frequently observed in immigrant populations (e.g., Fouka 2019).

Regarding the first type of assimilation, the Afrobarometer surveys have the fortunate feature that they enumerate the *ethnic self-identification* of as well as the *language(s)* spoken by respondents. In particular, round 4 of the survey enumerates *all* languages spoken by an individual, thus capturing the frequent multilingualism on the continent (Buzasi 2016). In addition, each survey asks respondents about their main language. To assess whether a respondent has partially assimilated, we can analyze deviations between respondents' language(s) and their ethnic self-identification. More precisely, we can test whether self-identified minority members speak the local plurality language as an important step in the assimilation process (Cohen and Middleton 1970).

Ethnic assimilation also occurs across generations, in particular through interethnic marriages between local minority and majority members that increase the identity choice set for their children (Cervellati, Chiovelli and Esposito 2016; Cohen and Middleton 1970; Fouka 2019). Studying ethnic assimilation in Northern Ghana, Goody (1969, p.137) concludes that ethnic out-marriage "is the main mechanism whereby integration is achieved." While only available in a minority of all surveys, some DHS surveys record the ethnic identities of spouses in surveyed households. These data allow me to construct whether a female respondent married a member of the local plurality or minority.[22]

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 spoken language in the DHS data corresponds to the local plurality (*Plur. group member*) in interaction with the treatment dummy. This setup lets the treatment dummy capture the change of the respective outcomes *among self-identified minorities* at the borders of regions and district.

I find generally significant and meaningful changes in minorities' assimilation towards the local majority at the border. The estimated border effect on 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 one's main language of 6.6 (2.8) percent points, though the latter estimate is statistically insignificant. These estimates show that minorities tend to assimilate by over-proportionally speaking the predominant local language.[23]

[22]Because of wide-spread polygamy, this is empirically cleanest. Looking at ethnic identities of men and their first spouses yields equivalent results, see Appendix Table A14.

[23]Additional cross-sectional results in Appendix Table A12 show that linguistic members of local

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

Models 3 and 6 find similar patterns of intergenerational assimilation fostered by interethnic marriage. The estimates suggest that female members of a local minority are 1.8 (2.6) percentage points more likely to be married to a region's (district's) local plurality member barely inside the unit than barely outside it. Controlling for the share of the plurality group in the local population shows that this effect can be explained by the post-treatment change in the 'supply' of plurality men at the border (Appendix D). Taken together, this analysis provides evidence that the effect of administrative units on the spatial distribution of ethnic groups partially works through the assimilation of local minorities.

The reader may note that the results in Table 2 cannot be *causally* interpreted, because conditioning on 'plurality group membership' introduces post-treatment bias. The existence of former minority members that have adopted the language of and also self-identify with the local plurality biases the estimated amount of language adoption downwards. The same occurs where minority-majority marriages

plurality groups identify more with their ethnic group (as compared to their national identity) than minority members.

are not accounted for because a spouse has already taken up the ethnic identity of her partner. The estimates thus likely constitute conservative estimates of the extent of ethnic assimilation among non-migrant minority members. Migrants are the subject of the following analysis.

### Ethnic migration patterns

The second mechanism that can cause a sharp decrease in the share of a unit's majority group at its borders consists in ethnic sorting through migration between administrative units, in particular (1) higher emigration rates of local minority members and (2) higher immigration rates of majority members. Furthermore, the local effects at units' borders could be driven by higher relocation rates of minority (majority) members towards the interior (periphery) of a unit. Disposing only of between-unit migration data, this section can only provide evidence on the first two patterns of ethnic sorting through migration between administrative units.

In order to assess ethnically 'biased' migration patterns across administrative units, I rely on census data samples from eleven countries in Sub-Saharan Africa[24] provided by IPUMS (Minnesota Population Center 2018). These records of about 33 million individuals contain individuals' region of birth and residence. The data from Burkina Faso, Mali, Senegal, Sierra Leone, and Zambia additionally contain the same variables at the level of districts. Jointly, this information on birth and residence units allows me to derive the full lifetime migration matrix of the population enumerated in each census.[25] To assess distinct migration patterns of local minority and plurality members, I draw on IPUMS' geographic data on administrative units and derive each unit's plurality group as that ethnic group from Murdock's (1959) map with the largest spatial intersection. The ethnic linkage data from Müller-Crepon, Pengl and Bormann (2019) provide the link of ethnic labels from IPUMS with Murdock's group names, necessary to code groups in the census as local plurality and minority groups in each administrative unit.

With the resulting dataset, I conduct three analyses (Table 3). I first estimate the

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

[25]Sorichetta et al. (2016) use the administrative unit of last residence for a similar purpose.

extent to which groups' status as local plurality acts as negative push factor that reduces emigration rates from individuals' regions and districts of birth (Models 1 and 4). Regions' (districts') Plurality group members show a emigration rate that is 12 (17) percentage points lower than that among local minorities. I then analyze the extent to which local plurality status acts as a pull factor, increasing immigration into migrants' co-ethnic regions and districts (Models 2 and 5). Again, the extent of such co-ethnic migration bias is extensive. Migrants move with a 6.3 (8.1) percentage points higher probability towards a region (district) in which they belong to the main ethnic group. Both patterns are identified in the presence of fixed effects that account for ethnic group and administrative unit characteristics.

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*** |
| | (0.019) | | (0.002) | (0.021) | | (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: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

As a last step of this assessment of spatial ethnic sorting, Models 3 and 6 present a full dyadic analysis of migration patterns. Here, the unit of analysis is the ethnic group nested in directed birth to residence unit dyads. The outcome consists in the share of the members of an ethnic group in the respective birth unit that has migrated towards the residence unit. Controlling for dyad and ethnic group fixed effects, the two main predictors assess the degree to which ethnic groups differentially migrate between administrative unit. The results coincide with those discussed above. The average migration rate between two regions (districts) amounts

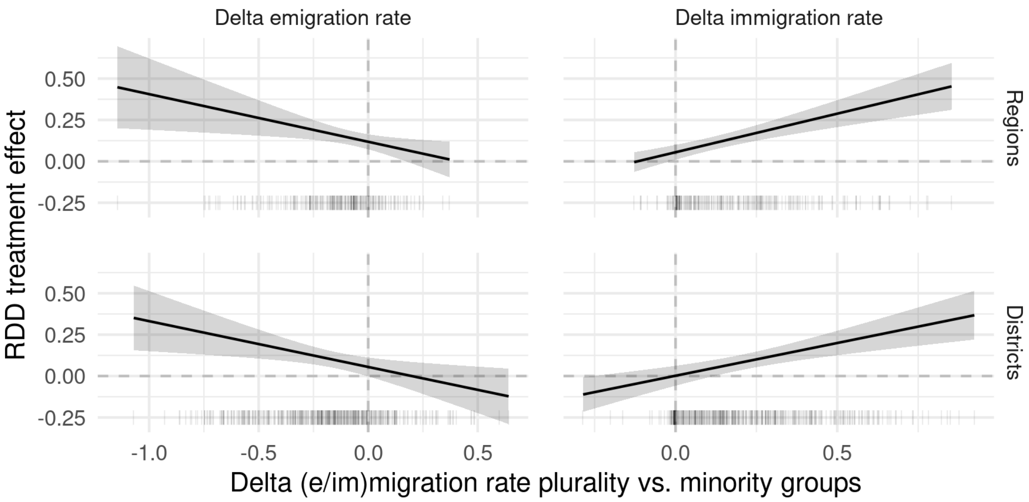



Figure 6: Effect of borders on ethnic identities increase in ethnic differences in e- and immigration.

Note: RDD estimates based on the DHS data on borders retrieved from the IPUMS database. Grey bars denote sample observations. All results are reported in Appendix E.2.

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. Given the low dyadic baseline migration rates, these effects are substantive. Additional results show that they are robust to varying model and sample specifications (see Appendix E.1).

Local minority members are thus more likely to exit, and migrants show a preference for migrating towards areas where they are part of the local ethnic plurality. Attributes of ethnic groups, source, or target units do not explain these patterns. 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 specification (Eq. 1).[26] The resulting interaction effect captures the correlation between the extent of ethnically biased migration and the

[26]To add the data to the main dataset, I merge the DHS data with border files made available by IPUMS and use the respective boundaries to construct the RDD design.

degree to which the share of a unit's plurality ethnic group increases at its border.

