Cultural Distance and Ethnic Civil Conflict*

Eleonora Guarnieri

November 19, 2021

Abstract: Ethnically diverse countries are more prone to conflict, yet we lack an understanding of why some groups engage in conflict and others do not. In this paper, I argue that civil conflict is explained by ethnic groups' cultural distance to the central government: an increase in cultural distance increases an ethnicity's propensity to fight over government power. To identify this effect, I leverage within-ethnicity variation in cultural distance to the government resulting from power transitions between ethnic groups over time. I validate my findings in a triple difference-in-differences design using ethnicities partitioned across countries, and through a novel instrumental variables approach. As an instrument for cultural distance, I use differences in ethnic homelands' exposure to the route of the Bantu expansion, a prehistoric migration that shaped culture in sub-Saharan Africa. I show that differences in preferences over public goods are the mechanism driving the effects of cultural distance on conflict. First, cultural distance triggers only conflict over government power, but not conflict over territory or resources. Second, using individual-level survey data, I find that respondents dislike the mix of public policies provided by a culturally distant government. By shedding light on which ethnic groups are more likely to rebel at a given point in time, these findings can inform strategies to target conflict prevention efforts.

Keywords: Ethnic civil war, culture, linguistic distance, Africa, Bantu expansion
JEL Classification: D74, Z1, O55

*Guarnieri: University of Exeter (e.guarnieri@exeter.ac.uk). I am indebted to Mathias Bühler, Davide Cantoni, Helmut Rainer, and Ana Tur-Prats for their helpful comments and patient guidance. I also thank Vojtěch Bartoš, Anke Becker, Graziella Bertocchi, Marc Fabel, Christian Fons-Rosen, Ariel Gomez, Elisabeth Grewenig, Paul Hufe, Andreas Kotsadam, Rainer Kotschy, Panu Poutvaara, Patrick Reich, Brendan Shanks, Fabian Waldinger, and seminar participants at LMU Munich, the EDGE Jamboree, NEUDC 2020, University of Glasgow, University of Milan Bicocca, ECARES, CUNEF, EI University, University of Alicante, IESEG, and University of Exeter for helpful suggestions. All remaining errors are my own.

1. Introduction

Since the end of World War II, civil conflicts have been the most common form of war around the world. Most civil conflicts have taken place in Africa, and the deadliest have been fought along ethnic lines.[1] During one year of the Rwandan Civil War between the Hutu and the Tutsi, at least 500,000 people were killed (Sundberg and Melander [2013]). Globally, civil conflict has caused five times as many deaths as inter-state war, and has led to long-lasting economic and social disruptions (Fearon and Laitin [2003]). Yet, what triggers ethnic civil conflict remains largely debated. A well-established literature has analyzed the role of ethnic-diversity measures at the country level using ethnic fractionalization (Fearon and Laitin [2003], Esteban *et al.* [2012]) and polarization indices (Montalvo and Reynal-Querol [2005], Esteban *et al.* [2012]), and has found that more diverse countries tend to experience more conflict. However, given that all ethnicities in a country face the same aggregate level of diversity, why do some rebel against the central government while others do not? And why do they rebel sometimes and not at other times?

To answer these questions, I proceed in two steps. First, I move the study of diversity and conflict from the country level to the ethnicity level and explore how ethnic groups' cultural *distance* to the central government affects their decision to rebel against it. Second, in line with existing theories of conflict (Esteban and Ray [2011], Spolaore and Wacziarg [2017]), I distinguish between conflicts fought over government power (i.e., public goods) and conflicts fought over territory and resources (i.e., rival goods).[2] I test the hypothesis that cultural distance—defined as differences in beliefs, values, and preferences (Mokyr [2016])—should only explain conflicts fought over government power. By nature, every group in a country is subjected to public policies provided by the government. Different preferences over public policies between an ethnic group and the government may give rise to disagreement and ultimately conflict. Therefore, a larger cultural distance to the government should increase an ethnicity's propensity to fight over government power, but not its propensity to fight over territory. If anything, ethnicities with different preferences and tastes should be less likely to compete over the same type of rival good.

While prominent in the theoretical literature on diversity and conflict (Esteban and Ray [2011], Caselli and Coleman [2013], Spolaore and Wacziarg [2017]), this hypothesis has never been empirically tested at the ethnic-group level in the context of civil war. To do so, I generate a dataset measuring how culturally distant each ethnic group is from the central government at a given point in time. I measure cultural distance using linguistic distance, as is standard in the literature.[3] I follow all ethnic groups over time and construct an indicator for whether they

[1] Ethnic civil conflicts, the focus of this paper, are "armed conflicts between the government of a state and one or more internal opposition groups that cause at least 25 battle-related deaths within a year in which armed groups (i) explicitly pursue ethno-nationalistic aims, motivations, and interests; and (ii) recruit fighters and forge alliances on the basis of ethnic affiliation" (Wimmer *et al.* [2009]).

[2] The data source I use in this paper categorizes each event into one of these two conflict types, i.e., either over government power or territory. Between 1961 and 2017, 34 percent of the whole prevalence of ethnic conflict in Africa was accounted for by conflict over territory, and 66 percent by conflict over power (Sundberg and Melander [2013]).

[3] See, for example, Fearon and Laitin [2003], Desmet *et al.* [2009], Desmet *et al.* [2012], Esteban *et al.* [2012]. The use of linguistic distance is motivated by the notion that language is a salient dimension of culture transmitted through generations (Spolaore and Wacziarg [2016a]). Different languages are the result of horizontal separations

rebel against the government in a specific year or not. I then classify each rebellion as either *conflict over government power* or *conflict over territory*. The resulting dataset is a panel of 236 distinct ethnic groups in 43 African countries observed over a period of 57 years (1961-2017).[4]

In the empirical analysis, I leverage changes in the ethnic identity of the government as a source of within-group time variation in cultural distance, and find support for my hypothesis. After a leadership change, the prevalence of conflict over power increases among ethnic groups that become culturally more distant to the government, and decreases among ethnic groups that become culturally closer. These effects cannot be attributed to differential trends in conflict prior to a government change across these two subgroups of ethnicities. The reaction to cultural distance occurs immediately after a change in leadership, and is large in magnitude: a one standard deviation increase in cultural distance increases the prevalence of conflict over government power by 0.36 standard deviations.[5] Conversely, I do not find a significant association between cultural distance to the government and conflict over territory.

My results hold after conditioning on a rich set of fixed effects. The data allows me to include ethnicity fixed effects, thus isolating the effect of cultural distance from any time-invariant ethnic-specific characteristic potentially associated with conflict: cultural norms, social structure, ancestral arrangements and geographic conditions, whether the group was split by country borders, or a general propensity to rebel. Since the source of within-ethnicity variation in cultural distance results from government changes, the inclusion of country-by-year fixed effects ensures that the coefficients are not confounded by government characteristics, or by country-level time shocks affecting all ethnicities within a country, including the overall effect of a leadership change. Moreover, country-by-year fixed effects allow me to keep constant the aggregate level of diversity (e.g., a country's level of ethnic fractionalization or polarization) and focus on the effect of cultural distance between groups.

While I interpret my findings as the effect of cultural distance on conflict, I consider alternative explanations. First, I show that the effect of cultural distance is not confounded by an ethnicity's representation in the government coalition, or by an ethnicity having lost or gained power. Second, the effect is not due to confounding geographic or climatic differences between an ethnic group and the ethnicities forming the government in a specific year. The estimates remain large and significant when controlling for ethnic differentials in geographic location, elevation, ruggedness, caloric suitability, mean temperature, and average yearly precipitation. Third, the results are not systematically driven by an ethnic group's conflict behavior prior to a government change. Taken together, these findings show that the estimates reflect the effect of cultural distance, and not of other confounding factors.

Next, I validate my findings using a tightly-controlled empirical exercise on a restricted sample

between populations, and these separations are likely to go hand in hand with cultural divergence.

[4] I focus on Africa for two reasons. First, it constitutes a unique setting due to the high degree of ethnic diversity within countries, which leads to the highest-quality data on ethnicity (e.g., the Linking Ethnic Data for Africa (LEDA) allows me to link several datasets on ethnicity and is only available for Africa). Second, for the purpose of identification, Africa offers two unique natural experiments I exploit in my analysis: the Bantu expansion and the random allocation of country borders.

[5] For comparison, this corresponds to three times the effect of an increase in resource inequality uncovered by Guariso and Rogall [2017]. They show that a one standard deviation increase in rainfall inequality between an ethnicity and the government increases the likelihood of conflict by 0.12 standard deviations.

of ethnic groups. My baseline specification does not control for unobserved time-varying, ethnic-specific shocks. If these occur simultaneously to a change in government and are correlated with an ethnic group's decision to rebel, they could confound my results. To address this issue, I restrict the analysis to the subset of ethnicities that were split by country borders during the Scramble for Africa and are therefore simultaneously exposed to different governments in different countries. While this subsample is arguably peculiar due to the documented long-lasting effects of ethnic partitioning on conflict (Michalopoulos and Papaioannou [2016]), focusing on this offers some advantage for the purpose of identification.

I use a triple difference-in-differences (DID) design that exploits within-ethnicity variation in linguistic distance to the government resulting from (i) a group's concurrent exposure to multiple governments in multiple countries and (ii) changes in the ethnic identity of the government in some countries but not others. Combined, these features allow me to enrich the baseline specification with the inclusion of ethnicity-year fixed effects.[6] Reassuringly, the results from the triple DID analysis are very similar to the ones from my baseline specification. A one standard deviation increase in cultural distance to the government increases the prevalence of conflict over government power by 0.39 standard deviations. A one standard deviation increase in cultural distance, conversely, reduces conflict over territory by 0.34 standard deviations. In line with the original hypothesis, this suggests that conflict over territory tends to be more likely between culturally close groups that share similar preferences.

One remaining potential concern with the analysis is the presence of omitted factors affecting both cultural distance and conflict. For instance, historical conflict between two ethnicities might have a direct impact on the probability of current conflict between them, and at the same time, might have affected their linguistic distance. I address these endogeneity concerns using a novel instrumental variables (IV) approach. To instrument for cultural distance, I exploit the Bantu expansion, a massive pre-historical migration that changed language and culture in some parts of sub-Saharan Africa but not others.[7] To my knowledge, this is the first paper exploiting this prehistoric cultural shock as a source of exogenous variation in cultural distance between today's African ethnicities.

I construct and assign to each ethnicity a *Bantu index*, which captures the extent to which ethnic homelands were exposed to the route of Bantu migration.[8] I then instrument for cultural distance between two ethnic groups using the absolute difference in their Bantu index. The idea behind this instrument is the following. Groups whose homelands were highly exposed to the Bantu migration route inherited Bantu culture, and should be culturally distant to those

[6]This approach is analogous to the one employed by Dickens [2018]. Michalopoulos and Papaioannou [2014] were the first exploiting the partition of African ethnicities in multiple countries as a source of within-ethnicity variation.

[7]Bantu farmers imported their culture capital into some of the territories originally inhabited by Pygmy and Khoisan hunter-gatherers. Archaeological evidence, despite being inconclusive, suggests that Pygmy people inhabited Central Africa, and Khoisan people inhabited sub-equatorial Africa (Diamond [1997]). Both were hunter-gatherer societies, unlike the Bantu people, who were farmers. In gradually expanding from their homeland in Cameroon, Bantu people followed emerging savannah corridors, which appeared in the rainforest due an exogenous climatic shock and then disappeared once climate re-adjusted (Grollemund *et al.* [2015]).

[8]To generate the Bantu index, I merge contemporary ethnicities to their ancestral homelands in the Murdock map, the most comprehensive ethnolinguistic map of Africa prior to European contact, and overlay ethnic groups' homelands on the historical migration route of the Bantu reconstructed by Grollemund *et al.* [2015].

whose culture remained unaffected. Conversely, two groups with a similarly high or similarly low Bantu exposure should be culturally close to each other, because they either both inherited Bantu culture, or they both kept their pre-existing one.[9] Consistent with this idea, the first stage documents a strong positive association between the absolute distance in the Bantu index and cultural distance. In the second stage, the effect of cultural distance on conflict over power becomes larger in magnitude when compared to the OLS estimates. A one standard deviation increase in cultural distance increases the prevalence of conflict by 0.86 standard deviations.[10] In the case of conflict over territory, the negative coefficient reaches statistical significance in some specifications. A one standard deviation increase in cultural distance decreases the prevalence of conflict over territory by 0.11 standard deviations.

In a last step, I explore factors that may explain why cultural differences trigger conflict over the control of the central government. I show that cultural differences are associated with diverging preferences over public policies. In particular, using individual-level data from seven rounds of the Afrobarometer survey for more than 80,000 individuals in 27 African countries, I find that respondents are more likely to oppose a wide range of current government policies if they belong to ethnicities that are culturally distant to the ethnic groups in power at the time of the survey. A one standard deviation increase in cultural distance increases discontent with government performance by 0.44 standard deviations.

This paper contributes to several strands of literature. First, it adds to the vast literature on civil war (see Blattman and Miguel [2010] for a review). Among studies focusing on ethnic conflict, some have highlighted time-invariant factors that make certain ethnicities more likely to experience conflict, like segmentary lineage organization (Moscona *et al.* [2020]), inter-personal diversity (Arbatlı *et al.* [2020]), and ethnic partitioning (Michalopoulos and Papaioannou [2016]). However, even after acknowledging these different propensities to experience conflict, what drives ethnic groups to fight against other specific ethnicities at a given point in time remains unclear. By focusing on cultural distance between ethnic groups and governments, I relate ethnicities to their potential opponents and show that the prevalence of conflict is a function of the characteristics of both combatants, and not only of the characteristics of one side. Two studies adopt a similar approach. Guarnieri and Tur-Prats [2020] focus on the intensive margin of violence, and show that cultural distance in gender norms between ethnic belligerents increase the use of conflict-related sexual violence. Guariso and Rogall [2017] show that income inequality between an ethnic group and the leading group in a country increases the likelihood of ethnic conflict between them. Whereas Guariso and Rogall's [2017] focus is on the economic drivers of conflict, my paper explores the role of deep-rooted cultural determinants.

By uncovering that cultural distance increases conflict over government power but not conflict over territory, this paper highlights the importance of considering the issue over which com-

[9]The assumption here is that, within contemporary countries, pre-Bantu groups were on average culturally similar to each other, and culturally dissimilar to the Bantu. This is partly supported by archaeological evidence showing that pre-Bantu societies shared the same mode of subsistence, i.e., hunting and gathering, while Bantu people were predominantly farmers (Diamond [1997]).

[10]This suggests that the original coefficients were biased downwards. This would be consistent, for instance, with the following: past conflict between two ethnicities is positively associated with contemporary conflict, and, at the same time, reduced two groups' linguistic distance through genocides eliminating linguistically distant subgroups.

batants fight when studying the determinants of conflict. In investigating the economic roots of local violence across Africa, McGuirk and Burke [2020] also distinguish between two types of conflict, and find that this distinction matters. They show that, in food-producing areas, the same positive global food price shock reduces conflict over the control of territory (what they call "factor conflict"), but increases conflict over the appropriation of surplus ("output conflict"). Campante *et al.* [2019] show that, in relatively non-democratic countries, conflict over government power is more likely to occur close to the capital city and that this does not hold for conflict over territory.

This study is closely related to theoretical and empirical work on ethnic diversity and conflict, summarized in the next section. Empirically, this literature has mostly focused on the relationship between diversity indices and conflict at the country level (Esteban and Ray [2011] and Esteban *et al.* [2012]). This country-level approach, however, remains silent on the group-specific heterogeneity in the decision to fight. By moving the analysis to the ethnicity-level, I unpack the country-level associations and identify precisely which ethnicities engage in conflict, when, and why.

More broadly, this paper adds to the literature on the consequences of ethnic diversity. In addition to studies focusing on conflict, others have analyzed a large range of economic outcomes (see Alesina and La Ferrara [2005] for a review), adopting as unit of analysis either countries, cities, or grid cells (see, for instance, Montalvo and Reynal-Querol [2017]). My study is part of a new line of work that keeps the aggregate level of diversity constant and focuses on distances between entities within aggregate units. While I study ethnicities within countries, Gomes [2020] examines individuals within geographical radii, and finds that children of mothers who are linguistically distant from their neighbors have worse health outcomes.

Finally, I contribute to the literature exploring the determinants of ethnolinguistic diversity. Michalopoulos [2012] finds that contemporary ethnic diversity is rooted into geographic variability. Work by Ashraf and Galor [2013] shows that genetic diversity, determined during the prehistoric migration of humans out of Africa, is a fundamental determinant of ethnic heterogeneity within countries today. Galor *et al.* [2018] explore the precise geographic roots of specific cultural and linguistic traits like the structure of the future tense, sex-based grammatical gender, and politeness distinctions. A recent contribution by Dickens [2020] studies linguistic distance between neighboring groups. This study finds that ethnicities separated across geographic regions with high variation in land productivity are more similar than ethnicities separated across more homogeneous regions, due to higher levels of historical trade. I speak to this literature by linking the prehistoric Bantu migration to cultural distance between ethnic groups today.

The remainder of the paper is structured as follows. Section 2 describes the conceptual framework linking cultural distance to inter-group conflict. Section 3 outlines the data sources and the procedure I adopted to construct the dataset. Section 4 describes the empirical strategy, and Section 5 shows the main results together with a battery of robustness tests. Sections 6 and 7 present the triple difference-in-differences and the instrumental variables estimation strategies, respectively, and the corresponding results. Section 8 discusses potential channels, and Section 9 concludes.

2. Conceptual Framework

The link between diversity and conflict has been debated by a large strand of interdisciplinary literature. This debate originates from the so-called primordialist view of conflict, according to which dissimilarities between groups spur conflict.[11] An anthropological formulation of this view posits that a society formed by culturally divergent groups can only be sustained through a political order in which one of the cultural groups politically dominates the others (Smith [1965]). In turn, this political structure inevitably generates dissensus, given that all groups within a jurisdiction are bound to the public policies provided by a culturally distant dominant group. According to Smith's [1965] theory of cultural pluralism, it is this *cultural dissensus* that generates ethnic conflict in pluralist societies. Yet, cultural differences might also impede conflict, by "focusing the ambitions of various groups on alternative sources of gratification, thereby preventing them from impinging on each other" (Horowitz [2000], p. 138). Taken together, these opposing views already hint at the potentially ambiguous relationship between cultural differences and conflict.

Literature in economics has recently shed further theoretical and empirical light on these issues. In their study of inter-state conflict, Spolaore and Wacziarg [2016b] provide empirical evidence in support of Horowitz's [2000] conjecture. They find that inter-state conflict is more likely between similar populations, because they tend to share preferences over the same type of resources or territories—the typical bone of contention in international conflict.

