BROWN
Orlando Bravo Center
for Economic Research

Diversity and Conflict*

Bravo Working Paper # 2020-005

‡Cemal Eren Arbatlı† Quamrul H. Ashraf‡ Oded Galor§ Marc Klemp¶

Abstract: This research advances the hypothesis and establishes empirically that interpersonal population diversity, rather than fractionalization or polarization across ethnic groups, has been pivotal to the emergence, prevalence, recurrence, and severity of intrasocietal conflicts. Exploiting an exogenous source of variations in population diversity across nations and ethnic groups, as determined predominantly during the exodus of humans from Africa tens of thousands of years ago, the study demonstrates that population diversity, and its impact on the degree of diversity within ethnic groups, has contributed significantly to the risk and intensity of historical and contemporary civil conflicts. The findings arguably reflect the contribution of population diversity to the non-cohesivnesss of society, as reflected partly in the prevalence of mistrust, the divergence in preferences for public goods and redistributive policies, and the degree of fractionalization and polarization across ethnic, linguistic, and religious groups.

Keywords: Social conflict, population diversity, ethnic fractionalization, ethnic polarization, interpersonal trust, political preferences
JEL codes: D74, N30, N40, O11, O43, Z13

---

*We thank the editor, four anonymous referees, Ran Abramitzky, Alberto Alesina, Yann Algan, Sascha Becker, Moshe Buchinsky, Matteo Cervellati, Carl-Johan Dalgaard, David de la Croix, Emilio Depetris-Chauvin, Paul Dower, Joan Esteban, James Fenske, Raquel Fern´andez, Boris Gershman, Avner Greif, Pauline Grosjean, Elhanan Helpman, Murat Iyigun, Noel Johnson, Garett Jones, Mark Koyama, Stelios Michalopoulos, Steven Nafziger, Nathan Nunn, John Nye, Omer ¨ Ozak, Elias ¨ Papaioannou, Sergey Popov, Stephen Smith, Enrico Spolaore, Uwe Sunde, Mathias Thoenig, Nico Voigtl¨ander, Joachim Voth, Romain Wacziarg, Fabian Waldinger, David Weil, Ludger W¨oßmann, Noam Yuchtman, Alexei Zakharov, and seminar participants at George Mason University, George Washington University, HSE/NES Moscow, the AEA Annual Meeting, the conference on "Deep Determinants of International Comparative Development" at Brown University, the workshop on "Income Distribution and Macroeconomics" at the NBER Summer Institute, the conference on "Culture, Diversity, and Development" at HSE/NES Moscow, the conference on "The Long Shadow of History: Mechanisms of Persistence in Economics and the Social Sciences" at LMU Munich, the fall meeting of the NBER Political Economy Program, the session on "Economic Growth" at the AEA Continuing Education Program, the workshop on "Biology and Behavior in Political Economy" at HSE Moscow, and the Economic Workshop at IDC Herzliya for valuable comments. Steven Brownstone, Gregory Casey, Ashwin Narayan, Daniel Prinz, and Jeffrey Wang provided excellent research assistance. Arbatlı acknowledges research support from the Center for Advanced Studies (CAS) at the National Research University Higher School of Economics (HSE). Ashraf acknowledges research support from the NSF (SES-1338738), the Hellman Fellows Program, and the Oakley Center for Humanities and Social Sciences at Williams College. Galor acknowledges research support from the NSF (SES-1338426) and the Population Studies and Training Center (PSTC) at Brown University. The PSTC receives core support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (5R24HD041020). Klemp acknowledges research support from the Carlsberg Foundation, the Danish Research Council (grant numbers 1329-00093 and 1327-00245), and the European Union's Horizon 2020 research and innovation programme under the Marie Sk lodowska-Curie Global Fellowship grant agreement number 753615).

†Faculty of Economic Sciences, National Research University Higher School of Economics, 20 Myasnitskaya Street, 101000, Moscow, Russia (e-mail: darbatli@hse.ru).

‡Department of Economics, Williams College, Schapiro Hall, 24 Hopkins Hall Drive, Williamstown, MA 01267 (email: qha1@williams.edu).