Plotted in Figure 6, the results 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 point yields a 3 percentage point 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,[27] the result suggests that ethnically biased migration has contributed to the crystallization of ethnic identities along administrative borders.

## Conclusion

John Iliffe (1979) has famously argued that "Europeans believed Africans belonged to tribes; Africans built tribes to belong to." This paper has analyzed the effect of the introduction of territorial rule through administrative units as one important mechanism behind this argument. I argue that colonial governance introduced clearly demarcated, inflexible administrative borders to the continent. The units they enclosed often cut across ethnic geography. Nevertheless, local governance was often ethnicized on the bedrock of partly invented, partly preexisting 'customary' institutions. In turn, local minorities created by haphazard 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 the ethnic identities exhibited across administrative borders. In particular, the share of regions' (districts') predominant ethnic group increases by about 13 (8) percentage points at the border to a unit dominated by a different group. Additional analyses support the argument that ethnic assimilation and emigration of minorities drive

[27] A host of factors, in particular ethnic discrimination, can simultaneously affect local ethnic change through assimilation and migration.

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. For one, I find that ethnic identities are not prehistorically given but politically shaped inside political arenas structured by territorial boundaries. 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, Reza Baqir and William Easterly. 1999. "Public goods and ethnic divisions." *Quarterly Journal of Economics* 114(4):1243–1284.

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. 2018. "Colonial Legacy, State-building and the Salience of Ethnicity in Sub-saharan Africa." *Economic Journal, forthcoming* .

Apthorpe, Raymond. 1968. "Does tribalism really matter?" *Transition* (37):18–22.

Argyle, W J. 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, A. 1983. "The Concept of Frontier in the Setting of States in Pre-colonial Africa." *Présence Africaine* 127:43.

Asiwaju, a. 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. 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.

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.

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.

Berman, Bruce J. 1998. "Ethnicitiy, Patronage, and the African State: The Politics of Uncivil Nationalism." *African Affairs* 388(97):305–341.

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.

Cervellati, Matteo, Giorgio Chiovelli and Elena Esposito. 2016. "Bite and divide: ancestral exposure to malaria and the emergence and persistence of ethnic diversity in Africa.".

Chandra, K. 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.

Cohen, William B. 1971. *Rulers of Empire: the French Colonial Service in Africa*. Stanford: Hoover Institution Press.

Conklin, Alice L. 1997. *A Mission to Civilize. The Republican Idea of Empire in France and West Africa, 1895–1930.* Stanford: Stanford University Press.

Crowder, M. 1964. "Indirect rule: French and British style." *Africa: Journal of the International African Institute* 34(3):197–205.

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.

Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica* 78(6):1863–1903.

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.

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.

Eubank, Nicholas. 2019. "Social Networks and the Political Salience of Ethnicity." *Quarterly Journal of Political Science* 14(1):1–39.

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.

Francois, Patrick, Ilia Rainer and Francesco Trebbi. 2015. "How Is Power Shared In Africa?" *Econometrica* 83(3):465–503.

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 p. 114..149.

Green, Elliott. 2018. "Ethnicity, national identity and the state: evidence from sub-Saharan Africa." *British Journal of Political Science* pp. 1–23.

Green, Elliott D. 2020. "The politics of ethnic identity in Sub-Saharan Africa." *Comparative Political Studies, Online First* .

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

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.

Hechter, Michael. 1975. *Internal Colonialism. The Celtic Fringe in British National Development, 1536-1966*. London: Routledge & Kegan Paul.

Henn, Soeren J. 2018. "The State, Chiefs, and Development: Evidence from Africa." (November).

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

Hirschman, Albert O. 1970. *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Cambridge, MA: Harvard University Press.

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

Ichino, Nahomi and Noah L. Nathan. 2013. "Crossing the Line: Local Ethnic Geography and Voting in Ghana." *American Political Science Review* 107(02):344–361.

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.

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

Kristof, Ladis K D. 1959. "The Nature of Frontiers and Boundaries." *Annals of the Association of American Geographers* 49(3):269–282.

La Fontaine, J. S. 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* pp. 622–634.

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

Larson, Jennifer M and Janet I Lewis. 2017. "Ethnic networks." *American Journal of Political Science* 61(2):350–364.

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.

MacArthur, Julie. 2016. *Cartography and the political imagination: mapping community in colonial Kenya*. Ohio University Press.

Mann, Michael. 1984. "The autonomous power of the state: its origins, mechanisms and results." *European Journal of Sociology* 25(2):185–213.

Mann, Michael. 2005. *The Dark Side of Democracy. Explaining Ethnic Cleansing*. Cambridge: Cambridge 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.

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

Michalopoulos, Stelios and Elias Papaioannou. 2016. "The Long-Run Effects of the Scramble for Africa." *American Economic Review* 106(7):1802–1848.

Miles, William F. S. 1994. *Hausaland Divided. Colonialism and Independence in Nigeria and Niger*. Ithaca: Cornell University Press.

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

Montalvo, José G. and Marta Reynal-Querol. 2005. "Ethnic polarization, potential conflict, and civil wars." *American Economic Review* 95(3):796–816.

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

Müller-Crepon, Carl, Guy Schvitz and Lars-Erik Cederman. 2020. "Shaping States into Nations: The Effects of Ethnic Geography on State Borders." *Unpublished Working Paper* .

Müller-Crepon, Carl, Yannick I Pengl and Nils-Christian Bormann. 2019. "Linking Ethnic Data from Africa." *Unpublished Working Paper, https://github.com/carl-mc/LEDA* .

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. 2020. "Cash Crops, Print Technologies and the Politicization of Ethnicity in Africa." *CEPR Discussion Paper No. DP15162* .

Peterson, Derek R. 2004. *Creative Writing: Translation, Bookkeeping, and the Work of Imagination in Colonial Kenya*. Social history of Africa Portsmouth, NH: Heinemann.

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 p. 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.

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. 2017. "The Political Logic of Cultural Revival: Ethnic Visibility and Electoral Politics in Malawi." *Unpublished Working Paper* .

Ruggie, John Gerard. 1983. "Continuity and transformation in the world polity: Toward a neorealist synthesis." *World Politics* 35(2):261–285.

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.

Sorichetta, Alessandro, Tom J Bird, Nick W Ruktanonchai, Elisabeth zu Erbach-Schoenberg, Carla Pezzulo, Natalia Tejedor, Ian C Waldock, Jason D Sadler, Andres J Garcia, Luigi Sedda and Others. 2016. "Mapping internal connectivity through human migration in malaria endemic countries." *Scientific data* 3(1):1–16.

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

Spear, Thomas. 2003. "Neo-Traditionalism and the Limits of Invention in British Colonial Africa." *Journal of African History* 44(1):3–27.

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.

Weidmann, Nils B. 2009. "Geography as Motivation and Opportunity: Group Concentration and Ethnic Conflict." *Journal of Conflict Resolution* 53(4):526–543.

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 University Press.

Wimmer, Andreas. 2018. *Nation building: why some countries come together while others fall apart*. Princeton: Princeton 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* .