In analyzing the relationship between alternative diversity measures at the country level and the equilibrium level of conflict, work by Esteban and Ray [2011] and Esteban *et al.* [2012] highlights that distinguishing between conflict over private goods and conflict over public goods is essential for understanding the interplay between diversity and conflict. Theoretically and empirically, they show that a country's severity of conflict can be approximated by a weighted average of three country-level measures of diversity: a Gini coefficient, an index of ethnic fractionalization, and a measure of ethnic polarization. The weights of each of these measures depend on a country-level measure of "publicness of the prize".

To elucidate this concept, Esteban and Ray [2011] discuss the following examples. In a situation in which groups fight for the control of an excludable private good, "primordial" inter-group distance—best captured by a measure of polarization—does not constitute a relevant explanation for this conflict. The only explanatory factor will be a country's distribution of group sizes, best captured by an index of fractionalization. At the other extreme, when a prize is fully public, differences in group preferences over public goods will emerge. This type of conflict will be better explained by indexes of polarization and Gini coefficients, since they capture the degree of inter-group distances within a country.

Along similar lines, Spolaore and Wacziarg's [2017] theoretical framework highlights that the impact of cultural distance on conflict depends on whether groups are fighting over the control of public goods or rival goods. Conflict over the control of public goods—e.g., public policies that all must share within a jurisdiction—is more likely between culturally distant groups that

[11] This view is often juxtaposed to the so-called instrumentalist view, which sees ethnic conflict as a result of grievances and inequality between groups, rather than the results of different ethnic identities.

hold different preferences. This explanation is entirely consistent with Esteban and Ray's [2011] theory. However, while in Esteban and Ray [2011] distance does not matter for conflict over excludable private goods, Spolaore and Wacziarg's [2017] framework claims that conflict over rival goods—e.g., territories and resources—will be more likely among culturally close groups that share similar preferences. I bring this hypothesis to the data and, for the first time, simultaneously test these two predictions at the ethnic group level in the context of ethnic civil conflict.

3. Data and Descriptive Statistics

3.1. Sources and Dataset Construction

To examine the relationship between cultural distance and conflict, I assemble a unique dataset measuring how culturally distant an ethnic group is from the government of a country in a specific year, which groups engage in violent conflict at a certain point in time, and whether they are fighting for government power or for exclusive control of a territory. I furthermore add information on climatic and geographical features, as well as on ancestral ethnic settlements of contemporary ethnic groups and their exposure to the route of the Bantu expansion. In this section, I describe the procedure I adopted to construct the dataset.

Ethnic groups. The first step consists of retrieving information on ethnic groups forming a country's population. To this end, I exploit the Ethnic Power Relations (EPR) Dataset Family (Wucherpfennig *et al.* [2012]). For each country of the world, EPR lists all politically relevant ethnic groups between 1946-2017 and their access to government power. A group is defined as *politically relevant* if one political actor claims to represent the interests of a group in the national political arena, or if group members are systematically and intentionally discriminated against. EPR codes the degree to which each ethnic group's representatives hold executive-level state power in each year, which ranges from total control of the government to political discrimination.[12] Each group falls into one of the following categories, in decreasing order of power: monopolist, dominant, senior partner, junior partner, self-exclusion, powerless, discriminated.[13] Based on this data, I construct a panel that follows each ethnic group over time.

Ethnic civil conflicts. To construct the dependent variable, I use the UCDP/PRIO Armed Conflict Dataset (Harbom *et al.* [2008]; Allansson *et al.* [2017]), which includes information on civil conflicts and the actors involved in them between 1946 and 2019. The UCDP/PRIO dataset defines civil conflicts as "armed conflicts between the government of a state and one or more internal opposition groups that cause at least 25 battle-related deaths within a year". I focus on *ethnic* civil conflicts in Africa fought between 1961 and 2017,[14] those where armed groups

[12] Whenever government changes happen in the same year as a conflict, EPR reports the power relations that were in place before the conflict outbreak.

[13] Additional categories include state collapse and irrelevant. I exclude the (rare) instances of state collapse from the analysis as, in those years, there is no government; I exclude irrelevant groups for the analysis as, for those, there is no information on their settlements.

[14] I restrict the analysis to this time period as most African countries gained independence in the 60s, and

"explicitly pursue ethno-nationalistic aims, motivations, and interests; and recruit fighters and forge alliances on the basis of ethnic affiliation" (Wimmer *et al.* [2009]).

UCDP/PRIO provides information on the issue over which rebel groups fight. First, a conflict can be about *government*, i.e., about the type of political system, the replacement of the central government, or the change of its composition. Second, a conflict can be about *territory*. While groups fighting for government intend to impose their authority over others, groups fighting for territory do not want to exercise power over other groups, but strive for exclusive control of a specific territory for own settlement, local resource use, or, in extreme cases, secession. In the first case, the nature of the bone of contention is public, while in the latter case, the good combatants are fighting over is rival and excludable.

Examples of struggles over state power include the case of the Liberians United for Reconciliation and Democracy (LURD) (2000-2003) against the government led by Charles Taylor, or the Tigray People's Liberation Front (TPLF) in Ethiopia (1976-1986) fighting against the hegemony of the Amharan government. Conflicts fought over territory include the one between the Movement of Democratic Forces of Casamance (MFDC) and the government of Senegal for the control of the Casamance region (1990-2011). Another example is the rebellion of the Liberation of the Enclave of Cabinda-Renewal (FLEC-R) and the Armed Forces of Cabinda (FLEC-FAC) against the government of Angola over the control of the Cabinda region (1991-2017). This category also includes cases where rebel groups seek control over specific resources linked to a certain territory, such as the Niger Delta People's Volunteer Force (NDPVF) fighting for controlling the petroleum resources of the delta region in Nigeria (2004).

Rebels' ethnic identity. To assign rebel groups an ethnic identity, I exploit the ACD2EPR dataset (Vogt *et al.* [2015]). This dataset assigns to each rebel group one or more ethnic groups in the EPR dataset. To continue some of the examples above, members of the Mandingo and Krahn ethnic groups formed the LURD rebel group in Liberia; the Tigry ethnic group formed the TPLF rebel group in Ethiopia; and the Ijaw group formed the NDPVF in Nigeria. Next, I merge this information to the panel of ethnicities, where each ethnic group constitutes a *potential rebel*, and construct a binary variable that is equal to 1 if a potential rebel is involved in a conflict against the central government in a certain year, and zero otherwise. All groups within a country can be potential rebels, apart from *monopolists* or *dominant* ones. By definition, these groups cannot rebel against themselves, because they have the exclusive control of the government.[15] Whenever an ethnicity becomes dominant or monopolist, it drops out of the panel.[16] In a robustness test, I re-run the analysis excluding these ethnicities, and show that the results remain unchanged.

Governments' ethnic identity. To assign an ethnic identity to the government in a specific year, I exploit information provided by the EPR dataset, which indicates which ethnic groups hold executive government power in a country at a certain point in time. EPR takes into

climatic data is available only starting from 1961. I exclude the most recent year because the EPR data is available only until 2017.

[15]Indeed, there are no instances in which a rebel group involved in a conflict is associated with a dominant or monopolist ethnicity.

[16]Note that these cases are quite rare. Only 14 ethnic groups temporarily exit the sample of potential rebels because they become monopolists or dominant.

consideration where executive power is exercised when coding ethnicities' access to it. This includes the presidency, the cabinet, and senior posts in the administration, including the army, or the ruling party leadership in one-party states. In some cases, government power is held exclusively by one ethnic group, that EPR classifies as either *monopolist* or *dominant*. In other cases, the government results from a coalition of ethnic groups, which can take either the role of *junior* or *senior partner*. Whether a group is classified as senior ot junior partner depends on the group's absolute influence in the executive, measured by the number and importance of the positions controlled by group members.

Cultural distance. To measure cultural distance between each potential rebel and the government, I use linguistic distance, as is standard in the literature. The source of information on languages is the EPR Ethnic Dimensions (EPR-ED) dataset. EPR-ED assigns up to three languages to each EPR ethnic group, indicating the three largest language segments spoken by group members in descending order and their relative size.[17] The advantage of this dataset is that it provides a nuanced description of the linguistic (and therefore, cultural) identity of an ethnicity also when the latter is not homogeneous. Therefore, my measure of cultural distance incorporates this intra-group heterogeneity.

I merge languages to linguistic trees in the Ethnologue database (Lewis *et al.* [2020]). For each language, the Ethnologue provides a classification starting with the language family (e.g. Afro Asiatic, Nilo-Saharan, Creole), followed by "nodes", i.e., the branching points of the linguistic tree, and ending with the language itself. To compute linguistic distance between each potential rebel and the government, I employ a three-steps procedure.

First, I calculate the distance between each pair of languages (x and y) as follows:

$$d_{xy} = 1 - \left(\frac{\#\text{ of common nodes between } x \text{ and } y}{\frac{1}{2}(\#\text{ of nodes of language } x + \#\text{ of nodes of language } y)} \right)^{\lambda}$$

This measure, called cladistic distance, is the most frequently used in the literature (Fearon and Laitin [2003]). Languages originating from different families have no nodes in common, and their distance will be equal to 1. The parameter λ ranges between 0 and 1, and is used to attribute higher weight to earlier common nodes, as early separations in the language tree are likely to signify larger cultural divergence on average than later separations. I follow Fearon and Laitin [2003] and assign to λ a value of 0.5.

Second, I calculate linguistic distance between potential rebels (r) and each ethnic group at the government (g_i):

$$LD_{rg_i} = \sum_{x=1}^{3} \sum_{y=1}^{3} (w_{rx} \times w_{g_i y} \times d_{xy}) \tag{1}$$

where w_{rx} and $w_{g_i y}$ are the fraction of population speaking language x in group r and language y in group g_i, respectively.

[17] These refer to indigenous African languages, not those imported by colonizers.

Third, I compute linguistic distance between each potential rebel and the government:

$$LD^W = \sum_{i=1}^{N} p_{g_i} \times LD_{rg_i} \tag{2}$$

where N is the total number of ethnicities forming the government, and p_{g_i} is a weight reflecting the position of power of group g_i in the government. When the government is composed only by one dominant or monopolist ethnic group, LD^W equals LD_{rg_i}. In the case of a government coalition, I assign a higher weight to linguistic distance between each potential rebel and the senior partner.[18] Senior partners will receive double the weight of each junior partner.[19]

Alternatively, I use an unweighted version of the just described linguistic distance measure:

$$LD^{UW} = \sum_{i=1}^{N} \frac{LD_{rg_i}}{N} \tag{3}$$

Ancestral settlements. I merge EPR ethnicities to ethnic groups in the Murdock Ethnographic Atlas (EA). This allows me to assign ethnic groups to the location of their ancestral homelands prior to European contact, which was recorded by Murdock. I merge contemporary ethnic groups to pre-colonial ethnic groups and respective settlements through the Linking Ethnic Data for Africa (LEDA) R-package (Müller-Crepon *et al.* [2020]), which systematically links ethnic categories across datasets exploiting the Ethnologue linguistic tree.[20] The definition of ethnic group in the EA is more fine-grained than the one in the EPR dataset. Therefore, one ethnic group in EPR is generally matched to multiple groups in the Murdock map.

Bantu expansion route. To estimate the exposure of each ethnic group to the Bantu expansion, I digitize the Bantu expansion route reconstructed by Grollemund *et al.* [2015] (see Figure 4, left). I then overlay the path onto the Murdock map. I leverage ethnic groups' ancestral homelands to calculate the exposure to the Bantu expansion for the following reasons. First, EPR provides ethnic settlements for contemporary ethnic groups, which are likely the result of recent phenomena endogenous to conflict, as well as of modern-day country borders imposed artificially by colonial powers. Second, the geographic locations of ethnic groups in EPR capture groups' regional presence and contemporary settlements, rather than homelands as in the Ethnographic Atlas.

Geographic and climatic controls. The GeoEPR dataset provides polygons describing each ethnic group's geographic location, allowing overlaps between different groups' settlements.[21]

[18] Rebels may be intentionally fighting against one specific group in the coalition, or against the government as a whole. Since I do not have systematic information on this issue, I consider all ethnic groups forming the government.

[19] In the example illustrated in Table 1, in the distance between Langi/Bakongo and the government, p_{g_i}=0.5 when g_i=Mbochi (the senior partner), and p_{g_i}=0.25 when g_i=Bateke or Kouyou (the junior partners).

[20] For additional details on the methodology, see Müller-Crepon *et al.* [2020]. For details on how I use the LEDA R-package, see Appendix B.

[21] GeoEPR does not provide the location of groups classified as living exclusively in urban areas. To these, I

To add information on geographic and climatic characteristics to each settlement, I combine GeoEPR with data on elevation, temperature, and precipitation at the grid level provided by Worldclim (Fick and Hijmans [2017]). Furthermore, to add information on the potential agricultural output, I use the Caloric Suitability Index (CSI) (pre-1500 average) from Galor and Özak [2016]. This measure reflects the potential caloric yield of a territory using a combination of geographical and climatic data unaffected by human activity.

Opinions on government performance. I merge ethnicity-level data to individual-level information on opinions about government performance collected in 7 waves of the Afrobarometer Survey (Afrobarometer Data [1999-2017]). I link EPR groups to the self-reported linguistic affiliation of individuals in the Afrobarometer through the LEDA R-package (Müller-Crepon *et al.* [2020]).

Figure 1 and Table 1 illustrate the structure of my data and the identifying variation with an example from Congo. First, as shown in Figure 1, there are six politically active ethnic groups: Lari/Bakongo, Kouyou, Mbochi (proper), Bateke, Bemba, Vili. The map illustrates where each ethnic group is located within the country. I follow each of these ethnicities over time, and track their involvement in conflict as well as their cultural distance to the government. Table 1 reports a data extract for the ethnic group Lari/Bakongo between 1995 and 2000. The within-ethnicity variation in cultural distance I exploit is driven by government changes: in this case, Lari/Bakongo experienced an increase in linguistic distance in 1998, following the Mbochi group replacing the Bembe as a senior partner in the coalition, and the Vili group dropping out of power. After this change, Lari/Bakongo started engaging in conflict against the government.

3.2. Summary Statistics

The resulting dataset includes 236 ethnic groups in 43 African countries over a period of 57 years (1961-2017) (N=9,827).[22] The source of within-ethnicity variation in cultural distance stems from 138 changes in the ethnic composition of the government. Table B-1 reports the number of changes experienced by each country in the sample. Ten countries out of 43 did not experience any change in the ethnic composition of the government in the period considered. The largest part of the variation comes from West Africa, with countries such as Niger, Nigeria, and Congo experiencing the highest number of government transitions.

Table B-2 reports descriptive statistics of the main variables used in the analysis. In Africa as a whole, 29 countries experienced at least one conflict over the period considered, with an overall prevalence of ethnic conflict of 0.073. Of the 236 potential rebels in the sample, 73 became rebels at least once, and 60 distinct governments were involved in an ethnic civil conflict at least once. 34 percent of the overall conflict prevalence is explained by conflicts fought for territory.

Within countries in Africa, linguistic distance between potential rebels and the government

assign the coordinates of the country's capital city. These constitute 2% of ethnicities in my sample.

[22] I exclude countries where there is only one ethnic group (Burkina Faso, Cabo Verde, Lesotho, Somalia, and Tunisia), and countries where there is only one potential rebel (Rwanda, Madagascar, Algeria). I do not include Sao Tome and Principe and Seychelles since they are not part of the EPR dataset family.

is on average 0.40-0.45, depending on the distance type considered. Potential rebels and the government are culturally close in Malawi, with an average linguistic distance of 0.06, while potential rebels are culturally very distant from the government in Liberia (0.90 on average). The Liberian example speaks to the accuracy of linguistic distance in capturing cultural differences. The Americo-Liberians were the group in power over almost the whole period considered. This group originates from free-born and formerly enslaved African Americans who emigrated in the 19th century and became the founders of Liberia. They imported African-American and Caribbean culture, and are therefore culturally different from the other Liberian groups (Gio, Mano, Mandingo, Kran (Guere), and Indigenous peoples).

4. Empirical Strategy

The objective of the empirical analysis is to investigate whether changes in cultural distance to the government have an effect on ethnic groups' involvement in ethnic civil conflict. To illustrate the idea behind the empirical strategy and get a first sense of how the prevalence of conflict reacts to government changes, in Figure 2 I plot the average residual variation in conflict over power and conflict over territory—net of country-year fixed effects—before and after a government change. I do so separately for two subsamples of ethnicities: groups that become culturally more distant to the government; and groups that become culturally less distant to the government following a change in leadership. Panel A shows that a group's involvement in conflict over power reacts to cultural distance: groups are more likely to rebel once they become culturally more distant to the government, and less likely once they become culturally closer. Conversely, Panel B illustrates that groups' involvement in conflict over territory does not respond to cultural distance. Importantly for identification, both panels indicate that the two subgroups of ethnicities do not display different trends in conflict involvement prior to a government change.

To measure the impact of changes in cultural distance on conflict in a regression framework, I estimate the following linear probability model:

$$\text{Conflict}_{rct} = \lambda_{c,t} + \zeta_r + \theta_r t + \beta \text{LD}_{rct} + \Gamma \text{G}_{rct} + \Omega \text{C}_{rct} + \epsilon_{rct} \tag{4}$$

where the dependent variable is an indicator that takes value 1 if there is a rebel group of ethnicity r fighting the government of country c in year t. LD_{rct} indicates linguistic distance between ethnic group r and the government of country c in year t.

$\lambda_{c,t}$ denotes a full set of country-year fixed effects. These capture time-invariant characteristics at the country-level—e.g., colonial history or geography—that might make some countries overall more or less prone to conflict than others. Furthermore, they account for time-specific shocks common to all ethnic groups in a country and for a specific government's characteristics. These include whether the government represents a large versus a small share of the population; whether it is a coalition of groups or formed by a single dominant one; whether it is a bellicose or peaceful government; the quality and type of its public policies; the strength of the army; or whether a certain government is established as a result of a conflict and thus is more vulnerable

to retaliation from opposing groups. Finally, country-year fixed effects allow to isolate the effect of cultural distance on conflict from the effect of a country's aggregate level of fractionalization and polarization.