§Department of Economics, Brown University, Robinson Hall, 64 Waterman Street, Providence, RI 02912 (email: Oded Galor@brown.edu).

¶Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark (email: marc.klemp@econ.ku.dk).

Diversity and Conflict*

Cemal Eren Arbatlı† Quamrul H. Ashraf‡ Oded Galor§ Marc Klemp¶

August 23, 2019

Abstract

This research advances the hypothesis and establishes empirically that interpersonal population diversity, rather than fractionalization or polarization across ethnic groups, has been pivotal to the emergence, prevalence, recurrence, and severity of intrasocietal conflicts. Exploiting an exogenous source of variations in population diversity across nations and ethnic groups, as determined predominantly during the exodus of humans from Africa tens of thousands of years ago, the study demonstrates that population diversity, and its impact on the degree of diversity within ethnic groups, has contributed significantly to the risk and intensity of historical and contemporary civil conflicts. The findings arguably reflect the contribution of population diversity to the non-cohesivnesss of society, as reflected partly in the prevalence of mistrust, the divergence in preferences for public goods and redistributive policies, and the degree of fractionalization and polarization across ethnic, linguistic, and religious groups.

Keywords: Social conflict, population diversity, ethnic fractionalization, ethnic polarization, interpersonal trust, political preferences

JEL codes: D74, N30, N40, O11, O43, Z13

*We thank the editor, four anonymous referees, Ran Abramitzky, Alberto Alesina, Yann Algan, Sascha Becker, Moshe Buchinsky, Matteo Cervellati, Carl-Johan Dalgaard, David de la Croix, Emilio Depetris-Chauvin, Paul Dower, Joan Esteban, James Fenske, Raquel Fernández, Boris Gershman, Avner Greif, Pauline Grosjean, Elhanan Helpman, Murat Iyigun, Noel Johnson, Garett Jones, Mark Koyama, Stelios Michalopoulos, Steven Nafziger, Nathan Nunn, John Nye, Ömer Özak, Elias Papaioannou, Sergey Popov, Stephen Smith, Enrico Spolaore, Uwe Sunde, Mathias Thoenig, Nico Voigtländer, Joachim Voth, Romain Wacziarg, Fabian Waldinger, David Weil, Ludger Wößmann, Noam Yuchtman, Alexei Zakharov, and seminar participants at George Mason University, George Washington University, HSE/NES Moscow, the AEA Annual Meeting, the conference on "Deep Determinants of International Comparative Development" at Brown University, the workshop on "Income Distribution and Macroeconomics" at the NBER Summer Institute, the conference on "Culture, Diversity, and Development" at HSE/NES Moscow, the conference on "The Long Shadow of History: Mechanisms of Persistence in Economics and the Social Sciences" at LMU Munich, the fall meeting of the NBER Political Economy Program, the session on "Economic Growth" at the AEA Continuing Education Program, the workshop on "Biology and Behavior in Political Economy" at HSE Moscow, and the Economic Workshop at IDC Herzliya for valuable comments. Steven Brownstone, Gregory Casey, Ashwin Narayan, Daniel Prinz, and Jeffrey Wang provided excellent research assistance. Arbatlı acknowledges research support from the Center for Advanced Studies (CAS) at the National Research University Higher School of Economics (HSE). Ashraf acknowledges research support from the NSF (SES-1338738), the Hellman Fellows Program, and the Oakley Center for Humanities and Social Sciences at Williams College. Galor acknowledges research support from the NSF (SES-1338426) and the Population Studies and Training Center (PSTC) at Brown University. The PSTC receives core support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (5R24HD041020). Klemp acknowledges research support from the Carlsberg Foundation, the Danish Research Council (grant numbers 1329-00093 and 1327-00245), and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Global Fellowship grant agreement number 753615).

†Faculty of Economic Sciences, National Research University Higher School of Economics, 20 Myasnitskaya Street, 101000, Moscow, Russia (e-mail: darbatli@hse.ru).

‡Department of Economics, Williams College, Schapiro Hall, 24 Hopkins Hall Drive, Williamstown, MA 01267 (email: qha1@williams.edu).