# Appendix

# Building Tribes: On the Administrative Origins of Ethnicity in Africa

## Table of Contents

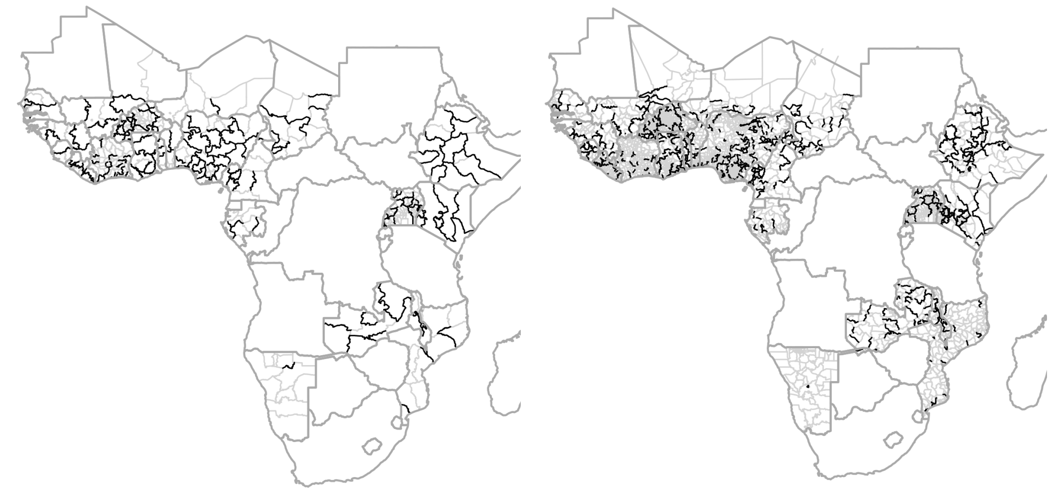
# A Data Appendix

## A.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.



(a) Regional borders (b) District borders

Figure A1: Borders analyzed in the baseline DHS sample

## A.2 Border alignment

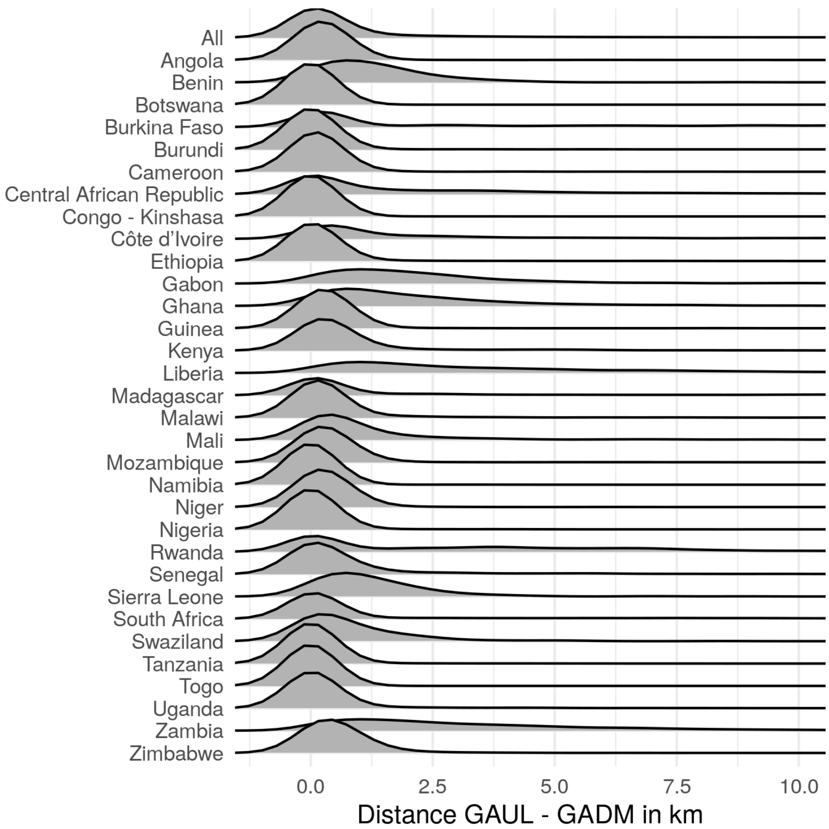



Figure A2: Minimum distance between points on district borders in GAUL (1990) to borders in GADM.

Note: Calculated by sampling 1 point on each GAUL border per 10 kilometers. For each point, I then calculate the distance to the closest border in the GADM data. Distributions are censored at 10 kilometers.

## A.3 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 $\epsilon$ 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 $\epsilon$ of 1/513 cells) that covers its bounding circle.

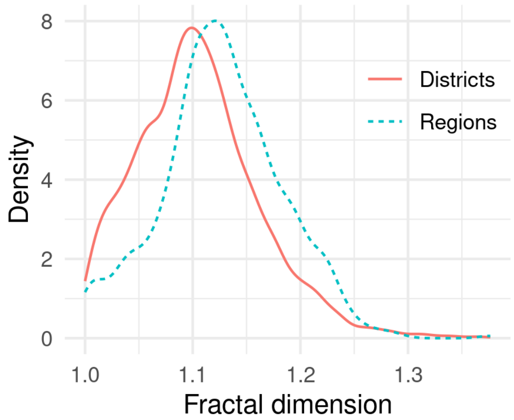



Figure A3: Distribution of fractal dimension values among *borders* in the baseline analysis. Each border receives equal weight.

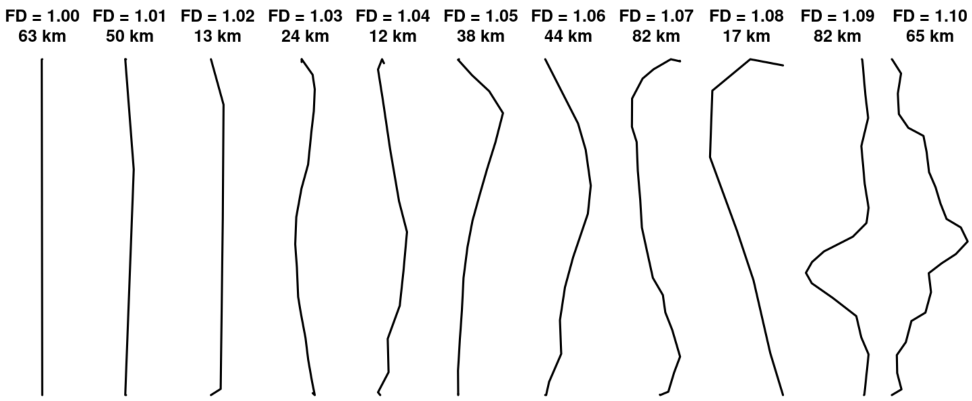



Figure A4: Lines with a fractal dimension between 1 and 1.1.

Note: Examples of lines used in Models 3 and 6 in Table 1.

2. Define a series of square grids $g \in G$ with a x and y resolution $\epsilon_g$ 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 intersect 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.[28]

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

# B Regression discontinuity analysis: Descriptive statistics

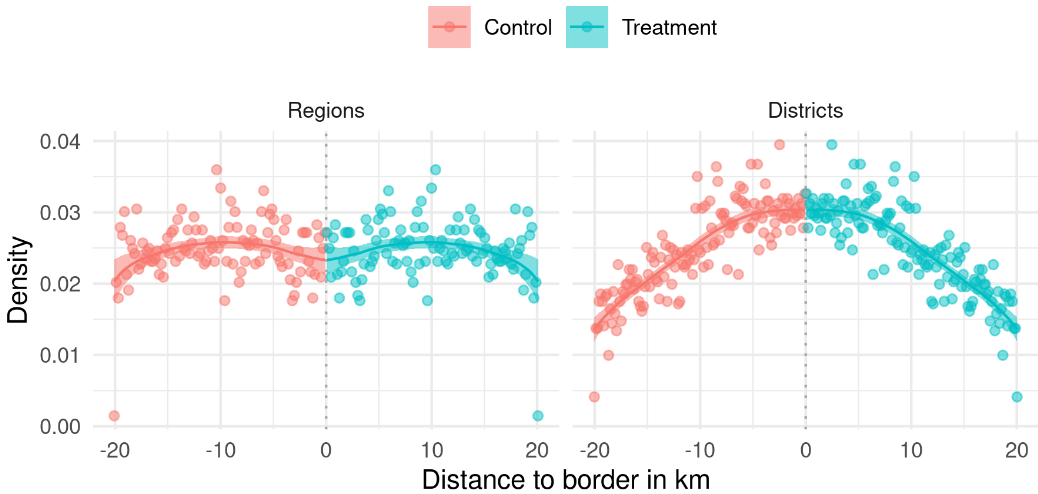



Figure A5: McCrary (2008) test.