Crucially, my dataset also allows me to include a full set of ethnicity fixed effects (ζ_r), as well as ethnicity-specific year trends ($\theta_r t$). By adding ethnicity fixed effects, I control for any time-invariant characteristics specific to an ethnic group that might affect its propensity to experience conflict. These include all ethnic traits that have been associated with conflict by the literature like social structure (Moscona *et al.* [2020]), within-group inter-personal diversity (Arbatlı *et al.* [2020]), or ethnic partitioning (Michalopoulos and Papaioannou [2016]). Moreover, these fixed effects ensure that the variation I exploit is within-ethnicity variation in cultural distance resulting from government changes (see example in Table 1). By controlling for an ethnicity-specific year trend, I can furthermore account for dimensions at the ethnicity level that change linearly over time, and allow each ethnic group to follow different trends.[23]

Finally, I control for a set of geographic and climatic controls. G_{rct} is a vector of differences between potential rebels and the government: geodesic distance, distance in elevation, ruggedness, and caloric suitability index. Dickens [2020] shows that geographical differences are an important determinant of linguistic distance, as neighboring groups that share homogeneous environments are more likely to be linguistically closer to each other. These findings speak to the importance of this vector of controls when seeking to isolate the effect of cultural distance from the effect of geography. I also include a set of differences in precipitation and mean temperature (C_{rct}) to account for the effect of rainfall inequality on ethnic conflict uncovered by Guariso and Rogall [2017]. I also control for time-variant climatic variables (temperature and precipitation) specific to each potential rebel to account for the potential confounding effect of climatic shocks, which are important determinants of conflict in the African continent according to Harari and La Ferrara [2018].[24]

5. Main Results

Table 2 reports estimates from the linear probability model described in equation 4, using as outcome variable a binary measure of conflict, either fought over government power or territory. Column 1 shows that within a country and a year, the larger the linguistic distance between an ethnic group and the government, the more likely it is that the group engages in conflict against the ethnicities in power. In this specification, without the inclusion of ethnicity fixed effects and ethnicity time trends, a one standard deviation increase in linguistic distance increases conflict prevalence by 1.3 percentage points (0.05 standard deviations). This specification, which ex-

[23]Note that this specification is equivalent to a difference-in-differences strategy where treatment is the change (positive or negative) in cultural distance stemming from government changes. I compare groups that experience a positive change in cultural distance to those that experience a negative change. The inclusion of unit-specific time trends allows me to relax the common trends assumption typical of DID designs. See a similar application in Pfeifer and Reutter [2020]. Another assumption for a causal interpretation of my results is that treatment effects have to be homogeneous across ethnic groups within a country, as well as over time (Goodman-Bacon [2018]).

[24]Given the dynamic nature of the GeoEPR dataset, ethnic settlements may vary over time, making geographic characteristics also time-variant. However, since ethnicity fixed effects absorb the variation almost entirely, I do not control for potential rebels' elevation, ruggedness, and CSI to avoid collinearity issues.

ploits within-country cross-ethnicity variation in cultural distance, does not account for plausibly relevant ethnic-specific confounding factors.

To account for these, I add ethnicity fixed effects in column 2, and now exploit within-ethnicity variation in cultural distance originating from changes in the ethnic composition of the government. The coefficient remains positive and significant, and increases in magnitude. A one standard deviation increase in cultural distance increases conflict prevalence by 6 percentage points (0.23 standard deviations). This coefficient remains significant and similar in magnitude when adding geographic and climatic controls (columns 3 and 4) and when using an unweighted measure of linguistic distance to the government (column 5).

In Table 3 I test my main hypothesis, i.e., whether these associations mask heterogeneous effects depending on what combatants are fighting over. I find that the just described effects are entirely driven by conflicts fought over government power. In my preferred specification in column 2 of Table 3, a one standard deviation increase in cultural distance increases conflict over power by 7.5 percentage points (0.36 standard deviations). Conversely, I do not find that increases in cultural distance to the government trigger conflict over territory (columns 5 to 8). The negative sign of the insignificant coefficients suggests that, if anything, the association reverses: a potential rebel is less likely to fight for territory when experiencing an increase in cultural distance to the government. These heterogeneous effects are in line with my hypothesis: an ethnicity rebels to gain power when disagreeing with the mix of public policies provided by a culturally distant government. In Section 8. I will provide direct evidence on this specific mechanism.

5.1. Leads and Lags and Anticipatory Effects

In this section, I explore the persistence of the effect of cultural distance on conflict, and test for anticipatory effects. Table 4 presents the results of these exercises. Column 1 reproduces the baseline estimation for linguistic distance in t for reference. To explore whether the effect persists, column 2 reports the estimate for a lagged measure of linguistic distance. The positive and significant coefficient suggests that the effect of a government change on conflict persists, but tends to fade away: the magnitude of the coefficient in column 2 almost halves, when compared to the one in column 1.[25] When including both contemporaneous and lagged linguistic distance simultaneously in column 3, only the contemporaneous measure remains positive and significant, suggesting that ethnic groups' reaction to a culturally distant (or close) government is immediate.

One concern could be that ethnic groups becoming more distant to the government were already following different trends in conflict involvement prior to a leadership change, when compared to ethnic groups becoming culturally closer. If this was the case, my estimates could be capturing pre-existing differences in trends between these two groups of ethnicities, and would not be the result of a change in cultural distance. While the visual inspection of the temporal variation in conflict in Figure 2 already suggests that this is not the case, I nevertheless formally test the common trends assumption in a regression framework.

[25]See Table A-11 in the Appendix for the full set of estimates using the lagged measure of cultural distance as main explanatory variable.

In column 4 of Table 4, I add a lead measure of linguistic distance. Should there be any pre-trends in the ethnic group(s) that become more distant to the government compared to those that become less distant after a leadership change, then this lead measure of linguistic distance should be estimated significantly different from zero. The coefficient is insignificant, pointing to absence of anticipatory effects and suggesting that the common trends assumption, crucial for identification, is satisfied. In columns 6-10, I report the same sequence of estimates with conflict over territory as dependent variable and, again, I do not find evidence of anticipatory effects (column 10). In the following subsection, I present a set of robustness tests.

5.2. Robustness Tests

Since the variation driving my results stems from government changes, a concern with my estimation strategy could be that ethnic groups enter or exit power in t as a result of their involvement in conflict (or lack thereof) in $t-1$. Given that ethnic conflict tends to be persistent over time, my results could be driven by a group's involvement in conflict prior to government transitions, and thus have nothing to do with changes in cultural distance. To assess this, I re-run the main specification controlling for a binary variable that is equal to 1 if a potential rebel was involved in a conflict over the previous year. Results in columns 1 to 4 in Table 5 show that lagged conflict is a powerful predictor of current conflict, but that the coefficients on linguistic distance—albeit smaller in magnitude—remain positive and significant in the case of conflict over government power, and negative and insignificant in the case of conflict over territory.[26] This suggests that my estimates are not an outcome of potential rebels' conflict involvement prior to a government change.

A potential explanation for my results could also be related to an ethnicity's representation in the government coalition. When becoming part of a coalition, an ethnicity's cultural distance to the government mechanically decreases. At the same time, this could prevent an ethnic group to rebel for reasons not necessarily related to cultural distance (e.g., the desire not to hurt members of the same group). Conversely, an increase in cultural distance might coincide with an ethnicity losing power, and this could trigger conflict aimed at regaining a position in the government. While I discuss these as potential mechanisms for my results in Section 8., I here perform three initial tests to verify whether this constitutes a threat to the interpretation of my results.

First, in Table 6, I test whether the effect of cultural distance on conflict occurs over and above the effect of being represented in the government. To assess this, I run a horse-race between cultural distance and a binary variable that equals 1 if a potential rebel is represented in the government coalition and 0 otherwise. Second, in Tables A-2 and A-3, I control for whether

[26] One caveat with this specification is that the inclusion of a lagged dependent variable in a fixed effects model might generate Nickell bias (Nickell [1981]). A solution to this problem could be to instrument the first lag of the dependent variable with the second lag. I report results of this exercise in Table A-1 in the Appendix. However, note that this solution requires arguably strong assumptions (i.e., that the second lag be uncorrelated with the first-differenced residuals). In line with Guryan's [2001] arguments, I expect that the estimates in Table 3 constitute an upper bound of the true effects, while those in Table 5 a lower bound. This would be consistent with treatment (i.e., an increase in cultural distance) being *negatively* selected on lagged conflict (see Appendix 1 in Guryan [2001]), meaning that a *decrease* in cultural distance might be the result of successful past conflict over government power.

an ethnicity has lost or gained power in the previous year, respectively. As expected, these variables are associated with conflict over government power: an ethnicity is less likely to rebel when holding a position in the government coalition, and more likely to rebel after having lost power. However, while the magnitude of the cultural distance estimates slightly decreases, the significance of the main results is unaffected by the inclusion of these controls. Third, I construct an alternative measure of cultural distance, which excludes the potential rebel from the distance computation if the latter is part of the government coalition. Results are reported in Table A-4. Again, all previous conclusions remain valid.

Next, I check the robustness of my results to alternative specifications, estimation samples and clustering of standard errors. First, I show that results are similar when excluding ethnic-specific time trends (Table A-7). In Figure A-1, I check for outliers by re-running my estimations dropping one country at a time. I also restrict the analysis to the 29 countries that experienced at least one conflict between 1961 and 2017 (Table A-5) and to the sample of 33 countries that experienced at least one government change (Table A-6). In Tables A-8 and A-10 I re-run the analysis with an alternative level of clustering of standard errors, by country, country-year or by ethnicity, respectively. My results are robust to all these tests.

Finally, I run the main specification on a balanced panel of 207 ethnicities (out of the original 236) that I observe uninterruptedly over the sample period. The original dataset is an unbalanced panel due to some missing country-years (e.g., Sierra Leone during the state collapse between 1993-2001), or due to ethnicities exiting the sample while holding monopoly or dominant power in the government.[27] Estimates on this subsample of ethnicities, reported in Table A-12, are similar to those from the full sample.

6. Triple Difference-in-Differences Approach

So far, my estimation strategy did not allow to account for unobserved time-varying ethnic-specific shocks. If these occurred simultaneously to a change in government and were correlated to an ethnic group's decision to rebel, they could confound my results. To address this, I re-run the analysis on a subset of ethnic groups that were split by country borders during the Scramble for Africa. This strategy, similar to the one employed by Michalopoulos and Papaioannou [2014] and analogous to the one adopted by Dickens [2018], exploits the fact that the same ethnicity is simultaneously exposed to different governments in different countries. As the ethnic identity of governments changes in some countries but not others, the quasi-random allocation of borders provides an exogenous source of within-group variation in cultural distance to the government.[28] As cultural distance of partitioned groups varies over time and between countries, this variation allows me to estimate a triple difference-in-differences specification and thus control for a full set of ethnicity-year fixed effects.[29]

[27] As mentioned in Section 3., these groups cannot be potential rebels as, by definition, they cannot be linked to rebel groups fighting the government.

[28] For additional evidence and details on the randomness in the allocation of colonial boundaries, see the discussion in Michalopoulos and Papaioannou [2016] and in Dickens [2018].

[29] The categories underlying the triple DID are: culturally distant/culturally close; treated (e.g., experiencing a government change)/non-treated; before/after a government change.

While this estimation strategy constitutes a more tightly-controlled empirical exercise, it has the disadvantage of restricting the sample to a peculiar set of ethnicities, i.e., those that were split into multiple countries. As Michalopoulos and Papaioannou [2016] show, partitioned groups are considerably more likely to experience political violence, ethnic wars, and government-led discrimination when compared to non partitioned ethnicities. For this reason, one should interpret the results presented in this section as a robustness exercise in support of the baseline findings.

I merge contemporary ethnicities to the Murdock Atlas, which contains information on ethnic settlements prior to European contact and therefore provides the most reliable source to identify groups that were partitioned during the Scramble for Africa.[30] The map in Figure 3 displays the partitioned ethnicities that I successfully merged with the EPR ethnic groups. This restricted sample includes 84 distinct Murdock ethnic groups partitioned across 31 African countries.[31] Table B-3 reports summary statistics. Compared to the full African sample, the prevalence of conflict over government power is slightly higher, which is consistent with Michalopoulos and Papaioannou's [2016] findings showing that partitioned groups are more likely to experience conflict compared to non-partitioned ones. On average, linguistic distance to the government is also higher in this restricted sample.

I estimate the following empirical model:

$$\text{Conflict}_{rct} = \eta_{c,r} + \lambda_{c,t} + \zeta_{r,t} + \beta \text{LD}_{rct} + \Gamma \text{G}_{rct} + \Omega \text{C}_{rct} + \epsilon_{rct} \tag{5}$$

The dependent variable is a measure of conflict for ethnic group r in country c and year t. The main independent variable, LD_{rct} is a measure of linguistic distance to the government in period t.

As in my previous specifications, I add a full set of country-year fixed effects ($\lambda_{c,t}$), as well as a set of ethnicity-country fixed effects ($\eta_{c,r}$). Moreover, since the same potential rebel is present in multiple countries, this specification controls for a full set of ethnicity-year fixed effects ($\zeta_{r,t}$), and thus accounts for any time-variant ethnic-specific shocks. G_{rct} and C_{rct} indicate the same set of geographic and climatic controls outlined in section 4.

6.1. Results

Table 7 reports the results. Columns 1 to 4 show estimates for conflict over government power, using both the weighted and the unweighted measure of cultural distance. In column 2, a one standard deviation increase in cultural distance to the government leads to a 8.7 percentage points (or 0.39 standard deviations) increase in disputes over power. The effect is large when compared to the mean prevalence of 0.054 in the sample considered, and remains quantitatively similar to the effect uncovered using the whole African sample.

When turning to conflict over territory in columns 5 to 8, the association reverses. The

[30] This is also the data source employed by Michalopoulos and Papaioannou [2016], who were the first exploiting partitioned groups for identification.

[31] Note that my merging procedure is successful for 41 percent of the ethnicities that Michalopoulos and Papaioannou [2016] classify as partitioned.

estimate in column 6 indicates that a one standard deviation increase in cultural distance to the government decreases conflict over territory by 5.3 percentage points (or 0.34 standard deviations), a sizeable effect that compares to a 0.015 conflict prevalence in the sample.

As in the estimating equation in the whole African sample, identification rests on the assumption that governments are not systematically elected because of the conflict behavior of ethnic groups prior to an election. In Table A-13, I find that the estimates remain similar when controlling for a lagged dependent variable. In Table A-14, I show that my main coefficients are robust to the inclusion of a binary variable indicating whether an ethnic groups is part of the government coalition or not as a control. Finally, in Table A-15, I report estimates for leads and lags of cultural distance and, similar to the results for the whole African sample, I do not find evidence of anticipatory effects.

7. Instrumental Variables Approach

A remaining concern with my strategy is that the measure of cultural distance might be endogenous to conflict. If there are omitted factors that affect linguistic distance between groups and contemporary conflict, then the estimates presented in the previous section are biased. An example is historical conflict. Old antagonisms between groups might have affected linguistic distance and, at the same time, can have a direct effect on conflict today.[32]

This is particularly a concern in my setting, where not all members of an ethnic group speak the same language. Recall that the linguistic distance measure is a weighted average of all pairwise linguistic distances, where weights reflect the percentage of group members speaking each language. Therefore, any source of endogeneity could operate through two channels: there could be omitted factors affecting both conflict and (i) how the linguistic tree of two languages evolved, or (ii) the relative importance of a language within an ethnicity's population.

The direction of the potential bias is ex-ante unclear. Suppose past conflict between two ethnicities is positively associated with current conflict. A history of violence might also have reduced two ethnicities' linguistic distance over time, e.g., due to genocides eliminating linguistically distant subgroups. If this was the case, then the main coefficients on conflict over government power would be biased downwards. On the other hand, if past conflict widened linguistic distance by reinforcing linguistic enclaves or ethnic boundaries, then my coefficients would be biased upwards.

7.1. The Instrument

To instrument for cultural distance between groups, I exploit the Bantu expansion, a natural experiment unique to African history. 5,000 years ago, a climatic shock generated a temporary loss of rainforest in Central Africa. Through increased seasonality of the monsoon, a lowering of the sea surface temperature in the Guinean Gulf, and less rainfall, the shock favored the emergence of savanna corridors (Bostoen *et al.* [2015]). After the climatic crisis ended, the new

[32]For evidence on how past conflict is correlated with current conflict in Africa, see Besley and Reynal-Querol [2014].

savanna environments disappeared, supplanted by rainforest.

Archaeological, anthropological, and linguistic evidence suggests that the temporary opening up of parts of the—beforehand impenetrable—rainforest facilitated the expansion of the Bantu, a tribe residing in Cameroon, throughout sub-Saharan Africa (see Bostoen *et al.* [2013] and Bostoen *et al.* [2015] for a summary of the interdisciplinary literature). The exact reason for why Bantu people started migrating is largely debated. One hypothesis is that the climate-driven opening of the forest gave hunters access to "naive" animal populations that were previously trapped in the forest and then became suddenly available (Bostoen *et al.* [2013]). However, there is by now consensus among scholars on how the *path* of the Bantu migration was exogenously shaped by climatic events, a crucial feature for my instrumental variables analysis (Grollemund *et al.* [2015], Bostoen *et al.* [2015]).

How is this massive prehistoric migration related to contemporary cultural distances between ethnic groups? The Bantu expansion has been defined as the most important linguistic, cultural and demographic process in Late Holocene Africa, and one of the most influential cultural events of its kind (Robbeets and Savelyev [2017]). Bantu farmers, in gradually expanding from their homeland in Northwestern region of Cameroon, spread their new language and culture throughout sub-Saharan Africa, assimilating or displacing earlier inhabitants of the regions they crossed, i.e., Pygmy and Khoisan hunter-gatherers (Diamond [1997]). Today, one out of three Africans is fluent in at least one of the approximately 500 existing languages belonging to the Bantu family.

Crucially, Bantu people did not move and settle everywhere in Sub-Saharan Africa, and this is exactly the idea behind my instrument. The Bantu expansion acted as a prehistoric "cultural shock"—in turn exogenously triggered by a climatic crisis—which affected the cultural capital of some areas but not others. Within today's countries, some territories—which then became ethnic homelands—happened to be crossed by the route of the Bantu expansion, while others did not. This is demonstrated, for instance, by the existence of pre-Bantu hunter-gatherers populations such as the Pygmies in Central Africa, the Hadza in Tanzania, and Khoisan people in southern Africa. The route of the Bantu expansion did not cross their ethnic homelands, and for this reason, these groups remain today culturally and linguistically distinct.

I exploit this unique event in African prehistory to construct a novel instrument for cultural distance between ethnic groups. Groups whose homelands were highly exposed to the Bantu migration route have inherited Bantu culture, and therefore should be culturally distant to those that remained unaffected. Instead, groups with a similarly high or low exposure to the Bantu expansion should be culturally close to each other, because they either both inherited Bantu culture, or kept their pre-existing one.[33] Based on this idea, I use the absolute difference in two groups' exposure to the Bantu expansion as an instrument for cultural distance.