§Department of Economics, Brown University, Robinson Hall, 64 Waterman Street, Providence, RI 02912 (email: Oded_Galor@brown.edu).

¶Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark (email: marc.klemp@econ.ku.dk).

1 Introduction

Over the course of the 20th century, in the period following World War II, civil conflicts have been responsible for more than 16 million casualties worldwide, well surpassing the cumulative loss of human life associated with international conflicts. Nations plagued by civil conflict have experienced significant fatalities from violence, substantial loss of productive resources, and considerable declines in their standards of living. While the number of countries experiencing conflict has declined from its peak in the early 1990s, as many as 35 nations have been afflicted by the prevalence of civil conflict since 2010, and more than a quarter of all nations encountered the incidence of civil conflict for at least a decade during the 1960–2017 time period.

This research explores the origins of the prevailing variation in the emergence, prevalence, recurrence, and severity of intrasocietal conflicts across countries, regions, and ethnic groups. It highlights one of their deepest roots, molded during the dawn of the dispersion of anatomically modern humans across the globe and its differential impact on the level of population diversity across regions. The study advances the hypothesis and establishes empirically that interpersonal diversity with each ethnic group, rather than fractionalization or polarization across ethnic groups, is pivotal for the understanding of civil conflicts. Exploiting an exogenous source of variations in population diversity across nations and ethnic groups, as determined predominantly during the exodus of Homo sapiens from Africa tens of thousands of years ago, the study establishes that interpersonal population diversity, and its impact on the degree of diversity within ethnic groups, has contributed significantly to conflicts in the course of human history. The study further suggests that the contribution of interpersonal population diversity to the non-cohesiveness of society, as reflected partly by the prevalence of mistrust, the divergence in preferences for public goods and redistributive policies, and the degree of fractionalization and polarization across ethnic, linguistic, and religious groups, has fostered social, political, and economic instability and magnified the vulnerability of society to internal conflicts.

Population diversity at the national or subnational level may contribute to intergroup as well as intra-group conflicts through several mechanisms. First, population diversity may have an adverse effect on the prevalence of mutual trust, and excessive diversity could therefore depress the level of social capital below a threshold that could have averted the emergence of social, political, and economic grievances and, thus, prevented violent hostilities. Second, to the extent that population diversity captures interpersonal divergence in preferences for public goods and redistributive policies, highly diverse societies may find it difficult to reconcile such differences through collective action, thereby intensifying their susceptibility to conflict. Third, insofar as population diversity reflects interpersonal heterogeneity in traits that are differentially rewarded, it can potentially cultivate resentments that are rooted in inequality, thereby magnifying the vulnerability to internal belligerence.

Moreover, the prehistorical variation in the level population diversity across regions and its potential role in facilitating the formation of ethnic groups may have contributed to the emergence of social conflicts. In particular, following the "out of Africa" migration of humans, the initial endowment of population diversity in each region may have influenced the process of group formation, reflecting the trade-off associated with the scale of the population. While a larger group may benefit from economies of scale, its productivity tends to be affected adversely by its incohesiveness. Thus, in light of the adverse impact of diversity on social cohesiveness, a larger initial endowment of population diversity have plausibly led to the emergence of a larger number of groups, and due to the forces of "cultural drift" and "biased transmission" of cultural markers (e.g., traditions, norms, and dialects), to the formation of distinct ethnic identities. The emergent fragmentation could have