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

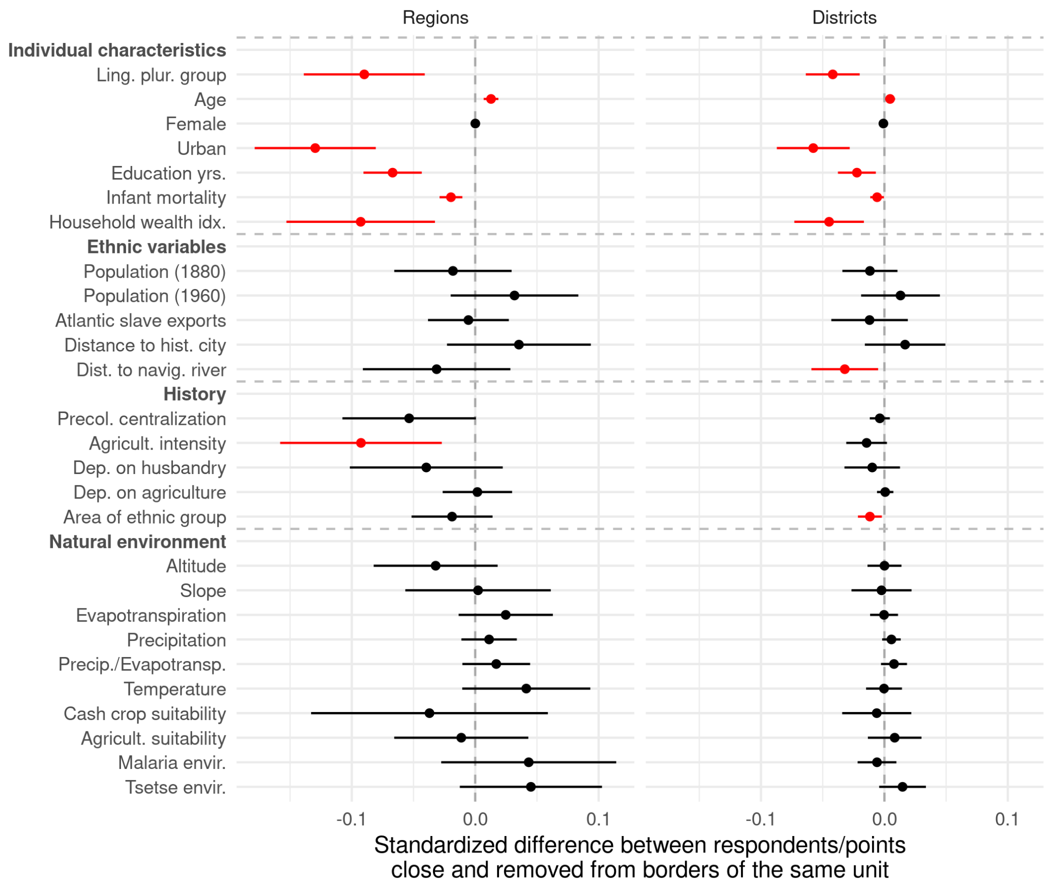



Figure A6: Comparison of DHS respondents and survey clusters close (<20km) to borders that divide units with differing plurality ethnic groups with average observations from the same units.

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 × 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.

# C 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 A7 combines the coefficients from most analyses discussed below.

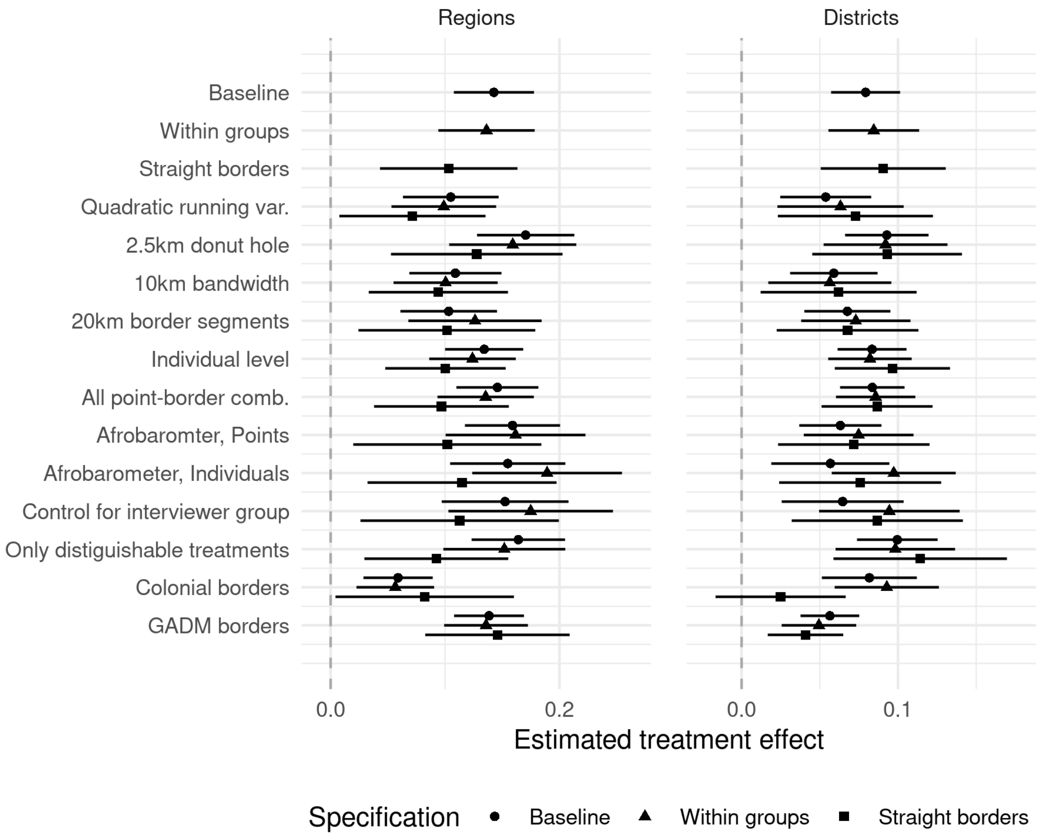



Figure A7: Summary of results from varying robustness checks.

## C.1 Fractal dimension and other measures of border arbitrariness

In order to gauge the effect of the particular and somewhat arbitrary fractal dimension cutoff of 1.1 used in the main analysis, this robustness check introduces alternative cutoff values and measures borders' arbitrariness via their alignment with rivers and watersheds. First, Figure A8 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.[29] I then split the resulting alignment variables, as well as, for comparative purposes,

[29]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.

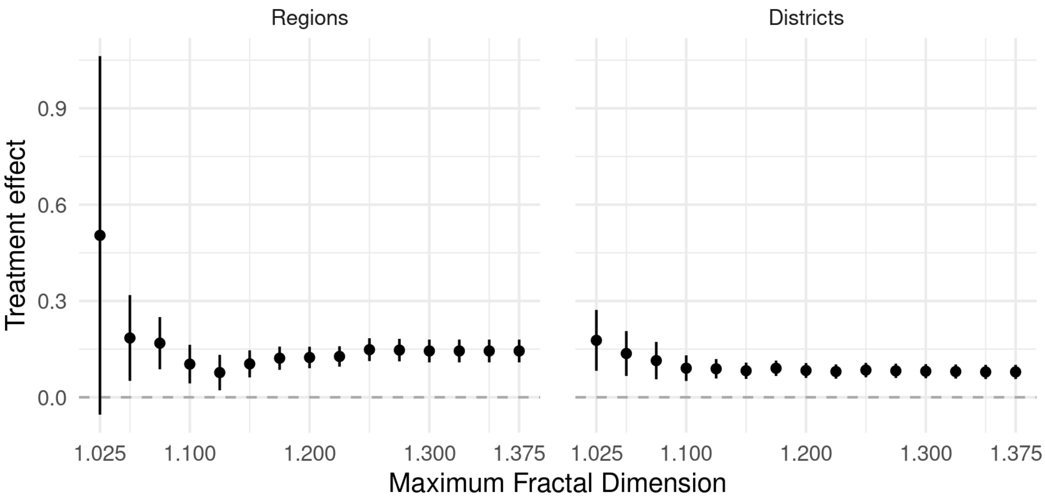



Figure A8: Estimate of border effect on local ethnic identities with different fractal dimension cutoffs.

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).

the fractal dimension values, into quartiles, with higher quartiles assigned to more natural borders, i.e. squiggly lines that align with rivers or watersheds. Finally, I re-estimate the baseline model for each quartile separately, resulting in the estimates presented in Figure A9. 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.