To obtain the instrument, I start by constructing a measure of exposure to the Bantu expansion for each ethnic group's homeland. I merge contemporary ethnicities to their ancestral homelands

[33]The assumption here is that, within contemporary countries, pre-Bantu groups were on average culturally similar to each other, and culturally dissimilar to the Bantu. This is partly supported by archaeological evidence showing that pre-Bantu societies shared the same mode of subsistence, i.e., hunting and gathering, while Bantu people were predominantly farmers (Diamond [1997]).

in the Murdock map, the most ancient record of ethnic settlements in Africa. I then overlay ethnic groups' ancestral homelands on the historical migration route of the Bantu reconstructed by Grollemund *et al.* [2015].[34] The left panel of Figure 4 illustrates the route of the Bantu migration and contemporaneous country borders. For each ethnic group in the Murdock map, I construct a Bantu Index capturing the intensity of exposure to the Bantu migration route. I calculate the length of the path crossing each ethnic homeland, divide it by the size of the settlement, and then normalize this index to range between 0 and 1:

$$\text{Bantu Index}_{EA} = \frac{\text{Path Length}}{\text{Homeland Area}} \tag{6}$$

Figure 4 (right) shows the distribution of the Bantu index across Africa. There is considerable within-country, cross-ethnicity variation in ethnic groups' exposure to the Bantu expansion. Since each EPR group is associated to multiple groups in the Ethnographic Atlas, I proceed as follows to obtain an index at the EPR-group level:

$$\text{Bantu Index}_{EPR} = \sum_{i=1}^{N} \alpha_i \text{Bantu Index}_{EA_i} \tag{7}$$

where α_i is a weight reflecting the accuracy of the match between each Murdock group and the EPR group.[35]

Next, I construct the instrument, i.e., the absolute distance in the Bantu Index between a potential rebel and the government. To this end, I follow the same procedure I adopted for the linguistic distance measure. First, I compute the Bantu distance between a potential rebel (r) and each ethnic group forming the government (g_i):

$$\text{BD}_{rg_i} = |\text{Bantu Index}_r - \text{Bantu Index}_{g_i}| \tag{8}$$

Second, I compute a weighted Bantu distance between each potential rebel and the government:

$$BD^W = \sum_{i=1}^{N} p_{g_i} \times BD_{rg_i} \tag{9}$$

where N denotes the total number of ethnicities forming the government and p_{g_i} is a weight reflecting the position of power of group g_i in the coalition. Alternatively, I construct the unweighted version:

$$BD^{UW} = \sum_{i=1}^{N} \frac{BD_{rg_i}}{N} \tag{10}$$

[34] See Appendix B for details on how Grollemund *et al.* [2015] reconstructed the route.

[35] The accuracy is given by the percentage of nodes in the EPR-group linguistic tree covered by the Murdock-group linguistic tree. $\sum_{i=1}^{N} \alpha_i = 1$. If the match is equally accurate between each Murdock group, then $\text{Bantu Index}_{EPR} = \sum_{i=1}^{N} \frac{\text{Bantu Index}_{EA_i}}{N}$. See Appendix B-2 for details on the LEDA package, the merging procedure, and the accuracy of the match.

7.2. Empirical Strategy

I use a two-stage least-square (2SLS) procedure to estimate equation 4. In the first stage, I estimate the effect of the distance in the exposure to the Bantu expansion route on linguistic distance between ethnic groups today:

$$\mathrm{LD}_{rct} = \lambda_{c,t} + \zeta_r + \theta_r t + \beta \mathrm{BD}_{rct} + \Gamma \mathrm{G}_{rct} + \Omega \mathrm{C}_{rct} + u_{rct} \tag{11}$$

I run the IV analysis for the subset of 17 sub-Saharan countries where the Bantu tribe migrated, and that are classified as Bantu countries according to the Guthrie classification of languages (see Figure 5).[36]

Table B-4 reports summary statistics for this subsample of 86 ethnic groups (N=3,628). The identifying variation stems from 41 government changes. In the Bantu region, the prevalence of conflict is slightly lower than in the continent as a whole (0.060) and only 13 percent of it is explained by conflicts fought for territory. Linguistic distance between groups is also, on average, smaller (0.257).

7.3. Results

Table 8 reports results from the instrumental variables analysis, which confirm the linear probability model estimates. In Panel A, columns 1 and 2 and columns 5 and 6 reports OLS estimates for conflict over government and conflict over territory, respectively, for the sub-sample of ethnicities in the Bantu region. Results are consistent with the estimates in the full African continent. The association between cultural distance and conflict over government is positive and significant, while the coefficients on conflict over territory are negative (and, only in one case, significant).

Panel C reports the first-stage results of equation 11 and shows how a larger difference in the exposure to the Bantu expansion is positively associated with cultural distance, measured through linguistic distance. The Kleibergen-Paap F-statistic indicates that the instrument is a powerful predictor of cultural distance. In Panel B (columns 3 and 4), I find that there is a positive reduced-form effect of Bantu distance on conflict over government power, which is reassuring for the validity of the instrument. A one standard deviation increase in Bantu distance increases conflict prevalence by 6.5-7 percentage points. Columns 3 and 4 in Panel A report the 2SLS result, which are consistent with the OLS estimates of columns 1 and 2. In all specifications, both for the weighted and unweighted measures of cultural distance, a one standard deviation increase in linguistic distance significantly increases the prevalence of conflict over power. The magnitude of the coefficients more than doubles when compared to the OLS specification. The original coefficients were biased downwards, which is consistent, for instance, with past conflict between two ethnicities being positively associated with contemporary conflict, and, at the same time, with past conflict reducing two groups' linguistic distance through genocides eliminating

[36]The countries covered by Bantu languages according to the Guthrie classification are: Angola, Botswana, Burundi, Cameroon, Congo, Congo DRC, Equatorial Guinea, Gabon, Kenya, Malawi, Mozambique, Rwanda, South Africa, Tanzania, Uganda, Zambia, Zimbabwe.

linguistically distant subgroups.

When turning to conflict over territory, columns 7 and 8 in Panel B document a negative reduced-form effect of Bantu distance on conflict over territory. In Panel A, columns 7 and 8 report the 2SLS results. There is a negative effect of cultural distance on conflict over territory, but only for the unweighted version of the linguistic distance measure not accounting for the relative importance of ethnic groups in the government (column 8). In this specification, a one standard deviation increase in cultural distance to the government decreases conflict by one percentage point, which is a large effect when considering the mean conflict prevalence in the sample (0.008).[37]

These results are robust to controlling for lagged conflict (Table A-17) and to controlling for whether the potential rebel is present in the government coalition or not (Table A-18).

7.4. Instrument Validity

The validity of the IV estimates rests on the assumption that the differential exposure to the Bantu expansion affects conflict today only through its impact on cultural distance, conditional on the controls included in the regression. The first concern is that geographical differences between ancestral homelands might have determined the path of the Bantu migration and, at the same time, might also affect the likelihood of conflict between groups today. First, Grollemund *et al.* [2015] emphasize how the route of the expansion was mostly determined by emerging savannah corridors, which appeared in the rainforest due to a climatic shock and then disappeared once the climatic crisis was over. Despite this suggestive evidence of exogeneity in the way the path evolved, my preferred specification addresses this concern by controlling for a set of geographic and climatic controls.

Another concern could be that the Bantu, in gradually expanding, intentionally and systematically avoided pre-existing tribes with certain characteristics that make them more (or less) prone to conflict today. Suppose the Bantu avoided particularly bellicose tribes, and favored territories inhabited by more peaceful populations. Then my instrument would be systematically picking up differences in bellicosity. If these differences have a direct effect on conflict today, the exclusion restriction is violated.

I examine whether this example is at play in my setting by delving deeper in the reduced-form analysis. If the Bantu systematically avoided certain territories along unobservable characteristics correlated with contemporary conflict, then only being more (or less) affected to the Bantu expansion should have a reduced-form effect on conflict, but not both. I test this by splitting the absolute distance in the Bantu index into two: being *more* exposed to the Bantu expansion than the ethnicities at the government, and being *less* exposed.

I run an alternative reduced-form estimation using these two measures, instead of the absolute difference as in my main specification. As shown in Table A-19, I do not find evidence of a

[37] A recent contribution by Lee *et al.* [2020] proposes a new test to assess the validity of the second-stage t-ratio inference when the first-stage F statistic is smaller than 104.7, as in this case. Table A-16 in the Appendix reports second-stage results with standard errors based on Lee *et al.*'s [2020] *tF* procedure and shows that inference considerations remain unchanged.

different reduced-form effect of being more versus less exposed to the Bantu expansion. Both coefficients hold the same sign, and the F statistics indicates that these coefficients are not different from each other. This suggests that what matters for conflict between ethnic groups is their *absolute* distance in the Bantu index—as conjectured in section 7.1.—and not the fact of being more (or less) affected by the Bantu migration compared to the other group.

8. Channels

Why are ethnic groups more likely to fight over power—but not over territory—when the government is culturally distant from them? In this section, I analyze potential mechanisms and explore pathways that have been advanced by existing theoretical work on relatedness and conflict.

My finding that cultural distance matters only for conflict fought over the control of the central government is consistent with Spolaore and Wacziarg's [2017] and Esteban and Ray's [2011] theoretical frameworks. Deep-rooted cultural differences shape diverging preferences over public policies, which trigger conflict over government power, but not over territory. The notion that cultural distance between ethnic groups generates diverging preferences over public policies has been often assumed by the literature on diversity and conflict, but has not been directly tested empirically.

To provide direct evidence of this mechanism, I exploit individual-level data from the Afrobarometer survey. The sample comprises individuals belonging to 95 different ethnic groups in 27 African countries surveyed between 1999 and 2017. Respondents report their opinions regarding the performance of their current government in matters related to the following broad policy domains: the economy, education, health, infrastructure, minorities, national security, and other social issues. Using PCA, I take the first component and group opinions about government performance into a single index ranging between 0-1, with 1 denoting the highest degree of disagreement. Table A-20 reports the results of the PCA for each survey round. The first principal component explains 34-61% of the common variance of the domains. Each domain always loads positively on the first principal component, suggesting that individuals who dislike government performance tend to do so along the whole range of listed public policies.

To establish whether cultural distance to the government comes with higher disagreement over public policies, I run the following specification:

$$\text{Opinion government}_{irct} = \lambda_{c,t} + \zeta_r + \theta_r t + \beta \text{LD}_{rct} + \Gamma \text{G}_{rct} + \Omega \text{C}_{rct} + \Phi \text{I}_i + \epsilon_{irct} \quad (12)$$

where $\text{Opinion government}_{irct}$ is a continuous variable that ranges between 0 and 1 and captures the extent to which individual i belonging to ethnic group r thinks that the government of country c in year t is performing fairly badly or very badly. LD_{rct} indicates linguistic distance between individual i's ethnic group r and government g in country c and year t. In addition to the set of fixed effects and ethnicity-level controls included in the baseline specification (see equation 4), I also add a set of individual-level controls I_i (a dummy for female, age, age squared,

and a dummy for urban residents) and a survey round fixed effect.[38]

Results in Table 9 show that individuals are more likely to disagree with a wide range of public policies implemented by the government when they are more linguistically distant to the ethnic groups in power. One standard deviation increase in an individual's linguistic distance to the government increases the degree of opposition to government policies by 13-20 percentage points, which is a large effect, considering an average degree of disagreement of 56 percent in the sample (a 23-36 percent increase). Equivalent results arise when constructing the dependent variable taking the mean disagreement across all policy domains.

Another potential explanation for my results is related to Caselli and Coleman's [2013] theory of ethnic conflict. In their framework, language is an ethnic marker that allows groups in power to exclude non-members from enjoying a country's resources. For example, linguistic distance facilitates marking non-group members and discriminating against them in the moment they apply for goods distributed through the central government, such as scarce places in higher education or jobs in the public sector. Ethnic distance should come with more conflict, because it is easier for a dominant group to mark and discriminate against ethnically distant groups.

According to this interpretation, it could be that discrimination—facilitated by linguistic distance—is what triggers conflict over power, and not a cultural mechanism. To shed light on this alternative explanation, I explore whether the effect of linguistic distance on conflict over power is larger among discriminated or powerless groups. To this end, I run the following:

$$\text{Conflict}_{rct} = \lambda_{c,t} + \zeta_r + \theta_r t + \beta \text{LD}_{rct} + \kappa \text{S}_{rt} + \rho \text{S}_{rt} \times \text{LD}_{rct} + \Gamma \text{G}_{rct} + \Omega \text{C}_{rct} + \epsilon_{rct} \quad (13)$$

where S_{rt} indicates one of the three following binary variables: (i) equal to 1 if a potential rebel r is in power in year t and 0 otherwise; (ii) equal to 1 if a potential rebel r is powerless in year t and 0 otherwise; (iii) equal to 1 if a potential rebel r is discriminated in year t and 0 otherwise. κ captures the direct effect of a potential rebel's status on its conflict behavior at time t, and ρ—the coefficient on the interaction term—measures to what extent the effect of linguistic distance on conflict over government power interacts with an ethnic group's status (in power, powerless or discriminated) at time t.

Table 10 reports estimates of equation 13. Columns 1 to 3 display coefficients when S_{rt} is a binary variable that equals 1 if a potential rebel is in power (i.e., part of the government coalition), and zero otherwise. As already shown in the robustness tests in section 5.2., and as expected, being in power has a direct negative effect on conflict. However, the effect of linguistic distance on conflict does not seem to interact with an ethnic group's position of power in the government coalition: the coefficient on the interaction term ρ is small and insignificant, and the coefficient β remains positive and significant. Similarly, being powerless does not have a direct effect on conflict (see columns 4 to 6), nor does the interaction term with linguistic distance.

Finally, in columns 7 to 9 I directly test whether the effect of linguistic distance on conflict is larger among discriminated groups, consistently with the explanation outlined above. As expected, being discriminated has a strong positive direct effect on the likelihood that a potential

[38]Note that each Afrobarometer round is conducted during different years in different countries.

rebel is fighting the government over power. However, the coefficient on linguistic distance remains large and significant, and the interaction term is insignificant. While I acknowledge the limitation of this empirical exercise due to the endogeneity of the discrimination measure, I take these results as additional suggestive evidence in favor of the cultural mechanism.

9. Conclusion

Conflict—and in particular, civil conflict fought along ethnic lines—is more prevalent in ethnolinguistically diverse societies. Yet, despite facing the same aggregate level of ethnolinguistic diversity, some ethnic groups rebel and others do not. This paper sheds light on why this is the case by studying the relationship between an ethnicity's cultural distance to the central government and its propensity to rebel against it. Consistently across estimation strategies, I find that a group's involvement in conflict increases with its cultural distance to the central government. However, this is true only for a specific type of conflict, i.e., one that is fought over the control of the central government. Ethnic rebellions over territory do not respond to changes in a group's distance to the government. If anything, larger distance comes with less conflict over territory and resources.

My paper relies on two key innovations. First, it draws a distinction between conflicts fought over government power and conflicts fought over territory. Second, it moves the analysis from the country level to the ethnicity level. Previous empirical analyses on diversity at the country level did not consider the ethnic identity of the combatants or the nature of the bone of contention, and therefore could not directly test the countervailing effects of diversity on conflict. My contribution highlights that the issue over which combatants fight is a crucial dimension to consider when studying the determinants of conflict.

The analysis at the ethnicity level allows me to delve into the mechanisms through which cultural distance generates conflict over government power. I document that the reaction to cultural distance occurs over and above the effect of exclusion from power or discrimination, suggesting that another mechanism is at play. I show that culturally distant groups are more likely to disagree over the mix of public policies, and are thus more likely to rebel to wrest control of the central government.

This paper focuses on ethnic civil conflict and, as a consequence, on a group's cultural distance to the ethnic groups holding government power. Civil conflict, while largely prevalent, is only one type of the many possible manifestations of inter-group violence. A promising avenue for future research is to test the hypothesis of this paper in the context of non-civil conflict, i.e., inter-group violence not involving the government, so far largely overlooked by the literature.[39] Another promising area for future research is improving the measurement of cultural distance. Linguistic distance is by now a well-established measure, but it does not allow to test which exact dimension of culture triggers conflict. Efforts to shed light on this issue are underway,[40] but exploring new metrics that allow to separately examine different cultural components is a

[39] A notable exception is the work by Depetris-Chauvin and Özak [2020].

[40] See, for example, Guarnieri and Tur-Prats [2020], who focus on distance gender norms and investigate its impact on the intensive margin of violence.

promising area for future lines of inquiry.

Taken together, my findings contribute to the literature on diversity and conflict by showing that, contrary to the most pessimistic primordialist views, cultural distance does not always and inevitably come with more conflict. Cultural distance triggers one specific type of dispute—i.e., the one over government power—, and it does so through a specific mechanism—i.e., by generating diverging preferences. As such, my analysis speaks against the widespread idea that ethnic conflict is the sole result of deep-rooted and insurmountable ethnic hatred between groups, and rather suggests that a mix of public policies accommodating diverging preferences might be powerful in alleviating conflict in multicultural societies.

References

Afrobarometer Data (1999-2017).

Alesina, A. and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of economic literature*, **43** (3), 762–800.

Allansson, M., Melander, E. and Themnér, L. (2017). Organized violence, 1989–2016. *Journal of Peace Research*, **54** (4), 574–587.

Arbatli, C. E., Ashraf, Q. H., Galor, O. and Klemp, M. (2020). Diversity and conflict. *Econometrica*, **88** (2), 727–797.

Ashraf, Q. and Galor, O. (2013). Genetic diversity and the origins of cultural fragmentation. *American Economic Review*, **103** (3), 528–33.

Besley, T. and Reynal-Querol, M. (2014). The legacy of historical conflict: Evidence from africa. *American Political Science Review*, pp. 319–336.

Blattman, C. and Miguel, E. (2010). Civil war. *Journal of Economic Literature*, **48** (1), 3–57.

Bostoen, K., Clist, B., Doumenge, C., Grollemund, R., Hombert, J.-M., Muluwa, J. K., Maley, J., Blench, R., Di Carlo, P., Good, J. *et al.* (2015). Middle to late holocene paleoclimatic change and the early bantu expansion in the rain forests of western central africa. *Current Anthropology*, **56** (3), 367–368.

—, Grollemund, R. and Muluwa, J. K. (2013). Climate-induced vegetation dynamics and the bantu expansion: Evidence from bantu names for pioneer trees (elaeis guineensis, canarium schweinfurthii, and musanga cecropioides). *Comptes Rendus Geoscience*, **345** (7-8), 336–349.