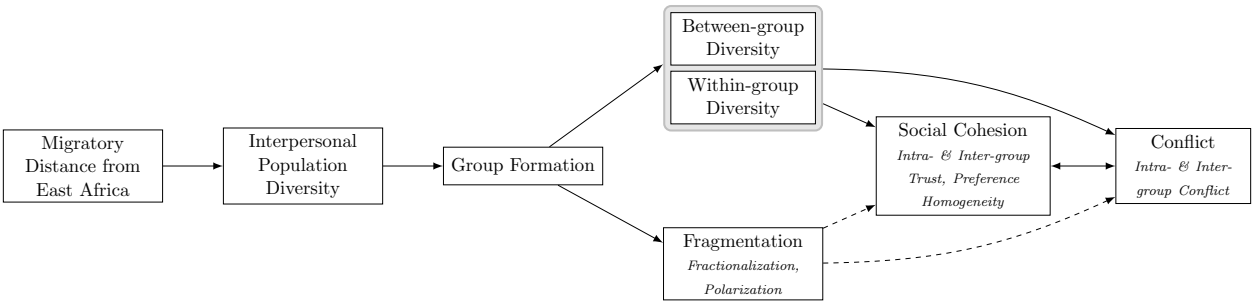


FIGURE 1: The Evolution of Population Diversity in a Location and Its Impact on Conflict

Notes: Solid arrows represent hypothesized links that are confirmed by the empirical analysis, whereas dashed arrows represent hypothesized links that do not gain consistent support. In particular, interpersonal diversity within as well as between groups affect both inter-group and intra-group conflict, partly via their adverse effect on social cohesion *within* and *across* ethnic groups.

fueled excessive inter-group competition and dissension, and could have created fertile grounds for the use of a divide-and-rule strategy by political elites, contributing to the emergence of conflict.

The exploration of the contribution of interpersonal population diversity to conflict within nations and ethnic groups relies on a novel measure that encompasses various dimensions of population diversity – proportional representation of ethnic groups, interpersonal diversity between groups, and interpersonal diversity within groups. While some aspects of population diversity at the national level can be captured by indexes of ethnolinguistic fractionalization and polarization, these measures predominantly reflect the proportional representation of ethnic groups in the population, disregarding the importance of the degree of interpersonal diversity within each ethnic group for the overall level of diversity at the national level. These deficient measures of population diversity may thus obfuscate the true impact of population diversity on civil conflicts within nations, and they do not permit the exploration of the role of diversity within an ethnic group on either intra-group or inter-group conflicts.

Exploiting variations across countries and ethnic homelands, the analysis demonstrates that interpersonal population diversity within and between ethnic groups has contributed fundamentally – as illustrated in Figure 1 – to the emergence, prevalence, recurrence, and severity of historical and contemporary intrasocietal conflicts across countries, regions, and ethnic groups. Furthermore, the country-level analysis documents that the contribution of population diversity to intrastate conflicts has plausibly operated partly via the number of ethnic groups in the population, the prevalence of mistrust, and the degree of dispersion in political preferences.

The dual analysis at the national and at the ethnic-homeland levels has several virtues. First, it permits the exploration of the impact of population diversity on the emergence of conflicts in societies of different scales, suggesting that population diversity reduces social cohesion and increases the likelihood of social conflicts within national as well as *subnational* populations. Second, since the boundaries of ethnic homelands largely predate the formation of modern nation states, the ethnic-homeland level analysis mitigates potential concerns regarding the impact of population diversity and internal conflicts on contemporary national borders (Alesina and Spolaore, 2003). Third, the focus on ethnic groups as well as on national populations permits the analysis to disentangle the impact of population diversity within an ethnic group, from the impact of ethnic diversity across groups, in the emergence of inter-group as well as intra-group conflicts. Fourth, because populations within ethnic homelands have been largely native to their locations,

the analysis at the ethnicity level diminishes potential concerns about the effect of conflicts on migrations across countries and on the global distribution of national population diversity.

The research employs several empirical strategies to mitigate concerns about the potential role of reverse causality, omitted cultural, geographical, and human characteristics, as well as sorting, in the observed association between population diversity and intrasocietal conflicts. In the course of human history, conflicts have plausibly altered the *observed* levels of diversity within ethnic groups, and the association between observed population diversity within an ethnic group and intra-group conflict may partly reflect reverse causality from conflict to diversity. Furthermore, the association between population diversity and internal conflicts at the ethnicity level may be governed by omitted cultural, geographical, and human characteristics. In order to mitigate these concerns, the empirical analysis exploits the negative association between the observed population diversity of an indigenous contemporary ethnic group and its migratory distance from East Africa, due to the serial founder effect (e.g., Harpending and Rogers, 2000; Ramachandran et al., 2005; Ashraf and Galor, 2013a), to *predict* population diversity for a globally representative sample of more than 900 ethnic groups.[1]