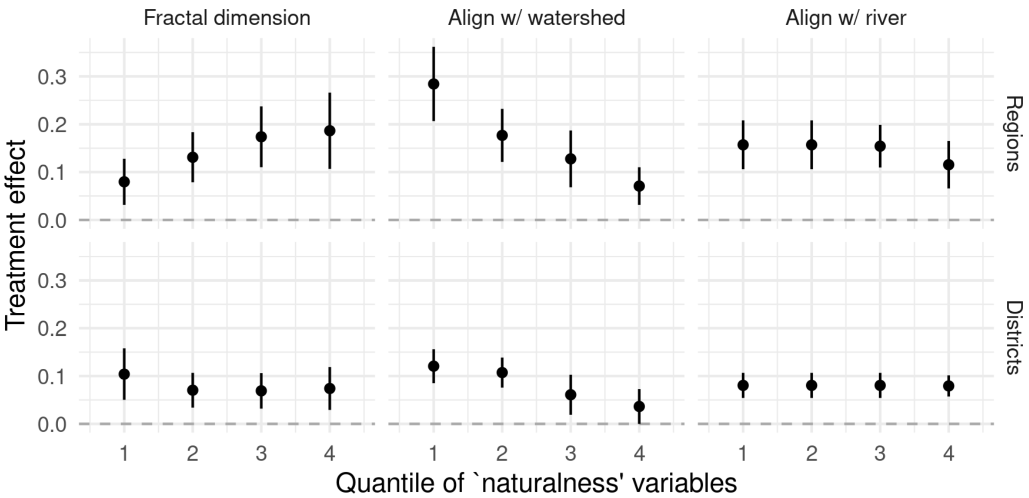



Figure A9: Estimate of border effect on local ethnic identities by degree of naturalness of border.

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.

## C.2 Distinguishing between regional and district borders

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

Table A2: Robustness check: Distinguishing pure district borders

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

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

## C.3 Running variable

Because a visual inspection of the data indicates a non-linear trend very close to the examined borders (below 5km), Table A3 assesses the 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 A10 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.

Table A3: Robustness check: Quadratic running variable specification

| | 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 | 0.105*** | 0.099*** | 0.071** | 0.054*** | 0.063*** | 0.073*** |
| | (0.021) | (0.023) | (0.033) | (0.015) | (0.021) | (0.025) |
| Cutoff | 20km | 20km | 20km | 20km | 20km | 20km |
| Running var linear | yes | yes | yes | yes | yes | yes |
| Running var quadratic | 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 | 0.667 | 0.679 | 0.648 | 0.700 | 0.655 |

*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

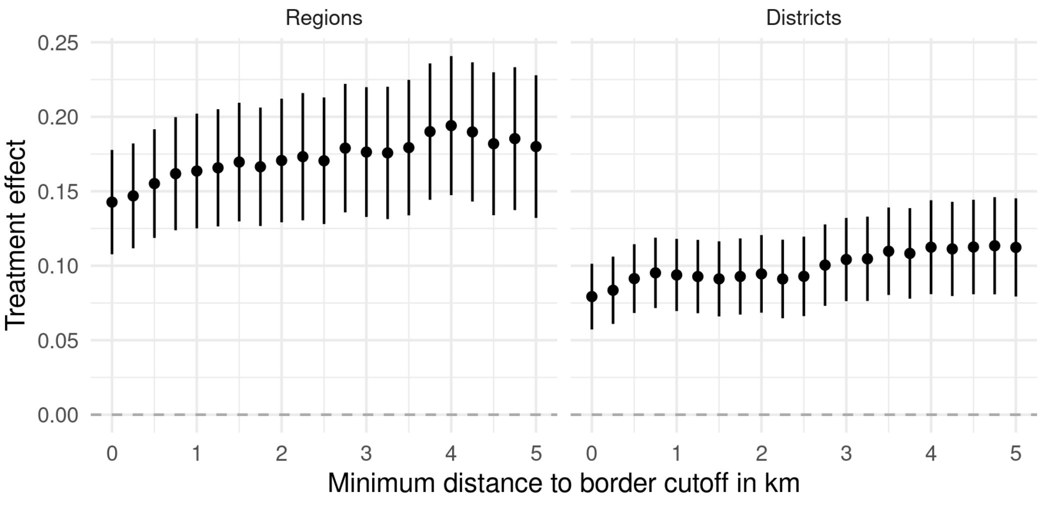



Figure A10: Estimate of border effect on local ethnic identities by size of 'donut'-hole around the border.

Note: x-axis signals the minimum distance of included observations to the border. For other specification details, see also baseline Models 1 and 4 in Table 1 in the main text.

### C.4 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 A11, 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 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.[30] 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 A12, 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.

[30]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.

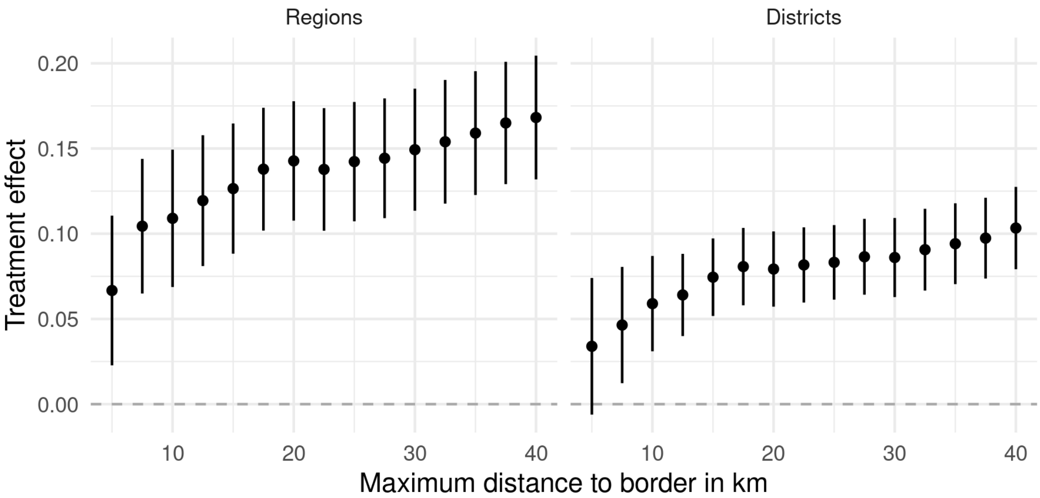



Figure A11: Estimate of border effect on local ethnic identities by bandwidth.

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

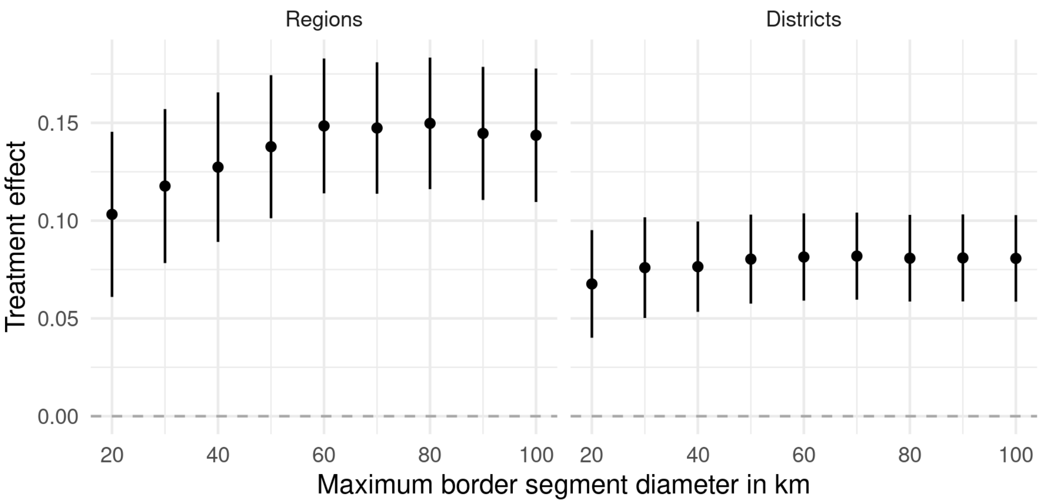



Figure A12: Estimate of border effect on local ethnic identities by maximum diameter of point clusters along border segments.

Note: Estimated with a stable bandwidth of 20km. See also baseline Models 1 and 4 in Table 1 in the main text.

## C.5 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. Table A4 presents the ensuing results which are next to identical to the main estimates.