Campante, F. R., Do, Q.-A. and Guimaraes, B. (2019). Capital cities, conflict, and misgovernance. *American Economic Journal: Applied Economics*, **11** (3), 298–337.

Caselli, F. and Coleman, W. J. (2013). On the theory of ethnic conflict. *Journal of the European Economic Association*, **11** (suppl_1), 161–192.

Depetris-Chauvin, E. and Özak, Ö. (2020). Borderline disorder:(de facto) historical ethnic borders and contemporary conflict in africa.

Desmet, K., Ortuño-Ortín, I. and Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of development Economics*, **97** (2), 322–338.

—, Weber, S. and Ortuño-Ortín, I. (2009). Linguistic diversity and redistribution. *Journal of the European Economic Association*, **7** (6), 1291–1318.

Diamond, J. (1997). Guns, germs, and steel: The fates of human societies.

Dickens, A. (2018). Ethnolinguistic favoritism in african politics. *American Economic Journal: Applied Economics*, **10** (3), 370–402.

— (2020). Understanding ethnic differences: The roles of geography and trade.

Esteban, J., Mayoral, L. and Ray, D. (2012). Ethnicity and conflict: An empirical study. *American Economic Review*, **102** (4), 1310–42.

— and Ray, D. (2011). Linking conflict to inequality and polarization. *American Economic Review*, **101** (4), 1345–74.

Fearon, J. D. and Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American political science review*, **97** (1), 75–90.

Fick, S. E. and Hijmans, R. J. (2017). Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, **37** (12), 4302–4315.

Galor, O. and Özak, Ö. (2016). The agricultural origins of time preference. *American Economic Review*, **106** (10), 3064–3103.

—, Özak, Ö. and Sarid, A. (2018). Geographical roots of the coevolution of cultural and linguistic traits. *Available at SSRN 3284239*.

Gomes, J. F. (2020). The health costs of ethnic distance: evidence from sub-saharan africa. *Journal of Economic Growth*, **25** (2), 195–226.

Goodman-Bacon, A. (2018). *Difference-in-differences with variation in treatment timing*. Tech. rep., National Bureau of Economic Research.

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C. and Pagel, M. (2015). Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences*, **112** (43), 13296–13301.

Guariso, A. and Rogall, T. (2017). Rainfall inequality, political power, and ethnic conflict in africa.

Guarnieri, E. and Tur-Prats, A. (2020). Cultural distance and conflict-related sexual violence.

Guryan, J. (2001). Desegregation and black dropout rates. *National Bureau of Economic Research*.

Harari, M. and La Ferrara, E. (2018). Conflict, climate, and cells: a disaggregated analysis. *Review of Economics and Statistics*, **100** (4), 594–608.

Harbom, L., Melander, E. and Wallensteen, P. (2008). Dyadic dimensions of armed conflict, 1946—2007. *Journal of peace research*, **45** (5), 697–710.

Horowitz, D. L. (2000). *Ethnic groups in conflict, updated edition with a new preface*. Univ of California Press.

Lee, D. L., McCrary, J., Moreira, M. J. and Porter, J. (2020). Valid t-ratio inference for iv. *arXiv preprint arXiv:2010.05058*.

Lewis, M. P., Simons, G. F. and Fennig, C. D. (2020). Ethnologue: Languages of the World. *Dallas, Texas: SIL International. Online version: http://www.ethnologue.com*.

McGuirk, E. and Burke, M. (2020). The economic origins of conflict in africa. *Journal of Political Economy*, **128** (10), 3940–3997.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *American Economic Review*, **102** (4), 1508–39.

— and Papaioannou, E. (2014). National institutions and subnational development in africa. *The Quarterly journal of economics*, **129** (1), 151–213.

— and — (2016). The long-run effects of the scramble for africa. *American Economic Review*, **106** (7), 1802–48.

Mokyr, J. (2016). *A culture of growth: the origins of the modern economy*. Princeton University Press.

Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American economic review*, **95** (3), 796–816.

— and — (2017). Ethnic diversity and growth: Revisiting the evidence. *Review of Economics and Statistics*, pp. 1–43.

Moscona, J., Nunn, N. and Robinson, J. A. (2020). Segmentary lineage organization and conflict in sub-saharan africa. *Econometrica*, **88** (5), 1999–2036.

Müller-Crepon, C., Pengl, Y. and Bormann, N.-C. (2020). Linking Ethnic Data from Africa.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society*, pp. 1417–1426.

Nunn, N. and Wantchekon, L. (2011). The slave trade and the origins of mistrust in africa. *American Economic Review*, **101** (7), 3221–52.

Pfeifer, G. and Reutter, M. (2020). The morning after: Prescription-free access to emergency contraceptive pills.

Robbeets, M. and Savelyev, A. (2017). *Language Dispersal Beyond Farming*. John Benjamins Publishing Company.

Smith, M. G. (1965). *The plural society in the British West Indies*. Univ of California Press.

Spolaore, E. and Wacziarg, R. (2016a). Ancestry, language and culture. In *The Palgrave handbook of economics and language*, Springer, pp. 174–211.

— and — (2016b). War and relatedness. *Review of Economics and Statistics*, **98** (5), 925–939.

— and — (2017). *The political economy of heterogeneity and conflict*. Tech. rep., National Bureau of Economic Research.

Sundberg, R. and Melander, E. (2013). Introducing the ucdp georeferenced event dataset. *Journal of Peace Research*, **50** (4), 523–532.

Vogt, M., Bormann, N.-C., Rüegger, S., Cederman, L.-E., Hunziker, P. and Girardin, L. (2015). Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution*, **59** (7), 1327–1342.

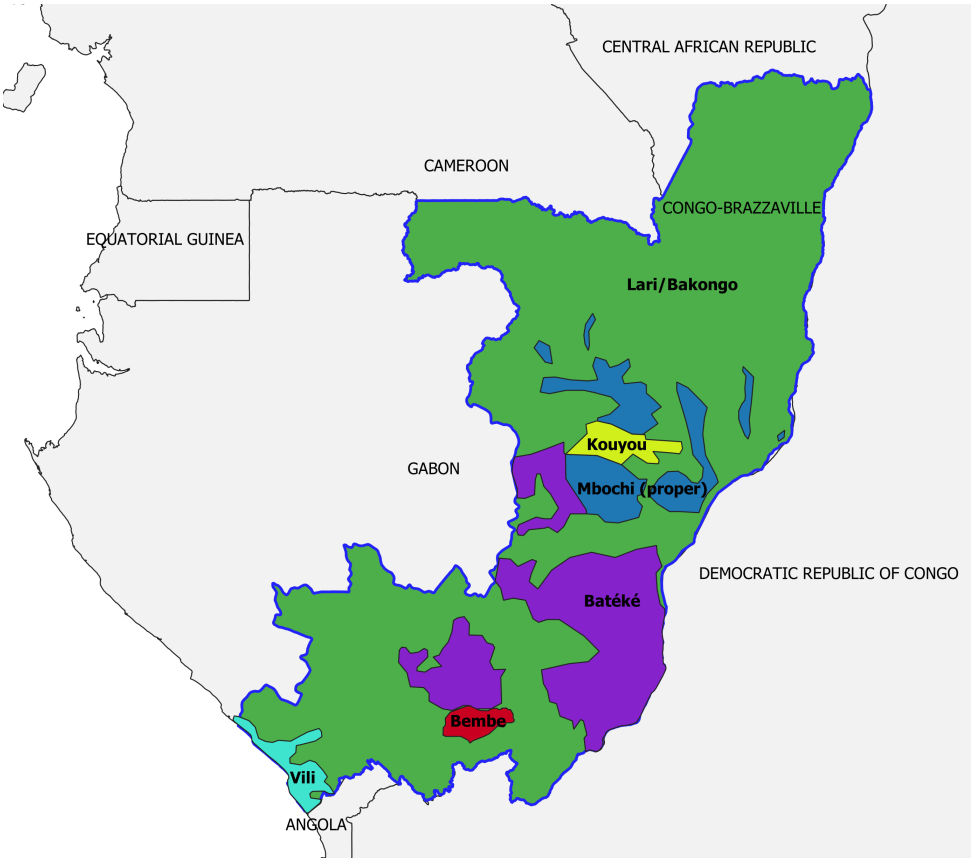
Wimmer, A., Cederman, L.-E. and Min, B. (2009). Ethnic politics and armed conflict: A configurational analysis of a new global data set. *American Sociological Review*, **74** (2), 316–337.

Wucherpfennig, J., Metternich, N. W., Cederman, L.-E. and Gleditsch, K. S. (2012). Ethnicity, the State, and the Duration of Civil War. *World Politics*, **64** (1), 79–115.

—, Weidmann, N. B., Girardin, L., Cederman, L.-E. and Wimmer, A. (2011). Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset. *Conflict Management and Peace Science*, **28** (5), 423–437.

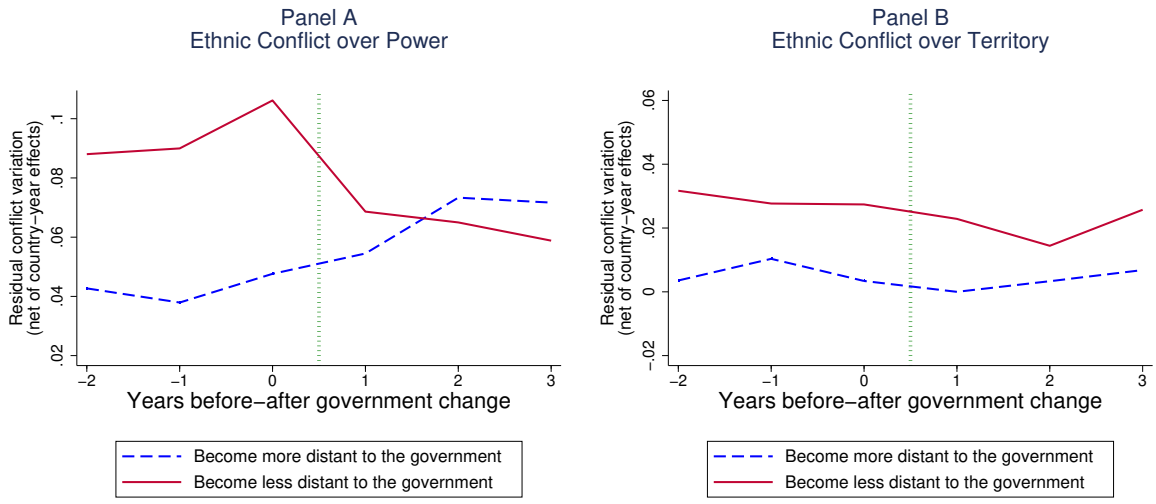
Figures and Tables

FIGURE 1: *Ethnic Groups and Ethnic Settlements in Congo*



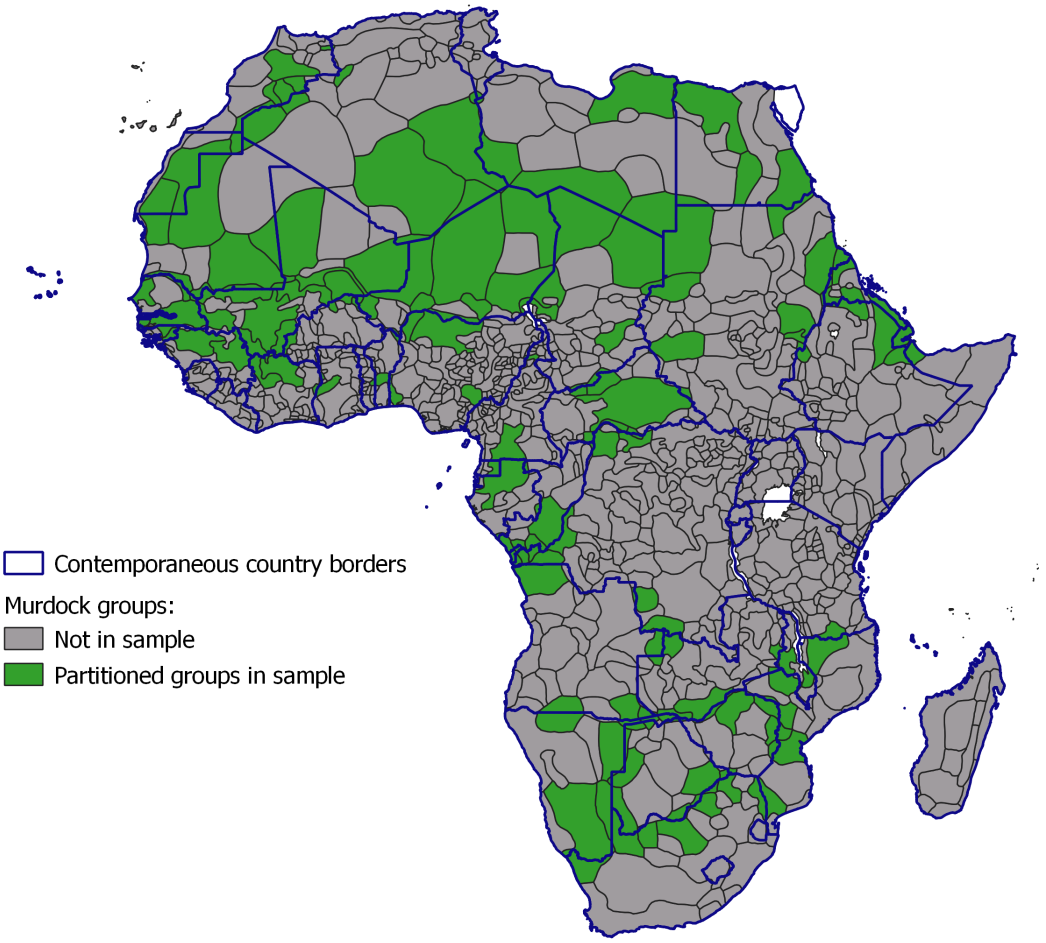
Notes: The figure illustrates the six politically relevant ethnic groups (Lari/Bakongo, Kouyou, Mbochi (proper). Bateke, Bemba, Vili) in Congo-Brazzaville and their settlements. Source: Ethnic Power Relations (EPR) Core (Wucherpfennig *et al.* [2012]) and GEO-EPR (Wucherpfennig *et al.* [2011]).

FIGURE 2: *Ethnic Conflict over Power and over Territory Before and After a Government Change*



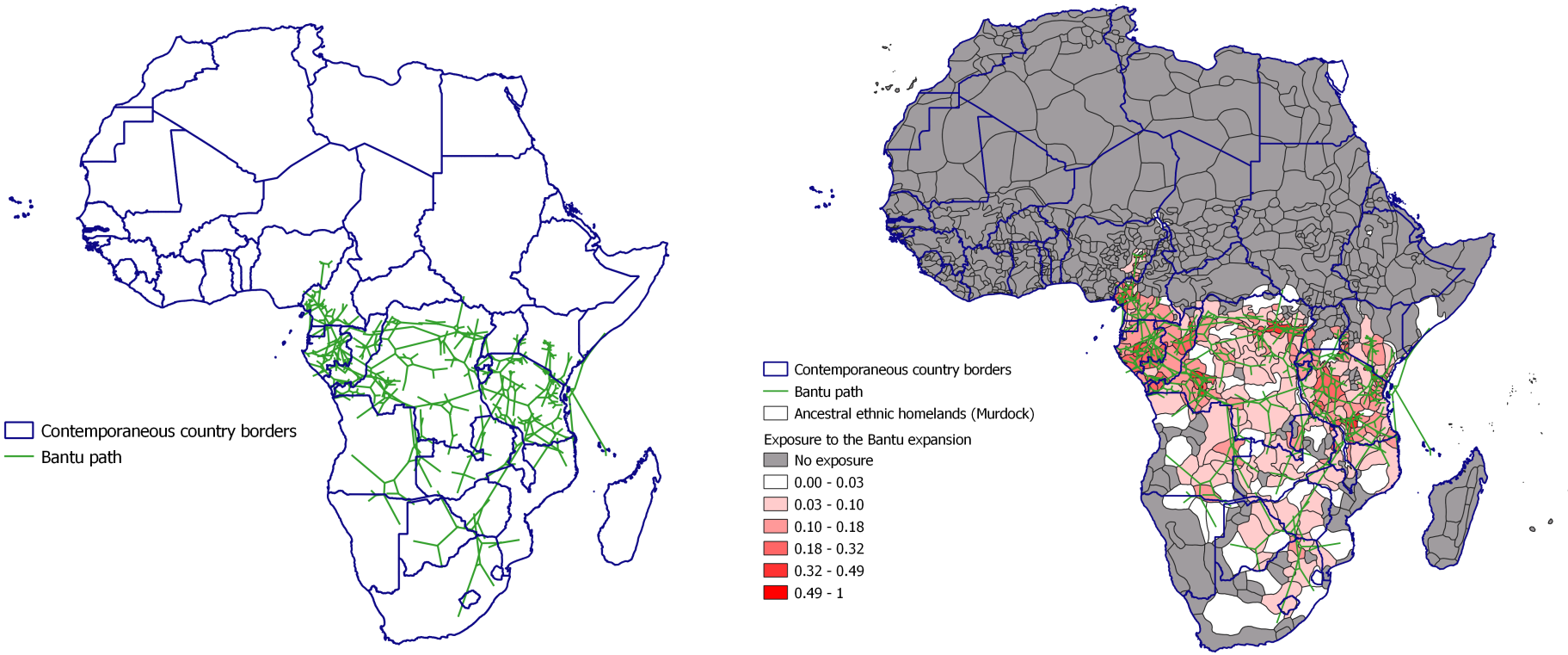
Notes: The figure plots the average residual variation in conflict over power (left) and conflict over territory (right) net of country-year fixed effects for two subsamples: ethnic groups that become culturally more distant to the government following a government change (blue dashed), and ethnic groups that become culturally less distant to the government following a government change (red solid).