Table A4: Robustness check: Individual-level analysis

| | 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 | 0.134*** | 0.124*** | 0.100*** | 0.083*** | 0.082*** | 0.096*** |
| | (0.017) | (0.019) | (0.027) | (0.011) | (0.014) | (0.019) |
| 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 |
| Group-Border FE: | no | yes | no | no | yes | no |
| Mean DV: | 0.34 | 0.35 | 0.31 | 0.39 | 0.4 | 0.38 |
| Borders: | 323 | 785 | 92 | 1019 | 1283 | 512 |
| Observations | 525,176 | 350,614 | 92,908 | 792,912 | 451,174 | 324,900 |
| Adjusted $R^2$ | 0.454 | 0.522 | 0.523 | 0.487 | 0.534 | 0.516 |

*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

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 Table A5 show that doing so does not change the results.

Table A5: Robustness check: All border-point combinations, weighted

| | 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 | 0.146*** | 0.135*** | 0.097*** | 0.084*** | 0.086*** | 0.087*** |
| | (0.018) | (0.021) | (0.030) | (0.011) | (0.013) | (0.018) |
| 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 | 0.35 | 0.31 | 0.39 | 0.39 | 0.37 |
| Borders: | 365 | 872 | 95 | 1434 | 1701 | 662 |
| Observations | 18,918 | 12,172 | 2,768 | 37,074 | 18,720 | 12,552 |
| Adjusted $R^2$ | 0.593 | 0.669 | 0.679 | 0.651 | 0.708 | 0.656 |

*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

## C.6 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 (2019). Finally, I re-estimate the main models using both survey clusters (Table A6) and respondents (Table A7) as the unit of analysis. Despite differences in the countries covered and ethnic groups enumerated by the Afrobarometer, the results closely coincide with those obtained from the DHS data.

Table A6: Robustness check: Afrobarometer data, Points

| | 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 | 0.159*** | 0.162*** | 0.102** | 0.063*** | 0.075*** | 0.072*** |
| | (0.021) | (0.031) | (0.042) | (0.013) | (0.018) | (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 |
| Group-Border FE: | no | yes | no | no | yes | no |
| Mean DV: | 0.39 | 0.46 | 0.39 | 0.38 | 0.38 | 0.37 |
| Borders: | 265 | 377 | 57 | 573 | 492 | 243 |
| Observations | 9,530 | 4,600 | 2,184 | 12,014 | 4,866 | 4,628 |
| Adjusted $R^2$ | 0.710 | 0.754 | 0.780 | 0.728 | 0.787 | 0.737 |

*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 A7: Robustness check: Afrobarometer data, Individuals

| | 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 | 0.155*** | 0.189*** | 0.115*** | 0.057*** | 0.097*** | 0.076*** |
| | (0.026) | (0.033) | (0.042) | (0.019) | (0.020) | (0.026) |
| 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 |
| Group-Border FE: | no | yes | no | no | yes | no |
| Mean DV: | 0.38 | 0.47 | 0.42 | 0.41 | 0.39 | 0.39 |
| Borders: | 265 | 377 | 57 | 573 | 492 | 243 |
| Observations | 156,290 | 79,354 | 44,744 | 196,406 | 82,582 | 84,652 |
| Adjusted $R^2$ | 0.577 | 0.625 | 0.689 | 0.579 | 0.628 | 0.621 |

*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

## C.7 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.[31] 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

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

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 C.6), and use information on interviewers' home language to code whether s/he belongs to the plurality ethnic group of the treatment unit.[32] 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[33], this correlation does not affect the estimated discontinuous jump in ethnic identities at district and regional borders. The main coefficient of interest remains stable in size and statistical significance.

Table A8: Robustness check: Controlling for interviewer language

| | 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 | 0.153*** | 0.175*** | 0.113** | 0.065*** | 0.094*** | 0.087*** |
| | (0.028) | (0.037) | (0.044) | (0.020) | (0.023) | (0.028) |
| Interviewer of plurality | 0.116*** | 0.140*** | 0.102*** | 0.067*** | 0.047*** | 0.065*** |
| | (0.014) | (0.024) | (0.022) | (0.011) | (0.014) | (0.011) |
| 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 |
| Group-Border FE: | no | yes | no | no | yes | no |
| Mean DV: | 0.37 | 0.45 | 0.42 | 0.41 | 0.39 | 0.38 |
| Borders: | 264 | 341 | 57 | 570 | 447 | 241 |
| Observations | 127,020 | 63,818 | 37,748 | 160,628 | 70,542 | 67,004 |
| Adjusted $R^2$ | 0.558 | 0.616 | 0.680 | 0.573 | 0.608 | 0.621 |

*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

A second caveat inherent in the data is that the language-based matching of ethnic groups mapped by Murdock (1959) and groups enumerated in the DHS is "many-to-many," with some groups from Murdock corresponding to multiple groups in the DHS surveys and vice-versa. Hence, for some borders with two differing plurality groups from Murdock at either side, the same DHS group(s) are attributed to both.[34] 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

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

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

[34] 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.

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, Table A9 presents the results from dropping all observations along borders for which the plurality groups on either side are not distinguishable in the DHS data. The respective results closely align with the baseline estimates, indicating that such biases are not severe.

Table A9: Robustness check: Drop borders with indistinguishable 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 | 0.164*** | 0.152*** | 0.092*** | 0.100*** | 0.098*** | 0.114*** |
| | (0.021) | (0.027) | (0.032) | (0.013) | (0.019) | (0.028) |
| 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 |
| Group-Border FE: | no | yes | no | no | yes | no |
| Mean DV: | 0.3 | 0.31 | 0.29 | 0.32 | 0.32 | 0.3 |
| Borders: | 278 | 1814 | 68 | 806 | 3383 | 340 |
| Observations | 12,904 | 12,772 | 1,668 | 17,138 | 17,060 | 5,440 |
| Adjusted $R^2$ | 0.585 | 0.681 | 0.712 | 0.641 | 0.708 | 0.621 |

*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

## C.8 Alternative border data

I rerun the baseline analysis using two alternative datasets on borders, one from the colonial era and the other from 2019. The colonial-era data 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) (see Figures A13 and A14). The main downside of these district data is that they 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 uncertainty. However, the results for colonial-era regional and district borders presented in Table A10 coincide with the main estimate. This is with the exception of the analysis of relatively straight district (but not regional) borders, where the estimate

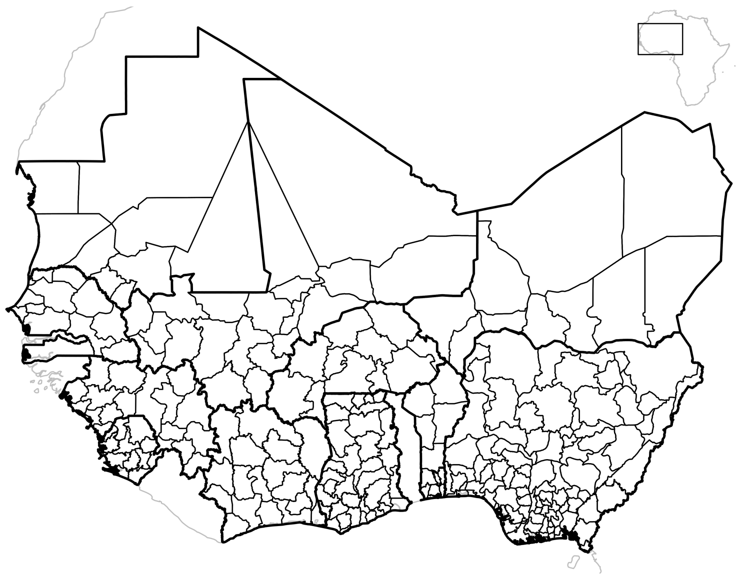


Figure A13: Colonial districts in West Africa.

shrinks and becomes statistically insignificant. This deviation is most likely due to the severely restricted sample, which consists of 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 smaller than in the baseline analysis. This is most likely due to the upgrading of ethnically distinct districts to regional status over time.

Table A10: Robustness check: Colonial border data

| | 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 | 0.059*** | 0.057*** | 0.082** | 0.082*** | 0.093*** | 0.025 |
| | (0.015) | (0.017) | (0.040) | (0.015) | (0.017) | (0.021) |
| Cutoff | 20km | 20km | 20km | 20km | 20km | 20km |
| Running var linear | yes | yes | yes | yes | yes | yes |
| Border × survey FE: | yes | yes | yes | yes | yes | yes |
| Mean DV: | 0.3 | 0.31 | 0.3 | 0.4 | 0.42 | 0.36 |
| Borders: | 197 | 551 | 45 | 278 | 605 | 106 |
| Observations | 11,688 | 8,882 | 1,796 | 13,608 | 9,182 | 4,470 |
| Adjusted $R^2$ | 0.539 | 0.644 | 0.686 | 0.644 | 0.696 | 0.724 |

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

The second alternative source of 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

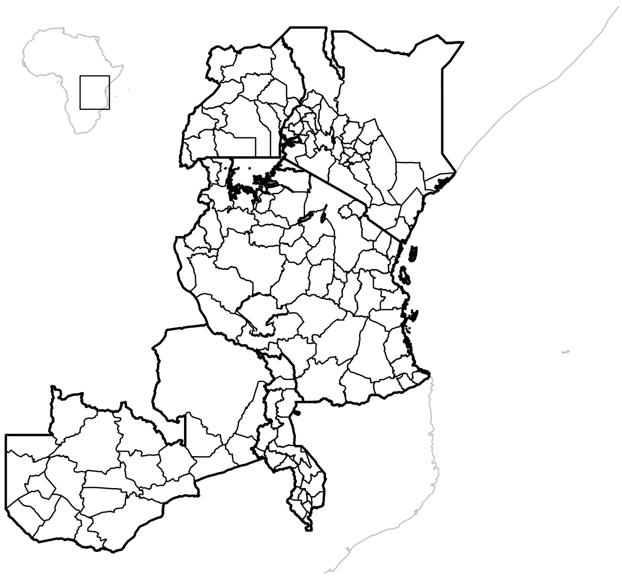


Figure A14: Colonial districts in Southern and East Africa.

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 Table A11 shows, the results closely align with the baseline estimates.

Table A11: Robustness check: GADM border data

| | 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 | 0.138*** | 0.136*** | 0.146*** | 0.056*** | 0.049*** | 0.041*** |
| | (0.016) | (0.019) | (0.032) | (0.010) | (0.012) | (0.012) |
| Cutoff | 20km | 20km | 20km | 20km | 20km | 20km |
| Running var linear | yes | yes | yes | yes | yes | yes |
| Border × survey FE: | yes | yes | yes | yes | yes | yes |
| Mean DV: | 0.37 | 0.38 | 0.37 | 0.39 | 0.39 | 0.41 |
| Borders: | 364 | 936 | 106 | 1096 | 1367 | 624 |
| Observations | 19,002 | 13,182 | 3,886 | 22,364 | 11,722 | 11,662 |
| Adjusted $R^2$ | 0.580 | 0.636 | 0.628 | 0.652 | 0.707 | 0.694 |

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

## C.9 Country-level jackknife

I use a county-level jackknife estimation to gauge the impact every country in the main sample has on the baseline estimates. Figure A15 plots the results of re-estimating the main model, in each iteration dropping observations from the

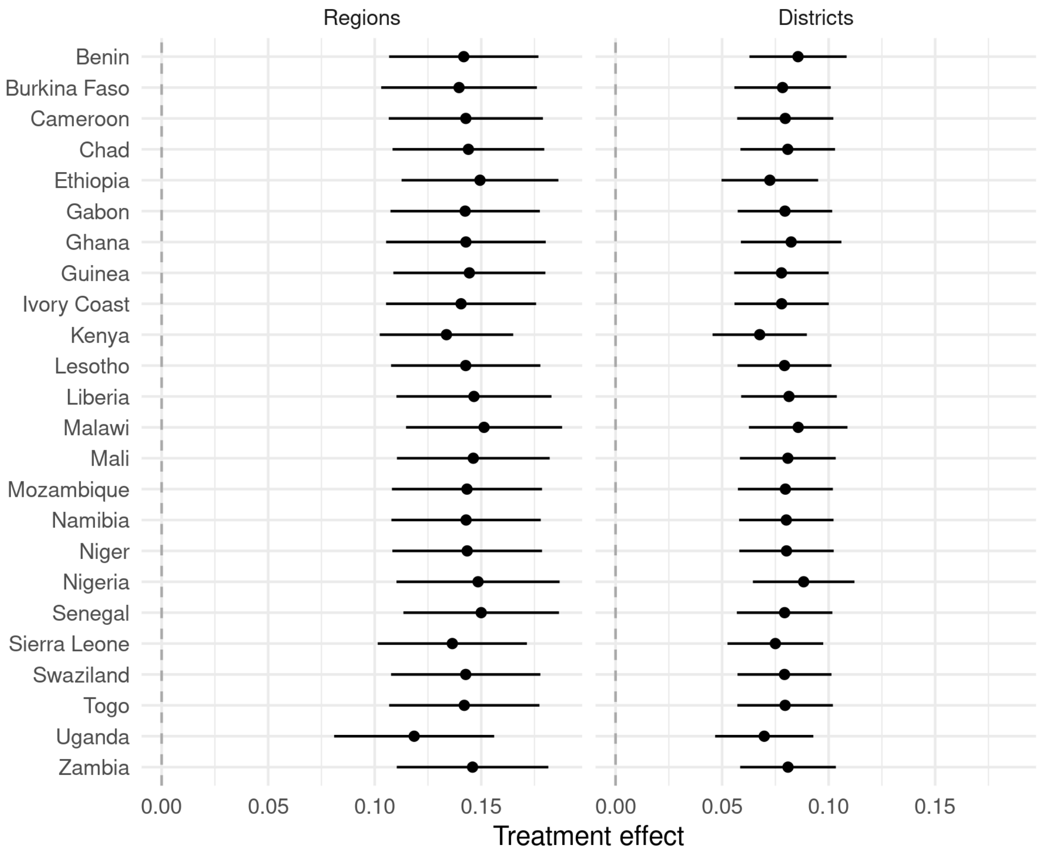



Figure A15: Estimate of border effect on local ethnic identities, dropping one country at the time.

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

country indicated on the y-axis. The results show that no single country drives the results in a significant manner.

### C.10 Split by former colonizer

To gauge whether the effects vary significantly between former colonizers – in particular the British and the French –, I split the sample accordingly and re-estimate the main RDD specification. Given its stronger indirect rule, it is not surprisingly the former British colonies show slightly larger effects than former French colonies (Figure A16). However, the French-British differences are not statistically significant.

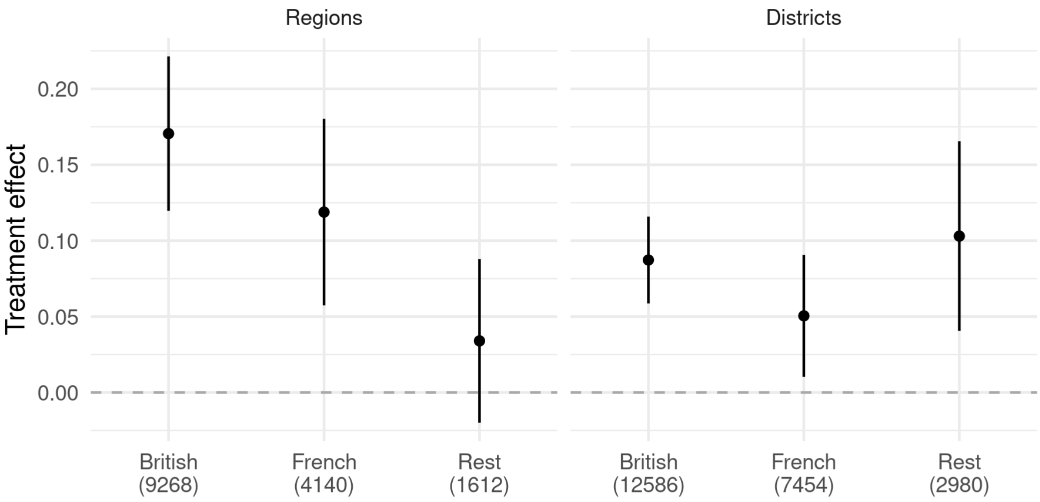



Figure A16: Estimate of border effect on local ethnic identities, split sample by former colonizer.