FIGURE 3: *Murdock Ethnic Groups Split by Country Borders*



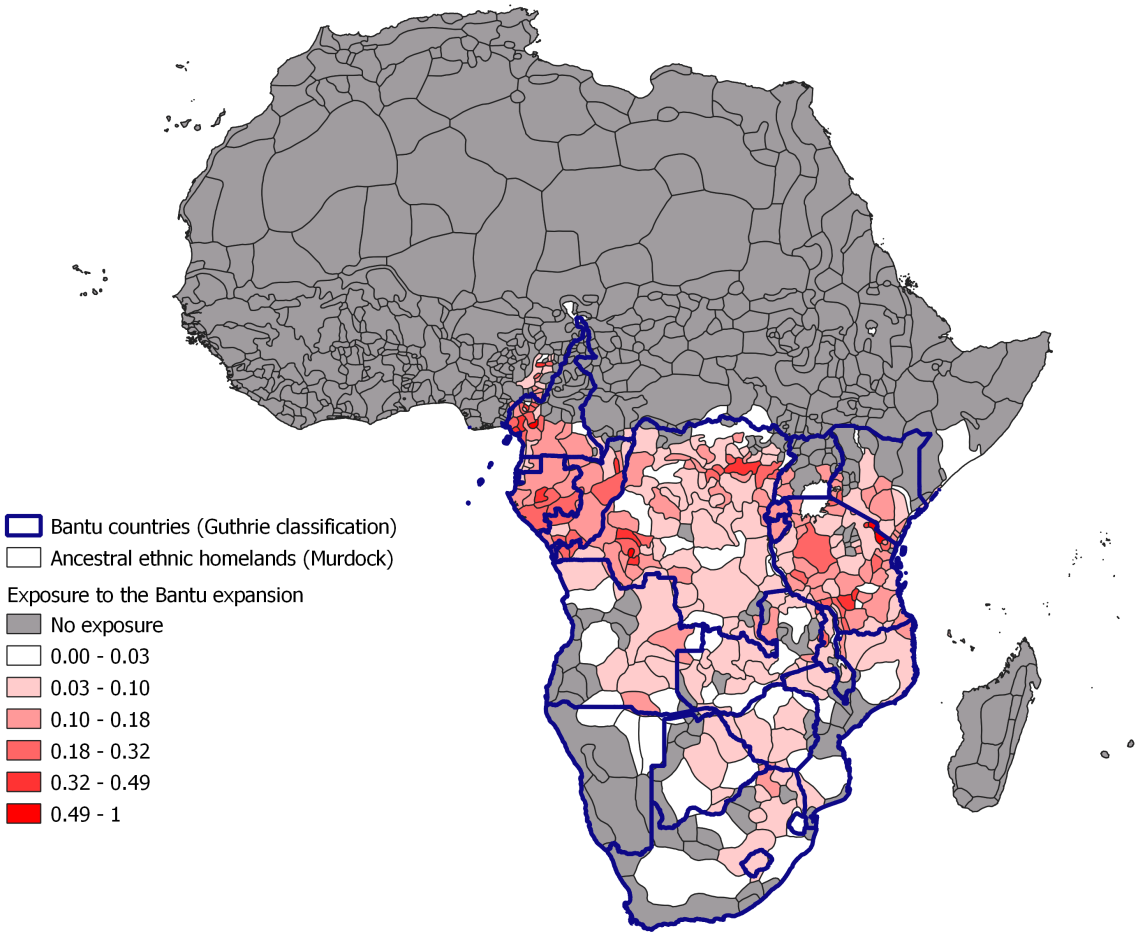
Notes: The figure displays the ethnic groups in the Murdock Ethnographic Atlas that were split by country borders and that I include in the difference-in-differences analysis. I do not include groups that I cannot successfully merge to at least two groups in the EPR dataset belonging to different countries.

FIGURE 4: *The Path of the Bantu Expansion and the Murdock Map*



Notes: The figure on the left shows the historical migration route of the Bantu tribe reconstructed by Grollemund et al. (2015) and contemporaneous country borders. See Appendix B for additional details on the route reconstruction. The figure on the right overlays historical migration route with the map of pre-colonial ethnic settlements constructed by Murdock. Different colors denote the intensity of exposure of each ethnic group to the Bantu expansion, computed as the the length of the path crossing an ethnic homeland divided by the size of the ethnic settlement, all normalized to range between 0 and 1.

FIGURE 5: *The Bantu Index and Countries in the Bantu Region*



Notes: The figure displays the index of exposure to the Bantu expansion based on the historical migration route reconstructed by Grollemund et al. (2015), as well as contemporaneous country borders of the 17 countries classified as belonging to the Bantu region according to the Guthrie classification. These include: Angola, Botswana, Burundi, Cameroon, Congo, Congo DRC, Equatorial Guinea, Gabon, Kenya, Malawi, Mozambique, Rwanda, South Africa, Tanzania, Uganda, Zambia, Zimbabwe.

TABLE 1: *Data extract and Illustration of the Identifying Variation*

country	**year**	**potential rebel**	**Government: ethnic group 1 [*status*]**	**ethnic group 2 [*status*]**	**ethnic group 3 [*status*]**	**ethnic group 4 [*status*]**	**linguistic distance**	**conflict for power**	**conflict for territory**
Congo	1995	Lari/Bakongo	Bembe [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]	Vili [*Junior partner*]	0.098	0	0
Congo	1996	Lari/Bakongo	Bembe [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]	Vili [*Junior partner*]	0.098	0	0
Congo	1997	Lari/Bakongo	Bembe [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]	Vili [*Junior partner*]	0.098	0	0
Congo	1998	Lari/Bakongo	Mbochi (proper) [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]		0.163	1	0
Congo	1999	Lari/Bakongo	Mbochi (proper) [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]		0.163	1	0
Congo	2000	Lari/Bakongo	Mbochi (proper) [*Senior partner*]	Bateke [*Junior partner*]	Kouyou [*Junior partner*]		0.163	0	0

Notes: The table illustrates a data extract for the potential rebel Lari/Bakongo in Congo between 1995 and 2000. The identifying variation is within-ethnicity changes in cultural distance resulting from government changes. In this case, the change in the government coalition occurred in 1998 and Lari/Bakongo experienced an increase in cultural distance and engaged in conflict against the government coalition.

TABLE 2: *Cultural Distance and Ethnic Civil Conflict*

	Dependent variable: ethnic conflict					
	(1)	(2)	(3)	(4)	(5)	(6)
Linguistic distance^{W}	0.013***	0.060***	0.063***	0.072***		
	(0.004)	(0.014)	(0.017)	(0.019)		
Linguistic distance^{UW}					0.051***	0.061***
					(0.015)	(0.022)
Mean of dep. var.	0.073	0.073	0.073	0.073	0.073	0.073
Country-year fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects		yes	yes	yes	yes	yes
Ethnicity year trend		yes	yes	yes	yes	yes
Geographic controls			yes	yes		yes
Climatic controls				yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.239	0.524	0.525	0.522	0.521	0.522

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government and 0 otherwise. *Linguistic distance*^{W} captures linguistic distance between each potential rebel and the ethnic groups at the government, weighted by the position of power of each ethnic group in case the government is a coalition. *Linguistic distance*^{UW} is an unweighted average of linguistic distance between each potential rebel and each ethnic group at the government. The linguistic distance measures are standardized. Geographic controls include the logged geodesic distance between each ethnic group and the government, absolute distance in elevation, absolute distance in ruggedness, and absolute distance in the caloric suitability index (CSI). Climatic controls include yearly average precipitation, yearly average temperature, absolute distance in mean precipitation, and absolute distance in mean temperature to the ethnic groups in power. The sample is an unbalanced panel on 236 distinct ethnic groups in 43 African countries over 57 years (1961-2017). Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE 3: *Cultural Distance and Ethnic Civil Conflict over Government Power and Territory*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.065***	0.075***			-0.004	-0.004		
	(0.015)	(0.017)			(0.008)	(0.011)		
Linguistic distance^{UW}			0.061***	0.072***			-0.009	-0.011
			(0.014)	(0.016)			(0.011)	(0.014)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.491	0.491	0.490	0.492	0.587	0.587	0.587	0.587

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE 4: *Cultural Distance and Ethnic Civil Conflict: Leads and Lags, Anticipatory Effects*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance$_{t-1}$		0.039**	-0.037	-0.041		-0.002	0.005	0.010
		(0.018)	(0.025)	(0.026)		(0.012)	(0.009)	(0.007)
Linguistic distance$_{t}$	0.075***		0.109***	0.079**	-0.004		-0.010	-0.004
	(0.017)		(0.028)	(0.036)	(0.011)		(0.010)	(0.005)
Linguistic distance$_{t+1}$				0.032				-0.005
				(0.029)				(0.008)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,565	9,565	9,305	9,827	9,565	9,565	9,305
Adjusted R-squared	0.491	0.500	0.503	0.510	0.587	0.595	0.595	0.626

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). *Linguistic distance* captures weighted linguistic distance between each potential rebel and the ethnic groups at the government, either in period t-1, t, or t+1. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE 5: *Cultural Distance and Ethnic Civil Conflict: Controlling for Lagged Conflict*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.034***	0.036**			-0.004	-0.004		
	(0.010)	(0.014)			(0.007)	(0.010)		
Linguistic distance^{UW}			0.030***	0.033**			-0.009	-0.010
			(0.011)	(0.015)			(0.010)	(0.012)
Conflict$_{t-1}$	0.572***	0.571***	0.572***	0.571***	0.405***	0.405***	0.405***	0.404***
	(0.081)	(0.082)	(0.081)	(0.082)	(0.115)	(0.115)	(0.115)	(0.114)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,565	9,565	9,565	9,565	9,565	9,565	9,565	9,565
Adjusted R-squared	0.672	0.672	0.672	0.672	0.662	0.662	0.662	0.662

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). Conflict$_{t-1}$ is a binary variable that is equal to 1 if an ethnic group was involved in a conflict over power (columns (1-4)) or over territory (columns (5 -8)) in the previous year. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE 6: *Cultural Distance and Ethnic Civil Conflict: Controlling for Representation in Government Coalition*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.040**	0.055***			-0.008	-0.008		
	(0.015)	(0.020)			(0.008)	(0.011)		
Linguistic distance^{UW}			0.037**	0.051**			-0.015	-0.015
			(0.017)	(0.020)			(0.012)	(0.014)
Ethnic group is part of coalition	-0.075**	-0.081**	-0.076**	-0.082**	-0.013	-0.013	-0.016	-0.017
	(0.033)	(0.038)	(0.034)	(0.039)	(0.012)	(0.015)	(0.012)	(0.015)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.501	0.501	0.500	0.501	0.593	0.593	0.593	0.593

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). *Ethnic group is part of coalition* is equal to 1 if an ethnicity is represented in the government coalition in a certain year and 0 otherwise. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE 7: *Triple Difference in Differences: Exploiting Ethnic Groups Split by Country Borders*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.042*	0.087**			-0.026	-0.053***		
	(0.025)	(0.038)			(0.016)	(0.020)		
Linguistic distance^{UW}			0.055*	0.115***			-0.029*	-0.061***
			(0.029)	(0.038)			(0.017)	(0.022)
Mean of dep. var.	0.054	0.054	0.054	0.054	0.015	0.015	0.015	0.015
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	6,742	6,742	6,742	6,742	6,742	6,742	6,742	6,742
Adjusted R-squared	0.908	0.909	0.908	0.911	0.707	0.714	0.707	0.714

Notes: the unit of observation is a Murdock ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns (1 -4)) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns (5 -8)). For a description of all explanatory variables, refer to notes in Table 2. The sample includes 31 African countries, 57 years (1961-2017), and 84 distinct Murdock ethnic groups. Following Dickens [2018], standard errors are clustered at the ethnic group-country level in parenthesis (corresponding to 181 clusters). Inference is similar when clustering two-way by country and year, or by ethnicity and year. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE 8: *IV: Bantu Exposure, Cultural Distance and Ethnic Civil Conflict*

	Ethnic conflict over government				Ethnic conflict over territory			
Panel A	OLS		2SLS		OLS		2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance	0.072***	0.070***	0.191***	0.139***	-0.000	-0.002**	-0.008	-0.010**
	(0.016)	(0.012)	(0.023)	(0.025)	(0.001)	(0.001)	(0.005)	(0.004)
Mean of dep. var	0.052	0.052	0.052	0.052	0.008	0.008	0.008	0.008
Panel B			Reduced Form				Reduced Form	
Bantu distance			0.070***	0.065***			-0.003*	-0.004**
			(0.013)	(0.013)			(0.002)	(0.002)
Mean of dep. var			0.052	0.052			0.008	0.008
Adjusted R-squared			0.537	0.534			0.330	0.330
Panel C			First Stage				First Stage	
			Linguistic distance				Linguistic distance	
Bantu distance			0.365***	0.468***			0.365***	0.468***
			(0.083)	(0.056)			(0.083)	(0.056)
Kleibergen-Paap F-statistic			14.62	53.64			14.62	53.64
Distance type	w	uw	w	uw	w	uw	w	uw
Country-year FE	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity FE	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity trends	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes	yes	yes	yes	yes
Climatic controls	yes	yes	yes	yes	yes	yes	yes	yes
Observations	3,628	3,628	3,628	3,628	3,628	3,628	3,628	3,628

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting for government power (columns (1-4)) or for territory (columns (5-8)) and 0 otherwise. *Linguistic distance* captures linguistic distance between each potential rebel and the ethnic groups at the government, either weighted or unweighted. *Bantu distance* denotes the absolute difference in the exposure to the Bantu expansion between potential rebels and the government. Linguistic distance and Bantu distance are standardized. For a description of geographic controls, refer to the notes in Table 2. The sample includes 17 African countries in the Bantu region, 57 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis, and are adjusted for the low number of clusters using the number of countries. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE 9: *Cultural Distance and Opinions on Government Performance*

	Dependent variable: Thinks the government is performing badly					
	First principal component			Average number of items		
	(1)	(2)	(3)	(4)	(5)	(6)
Linguistic distance	0.207***	0.207***	0.134**	0.178***	0.178***	0.126**
	(0.050)	(0.050)	(0.054)	(0.050)	(0.050)	(0.053)
Mean of dep. var.	0.559	0.559	0.559	0.555	0.555	0.555
Country-year fixed effects	yes	yes	yes	yes	yes	yes
Survey round fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity year trends	yes	yes	yes	yes	yes	yes
Individual controls		yes	yes		yes	yes
Geographic controls			yes			yes
Climatic controls			yes			yes
Observations	83,482	83,482	83,482	90,481	90,481	90,481
Adjusted R-squared	0.179	0.179	0.180	0.186	0.187	0.187

Notes: the unit of observation is an individual belonging to one of 95 distinct ethnic groups in 27 African countries interviewed between 1999 and 2017 in 6 rounds of the Afrobarometer survey. The dependent variable is a binary variable capturing the extent to which the person thinks the government is performing very badly or fairly badly in handling the following matters: managing the economy, creating jobs, keeping prices stable, narrowing income gap, reducing crime, improving basic health services, addressing educational needs, improving water and sanitation services, ensuring enough to eat, fighting corruption, reducing conflict, combating malaria, combating HIV, maintaining roads and bridges, providing reliable electric supply, managing rivers, promoting equal rights/opportunities for women, addressing needs of youth, protecting rights and promoting opportunities for disabled. Columns 1-3 adopt as dependent variable the first principal component in a principal component analysis grouping all items separately for each survey round. The first principal component loads positively on all items. Columns 4-6 adopt as dependent variable the mean of all items. *Linguistic distance* captures the weighted linguistic distance between a respondent and the ethnic groups at the government. *Linguistic distance* is standardized. For a description of geographic controls, refer to the notes in Table 2. Individual controls include age, age squared, a dummy for female and a dummy for residence in a rural area. Standard errors clustered at the ethnicity-year level are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

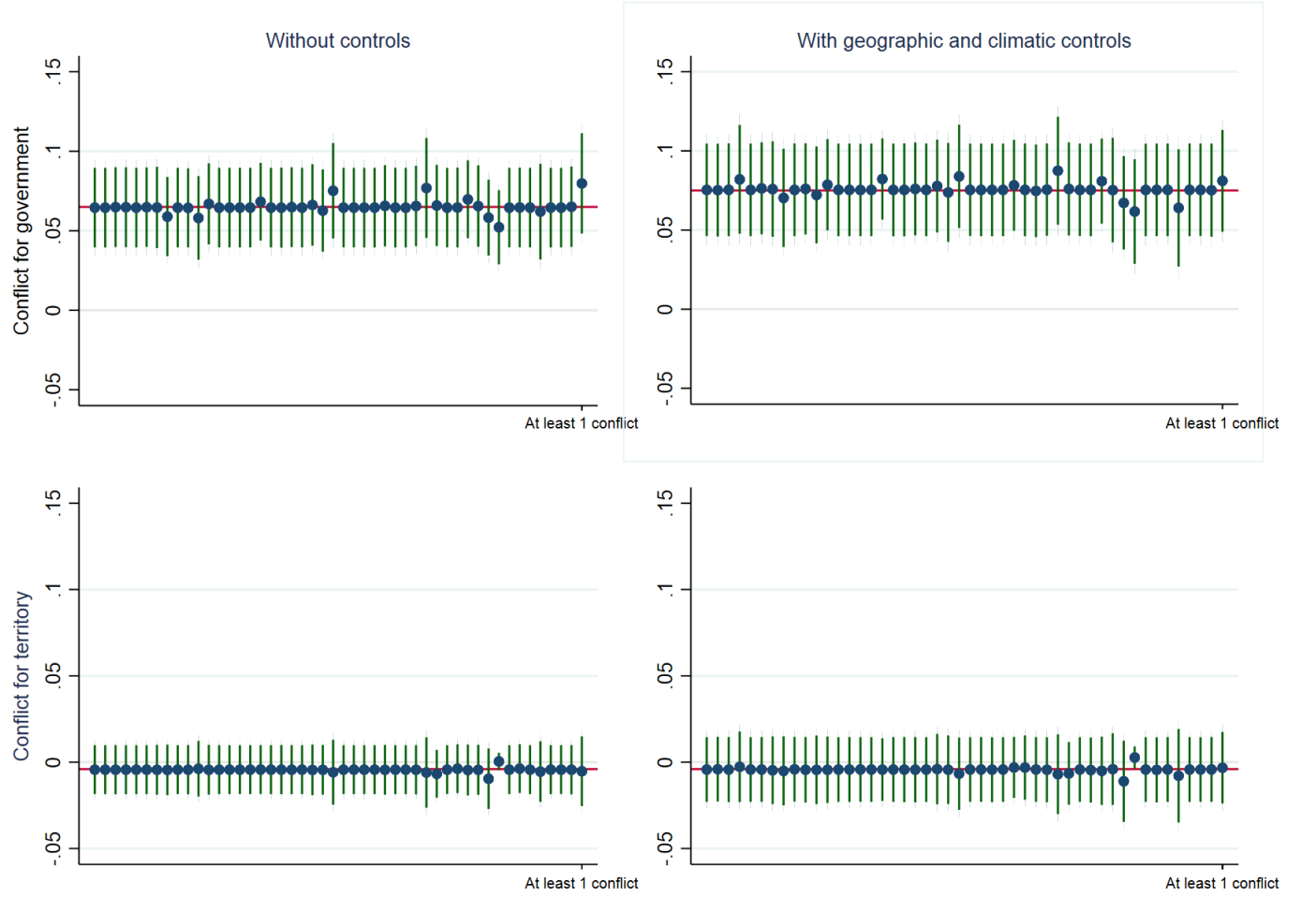
TABLE 10: *The Role of Power Relations*

	Dependent variable: Ethnic conflict over government								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Linguistic distance	0.038**	0.052**	0.057**	0.074***	0.076***	0.084***	0.053***	0.063***	0.069***
	(0.017)	(0.020)	(0.023)	(0.020)	(0.020)	(0.020)	(0.016)	(0.017)	(0.018)
In power	-0.038**	-0.041**	-0.040**						
	(0.017)	(0.019)	(0.019)						
Powerless				-0.017	-0.018	-0.019			
				(0.012)	(0.012)	(0.012)			
Discriminated							0.080**	0.082**	0.082***
							(0.030)	(0.030)	(0.030)
Linguistic distance	-0.002	0.002	0.002	-0.007	-0.006	-0.006	-0.010	-0.011	-0.011
× Status	(0.014)	(0.016)	(0.016)	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)
Mean of dep. var.	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trends	yes	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes	yes		yes	yes		yes	yes
Climatic controls			yes			yes			yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.494	0.494	0.494	0.499	0.499	0.499	0.512	0.512	0.512

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government and 0 otherwise. *Linguistic distance* captures linguistic distance between each potential rebel and the ethnic groups at the government. *Status* is a binary variable which takes value 1 if an ethnic group is *in power* (columns 1-3), *powerless* (columns 4-6), or discriminated by the government (columns 7-9), and 0 otherwise. For a description of geographic controls, refer to the notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

Appendix A: Additional Figures and Tables

FIGURE A-1: *Checking for outliers: dropping one country at a time and keeping only countries that experienced at least one conflict*



Notes: The figures show the stability of the estimates when dropping one country at a time from the sample, or when keeping only the 29 countries that experienced at least one conflict in the whole period (last coefficient on the right in each graph). The red horizontal line indicates the baseline coefficient; green and gray vertical lines indicate 90- and 95-percent confidence intervals, respectively. All specifications include country, year, country-year, and ethnicity fixed effects as well as ethnicity-specific time trends.

TABLE A-1: *Cultural Distance and Ethnic Civil Conflict: Controlling for Lagged Conflict Instrumenting the First Lag*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.027**	0.028*			-0.004	-0.004		
	(0.012)	(0.015)			(0.007)	(0.009)		
Linguistic distance^{UW}			0.025*	0.026			-0.008	-0.009
			(0.014)	(0.017)			(0.009)	(0.012)
Conflict$_{t-1}$	0.649***	0.649***	0.649***	0.649***	0.752***	0.752***	0.752***	0.751***
	(0.117)	(0.118)	(0.118)	(0.118)	(0.181)	(0.182)	(0.181)	(0.182)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,314	9,314	9,314	9,314	9,314	9,314	9,314	9,314

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1 -4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Conflict$_{t-1}$ is a binary variable that is equal to 1 if an ethnic group was involved in a conflict over power (columns 1-4) or over territory (columns 5 -8) in the previous year. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 55 years (1963-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-2: *Robustness: Controlling for Potential Rebels Having Lost Power*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.059***	0.072***			-0.004	-0.004		
	(0.013)	(0.017)			(0.008)	(0.011)		
Linguistic distance^{UW}			0.055***	0.068***			-0.010	-0.011
			(0.013)	(0.016)			(0.011)	(0.014)
Ethnic group lost power	0.038**	0.037**	0.038**	0.037**	-0.000	-0.000	0.001	0.000
	(0.015)	(0.015)	(0.015)	(0.015)	(0.006)	(0.006)	(0.006)	(0.006)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.499	0.500	0.499	0.500	0.593	0.593	0.593	0.593

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1 -4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5 -8). *Ethnic group lost power* is equal to 1 if an ethnicity has lost government power. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-3: *Robustness: Controlling for Potential Rebels Having Gained Power*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.063***	0.074***			-0.005	-0.004		
	(0.015)	(0.018)			(0.008)	(0.011)		
Linguistic distance^{UW}			0.059***	0.071***			-0.010	-0.011
			(0.015)	(0.017)			(0.011)	(0.014)
Ethnic group gained power	-0.010	-0.009	-0.010	-0.009	-0.001	-0.002	-0.002	-0.003
	(0.011)	(0.012)	(0.011)	(0.012)	(0.004)	(0.004)	(0.003)	(0.004)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.498	0.498	0.497	0.498	0.593	0.593	0.593	0.593

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). *Ethnic group lost power* is equal to 1 if an ethnicity has gained government power. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-4: *Robustness: Alternative Linguistic Distance Measure*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.043**	0.057**			-0.008	-0.008		
	(0.018)	(0.024)			(0.007)	(0.010)		
Linguistic distance^{UW}			0.039**	0.049**			-0.015	-0.016
			(0.017)	(0.023)			(0.009)	(0.011)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.495	0.496	0.495	0.496	0.593	0.593	0.593	0.593

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Linguistic distance indicates the (weighted or unweighted) linguistic distance between potential rebels and each ethnic group in the government coalition excluding the potential rebel itself from the computation if the latter is represented in the government. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-5: *Robustness: Only Countries that Experienced at Least One Conflict*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.080***	0.081***			-0.005	-0.003		
	(0.019)	(0.019)			(0.012)	(0.012)		
Linguistic distance^{UW}			0.074***	0.077***			-0.013	-0.012
			(0.018)	(0.017)			(0.016)	(0.016)
Mean of dep. var.	0.070	0.070	0.070	0.070	0.036	0.036	0.036	0.036
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	6,737	6,737	6,737	6,737	6,737	6,737	6,737	6,737
Adjusted R-squared	0.484	0.489	0.483	0.489	0.587	0.588	0.587	0.588

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Linguistic distance indicates the (weighted or unweighted) linguistic distance between potential rebels and each ethnic group in the government coalition. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 28 African countries, 57 years (1961-2017), and 166 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-6: *Robustness: Only Countries that Experienced at Least One Government Change*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.068***	0.074***			-0.012	-0.012		
	(0.018)	(0.019)			(0.014)	(0.016)		
Linguistic distance^{UW}			0.063***	0.068***			-0.018	-0.019
			(0.018)	(0.018)			(0.017)	(0.018)
Mean of dep. var.	0.053	0.053	0.053	0.053	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	8,132	8,132	8,132	8,132	8,132	8,132	8,132	8,132
Adjusted R-squared	0.481	0.482	0.481	0.482	0.586	0.587	0.587	0.587

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Linguistic distance indicates the (weighted or unweighted) linguistic distance between potential rebels and each ethnic group in the government coalition. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 31 African countries, 57 years (1961-2017), and 202 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-7: *Robustness: No Time Trends*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.023***	0.025***			-0.002	-0.003		
	(0.006)	(0.009)			(0.004)	(0.005)		
Linguistic distance^{UW}			0.020**	0.023*			-0.003	-0.005
			(0.008)	(0.011)			(0.006)	(0.007)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend								
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.429	0.429	0.429	0.430	0.487	0.488	0.487	0.488

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-8: *Robustness: Alternative Clustering (Country Level)*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.065***	0.075***			-0.004	-0.004		
	(0.016)	(0.017)			(0.009)	(0.011)		
Linguistic distance^{UW}			0.061***	0.072***			-0.009	-0.011
			(0.015)	(0.016)			(0.012)	(0.014)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.498	0.491	0.491	0.492	0.587	0.587	0.587	0.587

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Robust standard errors clustered at the country level are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-9: *Robustness: Alternative Clustering (Country-by-year Level)*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.065***	0.075***			-0.004	-0.004		
	(0.014)	(0.016)			(0.004)	(0.006)		
Linguistic distance^{UW}			0.061***	0.072***			-0.009*	-0.011
			(0.013)	(0.016)			(0.006)	(0.007)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.493	0.493	0.492	0.493	0.589	0.589	0.589	0.589

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Robust standard errors clustered at the country-year level are reported in parenthesis (corresponding to 2,000 clusters). *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-10: *Robustness: Alternative Clustering (Ethnic Group Level)*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.065***	0.075**			-0.004	-0.004		
	(0.024)	(0.034)			(0.008)	(0.012)		
Linguistic distance^{UW}			0.061**	0.072*			-0.009	-0.011
			(0.028)	(0.038)			(0.009)	(0.012)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,827	9,827	9,827	9,827	9,827	9,827	9,827	9,827
Adjusted R-squared	0.495	0.495	0.495	0.496	0.590	0.591	0.591	0.591

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 236 distinct ethnic groups. Robust standard errors clustered at the ethnicity level are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-11: *Robustness: Lagged Linguistic Distance*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance$^{W}{}_{t-1}$	0.042***	0.039**			-0.002	-0.002		
	(0.015)	(0.018)			(0.009)	(0.012)		
Linguistic distance$^{UW}{}_{t-1}$			0.038**	0.037*			-0.006	-0.007
			(0.015)	(0.018)			(0.013)	(0.015)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	9,565	9,565	9,565	9,565	9,565	9,565	9,565	9,565
Adjusted R-squared	0.499	0.500	0.499	0.500	0.595	0.595	0.595	0.596

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). *Linguistic distance* indicates the (weighted or unweighted) lagged linguistic distance between potential rebels and the government. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 207 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-12: *Robustness: Balanced Panel*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.051**	0.052**			-0.006	-0.009		
	(0.019)	(0.022)			(0.011)	(0.018)		
Linguistic distance^{UW}			0.048**	0.047**			-0.013	-0.020
			(0.019)	(0.020)			(0.015)	(0.020)
Mean of dep. var.	0.049	0.049	0.049	0.049	0.025	0.025	0.025	0.025
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	8,703	8,703	8,703	8,703	8,703	8,703	8,703	8,703
Adjusted R-squared	0.520	0.521	0.520	0.521	0.591	0.591	0.591	0.592

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). For a description of all explanatory variables, refer to notes in Table 2. The sample includes 43 African countries, 57 years (1961-2017), and 207 distinct ethnic groups observed uninterruptedly throughout the period. Two-way clustered standard errors by year and country are reported in parenthesis. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-13: *DID Robustness: Controlling for Lagged Conflict*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.028*	0.058***			-0.027*	-0.052***		
	(0.016)	(0.020)			(0.016)	(0.020)		
Linguistic distance^{UW}			0.035*	0.074***			-0.030*	-0.060***
			(0.019)	(0.020)			(0.018)	(0.022)
Conflict$_{t-1}$	0.394***	0.389***	0.390***	0.381***	-0.000	-0.012	0.001	-0.012
	(0.050)	(0.048)	(0.051)	(0.049)	(0.077)	(0.079)	(0.077)	(0.079)
Mean of dep. var.	0.054	0.054	0.054	0.054	0.015	0.015	0.015	0.015
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	6,589	6,589	6,589	6,589	6,589	6,589	6,589	6,589
Adjusted R-squared	0.923	0.924	0.923	0.924	0.716	0.720	0.716	0.721

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Conflict$_{t-1}$ is a binary variable that is equal to 1 if an ethnic group was involved in a conflict over power (columns 1-4) or over territory (columns 5-8) in the previous year. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 31 African countries, 57 years (1961-2017), and 84 distinct Murdock ethnic groups. Standard errors clustered at the ethnic group-country level in parenthesis (corresponding to 201 clusters). *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-14: *DID Robustness: Controlling for Potential Rebel Being Part of Government Coalition*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance^{W}	0.043*	0.087**			-0.030*	-0.053***		
	(0.025)	(0.038)			(0.016)	(0.020)		
Linguistic distance^{UW}			0.059**	0.114***			-0.033*	-0.061***
			(0.028)	(0.038)			(0.018)	(0.022)
Ethnic group is part of coalition	0.013	-0.010	0.023	-0.008	-0.022*	-0.010	-0.024*	-0.010
	(0.023)	(0.017)	(0.024)	(0.016)	(0.013)	(0.007)	(0.013)	(0.007)
Mean of dep. var.	0.054	0.054	0.054	0.054	0.015	0.015	0.015	0.015
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls		yes		yes		yes		yes
Climatic controls		yes		yes		yes		yes
Observations	7,874	7,874	7,874	7,874	7,874	7,874	7,874	7,874
Adjusted R-squared	0.858	0.860	0.858	0.861	0.794	0.799	0.794	0.799

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power and 0 otherwise (columns 1-4) and that takes value 1 if an ethnic group is fighting the government for the exclusive control of a territory and 0 otherwise (columns 5-8). Ethnic groups is part of coalition is equal to 1 if an ethnicity is represented in the government coalition and 0 otherwise. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 31 African countries, 57 years (1961-2017), and 84 distinct Murdock ethnic groups. Standard errors clustered at the ethnic group-country level in parenthesis (corresponding to 201 clusters). *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-15: *DID: Leads and Lags*

	Ethnic conflict over power				Ethnic conflict over territory			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Linguistic distance$_{t-1}$		0.039**		-0.005		-0.036*		-0.020
		(0.015)		(0.029)		(0.021)		(0.019)
Linguistic distance$_{t}$	0.087**		0.093***	0.098**	-0.053***		-0.055**	-0.040**
	(0.038)		(0.032)	(0.050)	(0.020)		(0.026)	(0.019)
Linguistic distance$_{t+1}$			-0.003	-0.005			0.007	0.006
			(0.019)	(0.017)			(0.015)	(0.016)
Mean of dep. var.	0.054	0.054	0.055	0.054	0.015	0.015	0.014	0.014
Country-ethn. fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Ethnicity-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Country-year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes	yes	yes	yes	yes
Climatic controls	yes	yes	yes	yes	yes	yes	yes	yes
Observations	6,742	6,589	6,558	6,405	6,742	6,589	6,558	6,405
Adjusted R-squared	0.909	0.907	0.909	0.909	0.714	0.720	0.713	0.720

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government over power (columns 1-5) or over territory (columns 6-10) and 0 otherwise. *Linguistic distance* captures linguistic distance between each potential rebel and the ethnic groups at the government, either in period t-1, t, or t+1. *Linguistic distance* is standardized. For a description of geographic and climatic controls, refer to the notes in Table 2. The sample includes 31 African countries, 57 years (1961-2017), and 84 distinct Murdock ethnic groups. Standard errors clustered at the ethnic group-country level in parenthesis (corresponding to 201 clusters). *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-16: *IV: Inference with Lee et al. [2020] tF Procedure*

	Conflict for government		Conflict for territory	
	(1)	(2)	(3)	(4)
Linguistic distance	0.191***	0.139***	-0.008	-0.010**
tF 0.05 se	(0.034)	(0.027)	(0.007)	(0.004)
Mean of dep. var	0.052	0.052	0.008	0.008
Distance type	w	uw	w	uw
Country-year FE	yes	yes	yes	yes
Ethnicity FE & trends	yes	yes	yes	yes
Controls	yes	yes	yes	yes
Observations	3,628	3,628	3,628	3,628

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government over power (columns 1-2) or over territory (columns 3-4) and zero otherwise. *Linguistic distance* captures linguistic distance between each potential rebel and the ethnic groups at the government. For a description of geographic controls, refer to the notes in Table 2. The sample includes 17 African countries in the Bantu region, 57 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country, adjusted for the low number of clusters using the number of countries, are reported in parenthesis and are further re-adjusted using Lee *et al.* [2020] *tF* procedure. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-17: *IV Robustness: Controlling for Lagged Conflict*

	Ethnic conflict over power			Ethnic conflict over territory		
	OLS	2SLS		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Linguistic distance	0.046***	0.129***	0.087***	0.000	-0.002	-0.002
	(0.011)	(0.010)	(0.013)	(0.008)	(0.008)	(0.008)
Conflict_{t-1}	0.446***	0.434***	0.439***	0.164	0.164	0.163
	(0.052)	(0.054)	(0.053)	(0.193)	(0.219)	(0.219)
Mean of dep. var.	0.052	0.052	0.052	0.008	0.008	0.008
Distance type	w	w	uw	w	w	uw
Country-year fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes	yes	yes
Climatic controls	yes	yes	yes	yes	yes	yes
Observations	3,537	3,537	3,537	3,537	3,537	3,537

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government over power (columns 1-3) or territory (columns 4-6) and 0 otherwise. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 17 African countries in the Bantu region, 56 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis, and are adjusted for the low number of clusters using the number of countries. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-18: *IV Robustness: Controlling for Representation in Government Coalition*

	Ethnic conflict over power			Ethnic conflict over territory		
	OLS	2SLS		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Linguistic distance	0.057***	0.180***	0.107**	-0.006	-0.029**	-0.034
	(0.019)	(0.051)	(0.046)	(0.004)	(0.012)	(0.021)
Ethnic group is part of coalition	-0.088	-0.022	-0.052	-0.032	-0.044*	-0.041
	(0.068)	(0.065)	(0.073)	(0.019)	(0.022)	(0.026)
Mean of dep. var.	0.052	0.052	0.052	0.008	0.008	0.008
Distance type	w	w	uw	w	w	uw
Country-year fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes	yes	yes
Ethnicity year trend	yes	yes	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes	yes	yes
Climatic controls	yes	yes	yes	yes	yes	yes
Observations	3,628	3,628	3,628	3,628	3,628	3,628

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government for gaining power or territory and 0 otherwise. *Ethnic group is part of coalition* is equal to 1 if an ethnicity is represented in the government coalition in a certain year and 0 otherwise. For a description of all explanatory variables, refer to notes in Table 2. The sample includes 17 African countries in the Bantu region, 56 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis, and are adjusted for the low number of clusters using the number of countries. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-19: *Reduced Form: Splitting Bantu Distance into Two*

	Ethnic conflict over power		Ethnic conflict over territory	
	(1)	(2)	(3)	(4)
Bantu more	0.059	0.063	-0.001	-0.001
	(0.046)	(0.050)	(0.005)	(0.005)
Bantu less	0.056	0.044	-0.002	-0.004
	(0.058)	(0.059)	(0.003)	(0.004)
F statistics (Bantu more - Bantu less=0)	0.00	0.03	0.00	0.13
Prob > F	0.976	0.861	0.977	0.722
Country-year fixed effects	yes	yes	yes	yes
Ethnicity fixed effects	yes	yes	yes	yes
Ethnicity year trends	yes	yes	yes	yes
Geographic controls	yes	yes	yes	yes
Climatic controls		yes		yes
Observations	3,628	3,628	3,628	3,628
Adjusted R-squared	0.531	0.534	0.329	0.329

Notes: the unit of observation is an ethnic group-country-year. The dependent variable is a binary variable that takes value 1 if an ethnic group is fighting the government and 0 otherwise. *Bantu more* and *Bantu less* is a continuous variable that denotes whether the rebel was more or less exposed to the Bantu expansion than the government, respectively. For a description of geographic controls, refer to the notes in Table 2. The sample includes 17 African countries in the Bantu region, 57 years (1961-2017), and 86 distinct ethnic groups. Two-way clustered standard errors by year and country are reported in parenthesis, and are adjusted for the low number of clusters using the number of countries. *** (**) (*) indicate significance at the 1% (5%) (10%) level.