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

## D Ethnic assimilation: Additional results

This section presents additional results on ethnic assimilation and integration only briefly mentioned in the main paper. First, Table A12 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 A12: 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 A13 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 A13 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 A14 shows very similar dynamics among the spouses of men interviewed by the DHS.

Table A13: (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 A14: (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

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

## E.1 Robustness checks

Figure A17 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

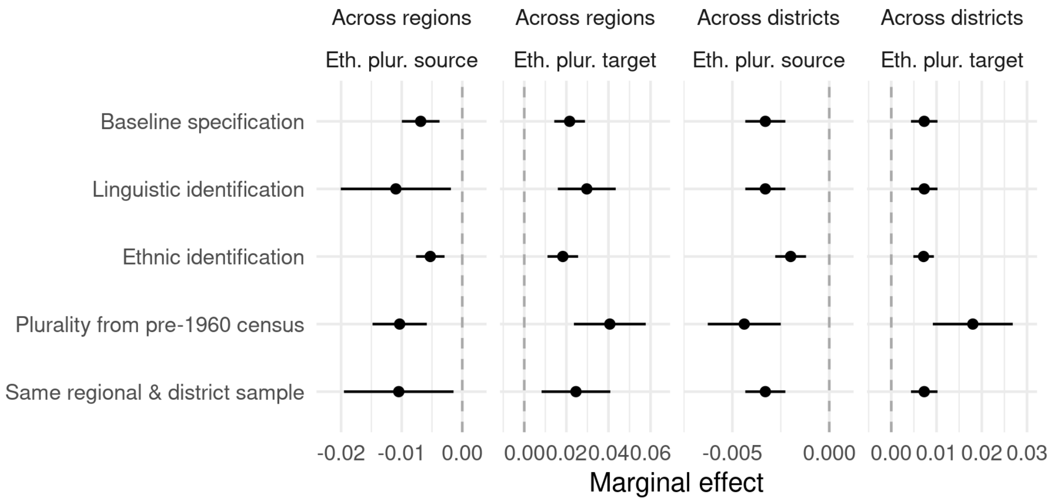



Figure A17: Dyadic migration analysis: Additional results.
Note: Models are based on the dyadic specifications (Models 3 and 6) in Table 3.

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 the region-level analysis.

An assessment of ethnic migration patterns by birth-decade in Figure A18 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 A19 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.

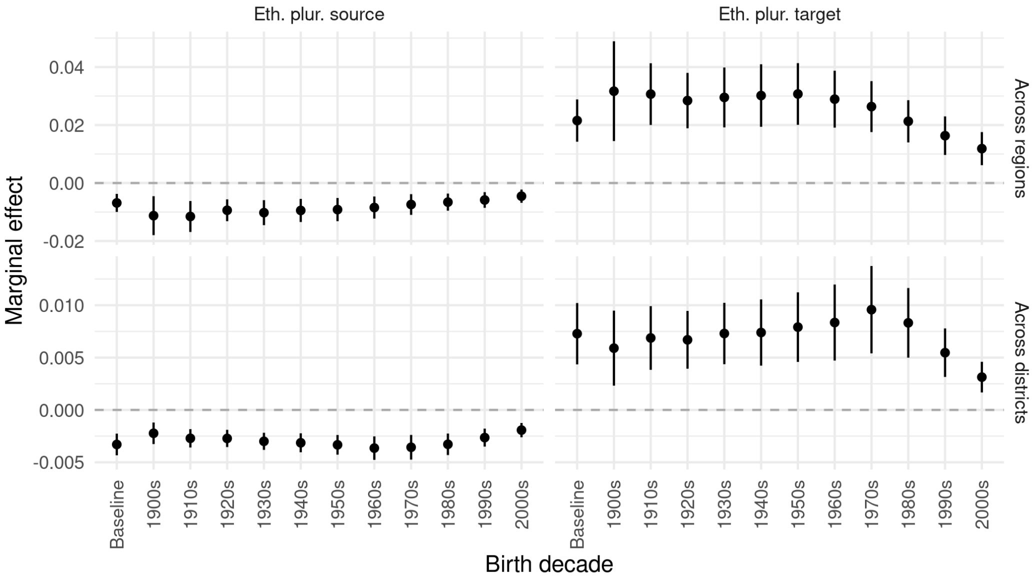



Figure A18: Dyadic migration analysis by birth decade

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

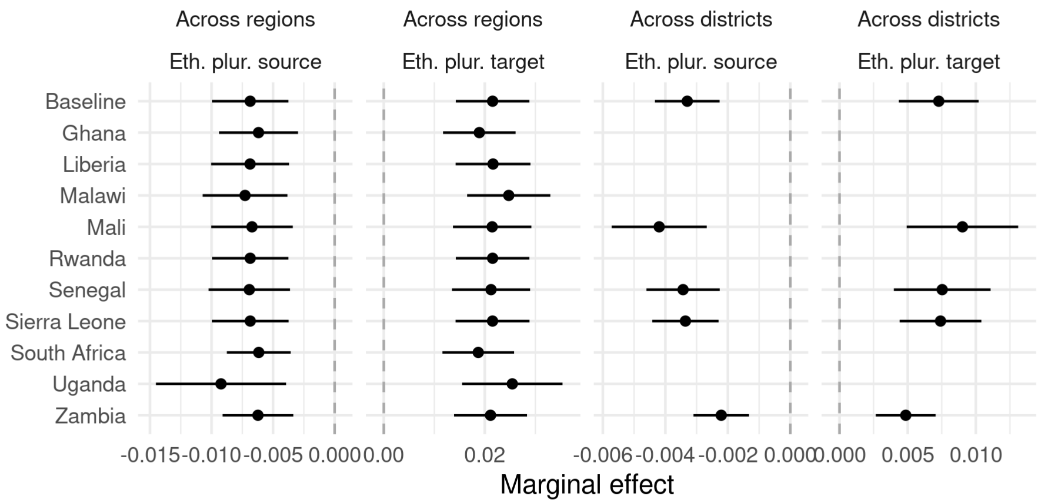



Figure A19: Dyadic migration analysis: Country-level jackknife.

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

## E.2 Correlation of biased migration with RD-estimates

As a last step in the analysis of ethnic migration patterns, I assess whether they credibly contribute to the effects of administrative borders on ethnic demography found in the main empirical section of the paper. To that intent, I proceed in three steps. First, I replicate the main analysis of the effect of regional and district borders on ethnic identities, now using the borders corresponding to the census data to assess discontinuities in DHS respondents' ethnic identification. As shown in Models 1 and 3 in Table A15, 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 $\Delta_u$, estimated in the presence of ethic group and unit fixed effects $\alpha_u$ and $\gamma_e$.[35] 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 $\beta_2$ captures the correlation of the border effect with the migration bias.

This analysis is carried out in Models 2-3 and 5-6 in Table A15 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 demog-

[35]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.

Table A15: Variation in border effects by extent of ethnically 'biased' migration

| | Outcome: Plurality group share (0-1) | | | | | |
|---|---|---|---|---|---|---|
| | Regions | | | Districts | | |
| | Baseline | Emigr. | Immigr. | Baseline | Emigr. | Immigr. |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treated | 0.153*** | 0.118*** | 0.055** | 0.091*** | 0.055* | 0.002 |
| | (0.023) | (0.023) | (0.023) | (0.027) | (0.029) | (0.031) |
| Treated × $\Delta_{u,emigr.}$ | | −0.288** | | | −0.277*** | |
| | | (0.115) | | | (0.104) | |
| Treated × $\Delta_{u,immigr.}$ | | | 0.466*** | | | 0.396*** |
| | | | (0.094) | | | (0.099) |
| Cutoff | 20km | 20km | 20km | 20km | 20km | 20km |
| Running var linear | yes | yes | yes | yes | yes | yes |
| Survey FE: | yes | yes | yes | yes | yes | yes |
| Border FE: | yes | yes | yes | yes | yes | yes |
| Mean DV: | 0.42 | 0.42 | 0.42 | 0.42 | 0.42 | 0.4 |
| 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

raphy 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.

# F 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, forthcoming* .

Müller-Crepon, Carl, Yannick I Pengl and Nils-Christian Bormann. 2019. "Linking Ethnic Data from Africa." *Unpublished Working Paper, https://github.com/carl-mc/LEDA* .

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