TABLE A-20: *Principal Component Analysis: Loadings and sampling adequacy*

	Loadings						
How well or badly would you say the current government is handling the following matters:	Round 1 1999-2001	Round 2 2002-2003	Round 3 2005-2006	Round 4 2008-2009	Round 5 2011-2013	Round 6 2014-2015	Round 7 2016-2017
Managing the economy		0.299	0.324	0.264	0.273	0.290	0.225
Improving living standards of the poor				0.283	0.283	0.303	0.237
Creating jobs	0.580	0.287	0.311	0.272	0.265	0.293	0.239
Keeping prices stable	0.589	0.286	0.305	0.243	0.233	0.277	0.217
Narrowing the income gap	0.562	0.281	0.312	0.269	0.266	0.286	0.232
Reducing crime		0.288	0.291	0.240	0.246	0.258	0.221
Improving basic health services		0.294	0.312	0.262	0.263	0.278	0.243
Addressing educational needs		0.290	0.306	0.241	0.253	0.271	0.241
Improving water and sanitation services		0.216	0.273	0.244	0.243	0.268	0.216
Ensuring enough to eat		0.272	0.317	0.272	0.264	0.284	0.231
Fighting corruption		0.299	0.309	0.250	0.261	0.277	0.227
Reducing conflict		0.287			0.247		
Combating HIV		0.242	0.249	0.190	0.190		
Combating malaria		0.251					
Maintaining roads and bridges				0.242	0.237	0.255	0.205
Providing reliable electric supply				0.246	0.235	0.258	0.203
Protecting rivers and forests				0.243			
Promoting equal rights and opportunities for women				0.224	0.223		0.226
Preventing election violence							0.238
Preventing or resolving violent community conflict							0.243
Countering violence from armed extremists							0.228
Addressing needs of youth							0.247
Protecting rights, promoting opportunities for disabled							0.236
Kaiser-Meyer-Olkin measure of sampling adequacy	0.664	0.880	0.887	0.918	0.920	0.928	0.938
Common variance explained by first principal component	61%	34%	37%	35%	35%	43%	36%
N	14,334	21,906	25,333	27,515	51,150	52,536	18,004

Notes: The table illustrates the loadings of each variable from the Afrobarometer survey used for the principal component analysis. Each variable is binary and is equal to 1 if the respondent thinks the government is performing badly or fairly badly in handling each policy goal. Sample sizes denote the full set of respondent answering each question. The final sample in Table 9 includes only individuals that were successfully merged with the conflict dataset.

Appendix B: Data Description, Sources, and Summary Statistics

B-1 Data Sources

Ethnic groups: Data on ethnic groups comes from the Ethnic Power Relations (EPR) dataset family (Vogt *et al.* [2015]). EPR lists each politically relevant ethnic group for each country in each year, and the respective access (or lack thereof) to executive government power. EPR defines an ethnic group as an identity group that defines itself or is defined by others along linguistic, religious or racial characteristics.
Source: https://icr.ethz.ch/data/epr/core/

Linguistic distance: I retrieve data on languages (up to 3) spoken by each ethnic group from the Ethnic Dimensions Dataset. I compute a measure of linguistic distance (called cladistic distance) using information on linguistic trees coming from the Ethnologue database. I describe the methodology for the construction of the linguistic distance measure in section 3..
Sources: https://icr.ethz.ch/data/epr/ed/ and https://www.ethnologue.com/browse/names

Conflict: Data on ethnic civil conflicts comes from the UCDP/PRIO Armed Conflict Dataset (version 20.1), with information on the identity of the combatants involved in a conflict (rebel groups and governments). Information on the ethnic identity of rebel groups comes from the ACD2EPR dataset.
Sources: https://ucdp.uu.se/downloads/ and https://icr.ethz.ch/data/epr/acd2epr/

Geodesic distance: I compute geodesic distances between each group's centroid, which I calculate based on ethnic settlements provided by the GeoEPR dataset (Wucherpfennig *et al.* [2011]). Throughout the analysis, I use log geodesic distance.
Source: https://icr.ethz.ch/data/epr/geoepr/

Absolute difference in elevation: I combine GeoEPR with data on elevation at the grid level (2.5 arc-minute resolution) provided by the Worldclim Global Climate Database (Fick and Hijmans [2017]). I measure average elevation of each ethnic group, and then calculate the absolute distance in elevation between each ethnic group and the ethnicities forming the government.
Source: https://www.worldclim.org/data/worldclim21.html

Absolute difference in ruggedness: I produce a measure of ruggedness for each ethnic group using the standard deviation of the Worldclim elevation data. I calculate the absolute distance in ruggedness between each ethnic group and the ethnicities forming the government.
Source: https://www.worldclim.org/data/worldclim21.html

Absolute difference in the Caloric Suitability Index: Data on the potential agricultural output come from the Caloric Suitability Index (CSI) provided by Galor and Özak [2016]. This measure reflects the potential caloric yield of a grid cell based on the Global Agro-Ecological Zones (GAEZ) project of the Food and Agriculture Organization (FAO). The Index results from a combination of climatic and grographic variables unaffected by human activity. I use the pre-1500 average CSI measure that includes cells with zero productivity. I calculate the average CSI for each ethnic group, and calculate the absolute distance in CSI between each ethnic group and the ethnicities forming the government.

Source: https://ozak.github.io/Caloric-Suitability-Index/

Absolute difference in temperature: Temperature data comes from the WorldClim Global Climate Database. I compute a time-varying measure of mean temperature for each group, and calculate the absolute distance in mean temperature between each ethnic group and the ethnicities forming the government.
Source: https://www.worldclim.org/data/worldclim21.html

Absolute difference in rainfall: Rainfall data comes from the WorldClim Global Climate Database. I compute a time-varying measure of average rainfall for each group, and calculate the absolute distance in mean rainfall between each ethnic group and the ethnicities forming the government.
Source: https://www.worldclim.org/data/worldclim21.html

Ancestral homelands of contemporary ethnicities: I retrieve the geolocation of ethnic groups' ancestral homelands from the Murdock Map, digitized by Nunn and Wantchekon [2011]. I merge EPR ethnicities to the corresponding Murdock groups' counterparts using the LEDA R-package (see details in Section B-2.)
Source: https://scholar.harvard.edu/nunn/pages/data-0

The Bantu expansion route. Data on the Bantu expansion route comes from work by Grollemund *et al.* [2015]. They reconstructed the route by collecting the indigenous geographic location of 424 Bantu languages, including now-extinct languages. Using Bayesian techniques on a sample of 100 lexical items, they construct a pyhlogenetic tree connecting all these languages. Based on this, they reconstruct the probable ancestral geographical locations of each of the internal nodes of the phylogenetic tree through a model calibrated using archaeological evidence (e.g., node 1 is dated back to 4000-5000 before present in the Grassfields region of Cameroon, on the basis of an archaeological site called Shum Laka, the principal site associated with Bantu homelands). Using other archeological sites, they calibrate other branching points of the tree. using a Brownian motion model, they infer the ancestral latitude and longitude for each internal node of the tree, which they then connect using straight lines. Additional methodological details can be found in Grollemund *et al.* [2015] and in the article's Supplementary Appendix.

Opinions on government performance: Data on individuals' opinion on the performance of the current government comes from seven rounds of the Afrobarometer survey, conducted between 1999 and 2017 in 27 African countries. See Table A-20 for additional details on the questions asked in each round. Refer to Section 8. for details on how I construct a measure of an individual's degree of disagreement with policies implemented by the current government. I merge the self-reported ethnolinguistic affiliation of Afrobarometer respondents ethnicities to the corresponding EPR groups using the LEDA R-package (see details in Section B-2.)
Source: http://afrobarometer.org/data

Age: Age of respondent at the time of survey.

Gender: An indicator variable equal to one if a respondent is female.

Rural: An indicator variable for rural locations.

B-2 Merging Ethnic Groups Across Datasets

To merge ethnicities across datasets, I use a recent R-package created by Müller-Crepon *et al.* [2020] called Linking Ethnic Data for Africa (LEDA). Since ethnic categories vary considerably across datasets, Müller-Crepon *et al.* [2020] created an algorithm that systematically links ethnic groups across 11 dataset. Using a dictionary-based linking procedure, they match more than 8,100 ethnicities via the list of known language families, languages, and dialects from the 16th edition of the Ethnologue database.

Since links are formed through the linguistic tree, the LEDA package allows the researcher to choose the level of precision for each match. One option is to use the so-called *set overlap* rule, which generates a link between any two groups that share at least one language node at a specified level of the language tree. The higher the level specified, the higher is the completeness, but the lower the precision of the match. Throughout the analysis, I opt for precision of matches, and choose the *dialect* level (i.e., the lowest level) of the linguistic tree of ethnic group in dataset A for generating matches to the ethnic group(s) in dataset B.

Even after choosing the matching procedure that favors precision, some matches remain more precise than others. Conveniently, for each match, the package provides an indication of the degree of precision by indicating the extent to which the linguistic tree of an ethnic group in dataset A overlaps with the linguistic tree of an ethnic group in dataset B. Whenever LEDA matches an ethnic group in dataset A (e.g., the EPR dataset) to multiple groups in dataset B (e.g., the Murdock Map), I employ this accuracy measure as a weight when calculating statistics for group in dataset A based on multiple groups in dataset B (e.g., when computing a measure of Bantu exposure for each EPR group based on the Bantu Index of multiple Murdock groups). Groups in dataset B that constitute more accurate matches will get more weight than groups constituting weaker links.

For additional information on the LEDA package, see: http://www.carlmueller-crepon.org/project/ethnic_matching/

B-3 Summary Statistics

TABLE B-1: *Number of Government Changes by Country*

Country	Number of changes	Country	Number of changes
Angola	0	Malawi	1
Benin	8	Mali	1
Botswana	0	Mauritania	0
Burundi	0	Mauritius	8
Cameroon	2	Morocco	0
Central African Republic	5	Mozambique	1
Chad	8	Namibia	1
Comoros	0	Niger	10
Congo	9	Nigeria	9
Congo, DRC	8	Senegal	2
Cote d'Ivoire	5	Sierra Leone	8
Djibouti	2	South Africa	1
Egypt	0	South Sudan	0
Equatorial Guinea	0	Sudan	2
Eritrea	1	Tanzania	1
Ethiopia	2	The Gambia	0
Gabon	2	Togo	5
Ghana	6	Uganda	5
Guinea	5	Zambia	1
Guinea-Bissau	6	Zimbabwe	4
Kenya	6		
Liberia	3		
Libya	0	**Total**	**138**

Notes: The table reports the number of changes in the ethnic identity of governments experienced by each country in the sample over the period 1961-2017.

TABLE B-2: *Summary Statistics: Africa*

	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)	Obs. (5)
Conflict Variables					
Ethnic conflict	0.073	0.259	0	1	9,827
Conflict over government	0.049	0.215	0	1	9,827
Conflict over territory	0.025	0.155	0	1	9,827
Linguistic Distance					
Linguistic Distance^{W}	0.402	0.321	0	1	9,827
Linguistic Distance^{UW}	0.405	0.317	0	1	9,827
Geographic Controls					
Log geodesic distance^{W}	5.530	1.272	0	7.284	9,827
Log geodesic distance^{UW}	5.524	1.287	0	7.274	9,827
Absolute distance ruggedness^{W}	131.0	153.8	0	1,223	9,827
Absolute distance ruggedness^{UW}	133.5	155.4	0	1,238	9,827
Absolute distance elevation^{W}	214.9	242.4	0	1,442	9,827
Absolute distance elevation^{UW}	215.0	240.5	0	1,384	9,827
Absolute distance CSI^{W}	588.5	1,157	0	9,209	9,827
Absolute distance CSI^{UW}	589.9	1,135	0	9,206	9,827
Climatic Controls					
Average precipitation	92.27	61.62	0.280	353.3	9,827
Absolute distance precipitation^{W}	22.83	24.15	0	254.9	9,827
Absolute distance precipitation^{UW}	22.66	23.64	0	254.5	9,827
Average temperature	24.13	2.977	15.29	29.35	9,827
Absolute distance temperature^{W}	1.256	1.247	0	7.545	9,827
Absolute distance temperature^{UW}	1.257	1.235	0	7.516	9,827

Notes: The sample includes 236 distinct ethnic groups in 43 African countries over a period of 57 years (1961-2017). The superscript W indicates average distances between potential rebels and each ethnic group in the government coalition weighted by the role of ethnic groups in power (where senior partners receive double the weight of each junior partner); the superscript UW indicates unweighted average distances between potential rebels and each ethnic group in the government coalition.

TABLE B-3: *Summary Statistics: Split Ethnic Groups*

	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)	Obs. (5)
Conflict Variables					
Ethnic conflict	0.069	0.252	0	1	6,742
Conflict over government	0.054	0.225	0	1	6,742
Conflict over territory	0.015	0.122	0	1	6,742
Linguistic Distance					
Linguistic Distance^{W}	0.486	0.306	0	1	6,742
Linguistic Distance^{UW}	0.497	0.302	0	1	6,742
Geographic Controls					
Log geodesic distance^{W}	5.843	0.966	0	7.254	6,742
Log geodesic distance^{UW}	5.882	0.948	0	7.254	6,742
Absolute distance ruggedness^{W}	112.6	132.9	0	1,191	6,742
Absolute distance ruggedness^{UW}	118.0	136.6	0	1,191	6,742
Absolute distance elevation^{W}	172.5	147.7	0	1,196	6,742
Absolute distance elevation^{UW}	175.4	144.6	0	1,196	6,742
Absolute distance CSI^{W}	420.7	696.2	0	6,655	6,742
Absolute distance CSI^{UW}	450.9	723.4	0	6,655	6,742
Climatic Controls					
Average precipitation	56.73	46.28	0.295	288.9	6,742
Absolute distance precipitation^{W}	16.95	15.48	0	101.7	6,742
Absolute distance precipitation^{UW}	17.44	15.65	0	97.09	6,742
Average temperature	24.18	3.341	17.08	29.13	6,742
Absolute distance temperature^{W}	1.076	0.886	0	6.576	6,742
Absolute distance temperature^{UW}	1.094	0.877	0	6.576	6,742

Notes: The sample includes 31 African countries, 57 years (1961-2017), and 84 distinct Murdock ethnic groups. The superscript W indicates average distances between potential rebels and each ethnic group in the government coalition weighted by the role of ethnic groups in power (where senior partners receive double the weight of each junior partner); the superscript UW indicates unweighted average distances between potential rebels and each ethnic group in the government coalition.

TABLE B-4: *Summary Statistics: Bantu Region*

	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)	Obs. (5)
Conflict Variables					
Ethnic conflict	0.060	0.238	0	1	3,628
Conflict over government	0.052	0.222	0	1	3,628
Conflict over territory	0.008	0.089	0	1	3,628
Linguistic Distance					
Linguistic Distance^{W}	0.257	0.273	0	1	3,628
Linguistic Distance^{UW}	0.264	0.274	0	1	3,628
Bantu Distance					
Bantu Distance^{W}	0.048	0.048	0	0.435	3,628
Bantu Distance^{UW}	0.054	0.274	0	0.596	3,628
Geographic Controls					
Log geodesic distance^{W}	5.603	1.349	0	7.284	3,628
Log geodesic distance^{UW}	5.600	1.349	0	7.274	3,628
Absolute distance ruggedness^{W}	148.1	120.7	0	676.1	3,628
Absolute distance ruggedness^{UW}	151.8	126.9	0	688.7	3,628
Absolute distance elevation^{W}	265.7	245.1	0	1,442	3,628
Absolute distance elevation^{UW}	269.8	243.2	0	1,384	3,628
Absolute distance CSI^{W}	331.7	597.6	0	3,646	3,628
Absolute distance CSI^{UW}	360.0	646.4	0	3,440	3,628
Climatic Controls					
Average precipitation	106.7	47.46	10.94	304.1	3,628
Absolute distance precipitation^{W}	19.41	16.57	0	101.7	3,628
Absolute distance precipitation^{UW}	19.42	16.29	0	97.09	3,628
Average temperature	22.47	2.319	15.29	25.96	3,628
Absolute distance temperature^{W}	1.312	1.079	0	5.911	3,628
Absolute distance temperature^{UW}	1.327	1.057	0	5.757	3,628

Notes: The sample includes 86 distinct ethnic groups in 17 African countries corresponding to regions covered by Bantu languages according to the Guthrie classification (Angola, Botswana, Burundi, Cameroon, Congo, Congo DRC, Equatorial Guinea, Gabon, Kenya, Malawi, Mozambique, Rwanda, South Africa, Tanzania, Uganda, Zambia, Zimbabwe) over a period of 57 years (1961-2017). The superscript W indicates average distances between potential rebels and each ethnic group in the government coalition weighted by the role of ethnic groups in power (where senior partners receive double the weight of each junior partner); the superscript UW indicates unweighted average distances between potential rebels and each ethnic group in the government coalition.

TABLE B-5: *Summary Statistics: Individual-Level Data*

	Mean (1)	Std. Dev. (2)	Min (3)	Max (4)	Obs. (5)
Opinion on government performance					
Bad opinion (first PC)	0.567	0.307	0	1	83,482
Bad opinion (average)	0.555	0.313	0	1	90,481
Linguistic Distance					
Linguistic distance^W	0.345	0.253	0	1	90,481
Individual Controls					
Age	36.65	14.48	15	105	90,481
Female	0.499	0.500	0	1	90,481
Rural	0.562	0.496	0	1	90,481
Geographic Controls					
Log geodesic distance^W	5.430	1.123	0	6.974	90,481
Absolute distance ruggedness^W	98.21	100.7	0	1223	90,481
Absolute distance CSI^W	314.8	461.4	0	6692	90,481
Absolute distance elevation^W	147.5	155.3	0	1214	90,481
Climatic Controls					
Average precipitation	83.66	43.14	3.618	330.7	90,481
Absolute distance precipitation^W	20.15	18.93	0	172.6	90,481
Average temperature	23.59	3.644	15.29	29.13	90,481
Absolute distance temperature^W	0.898	0.857	0	6.429	90,481

Notes: The sample includes individuals from 95 distinct ethnic groups in 27 African countries surveyed between 1999 and 2017. The superscript W indicates average distances between potential rebels and each ethnic group in the government coalition weighted by the role of ethnic groups in power (where senior partners receive double the weight of each junior partner).