Nevertheless, several scenarios could a priori weaken the credibility of this methodology. First, selective migration out of Africa, or natural selection along the migratory paths, could have affected human traits and, therefore, conflict independently of the impact of migratory distance from Africa on the degree of *diversity* in human traits. However, while migratory distance from Africa has a significant negative association with the degree of diversity in human traits, it appears to be uncorrelated with the *mean* level of traits in a population, such as height, weight, and skin reflectance, conditional on distance from the equator (Ashraf and Galor, 2013a). Second, migratory distance from Africa could be correlated with distances from focal historical locations (e.g., technological frontiers) and could, therefore, capture the effect of these other distances on the process of development and the emergence of conflicts, rather than the effect of these migratory distances via population diversity. Nevertheless, conditional on migratory distance from East Africa, distances from historical technological frontiers in the years 1, 1000, and 1500 do not qualitatively alter the impact of predicted diversity on internal conflicts, further justifying the reliance on the "out of Africa" hypothesis and the serial founder effect for identifying the influence of population diversity on intrasocietal conflicts.

Moreover, a threat to identification would emerge if the actual migratory paths from Africa would have been correlated with geographical characteristics that are directly conducive to conflict (e.g., soil quality, ruggedness, climatic conditions, and propensity to trade). This would have involved, however, that the conduciveness of these geographical characteristics to conflicts would be aligned along the main root of the migratory path out of Africa as well as along each of the main forks that emerge from this primary path. In particular, in several important forks of this migration process (e.g., the Fertile Crescent and the associated eastward migration into Asia and westward migration into Europe), geographical characteristics that are conducive to conflicts would have to diminish symmetrically along these divergent secondary migratory paths. Nevertheless, the analysis establishes that the results are qualitatively unaffected when it accounts for a wide range of potentially confounding geographical characteristics of ethnic homelands, spatial dependence, as

[1]The contemporary worldwide distribution of observed population diversity across indigenous ethnic groups overwhelmingly reflects a serial founder effect – i.e., a chain of ancient population bottlenecks – originating in East Africa. In particular, because the spatial diffusion of humans to the rest of the world occurred in a stepwise migration process beginning around 90,000–60,000 BP, where in each step, a subgroup of individuals left their parental colony to establish a new settlement farther away, carrying with them only a subset of the diversity of their parental colony, the population diversity of a prehistorically indigenous ethnic group as observed today decreases with the distance along ancient human migratory paths from East Africa.

well as time-invariant unobserved heterogeneity in each region, identifying the association between interpersonal population diversity and internal conflicts across societies in the same region.

The observed association between population diversity and internal conflict at the ethnic-homeland level may further reflect the sorting of less diverse populations into geographical niches that are less conducive to conflicts. While sorting would not affect the existence of a positive association between population diversity and conflicts, it would weaken the proposed interpretation of this association. However, such sorting would require that the spatial distribution of ex-ante conflict risk would have to be negatively correlated with migratory distance from Africa and the conduciveness of geographical characteristics to conflicts would have to be negatively aligned with the primary migratory path out of Africa as well as with each of the main subsequent forks and their associated secondary migratory paths. These concerns are further mitigated by accounting for heterogeneity in a wide range of geographical characteristics across ethnic homelands, spatial autocorrelation, and regional fixed effects.

Further, to the extent that interregional migration flows in the post-1500 era, and thus the proportional representation of ethnic groups within each national population, may have been affected by historically persistent spatial patterns of conflict risk, contemporary national population diversity may be endogenous to intrastate conflicts. Thus, to mitigate these concerns two alternative empirical strategies are developed, yielding remarkably similar results. The first strategy confines the analysis to variations in a sample of countries that only belong to the Old World (i.e., Africa, Europe, and Asia), where diversity of contemporary national populations predominantly reflects the diversity of indigenous populations that became native to their current locations well before the colonial era. This strategy rests on the observation that post-1500 population movements within the Old World did not result in the significant admixture of populations that were very distant from one another. The second strategy exploits variations in a globally representative sample of countries using an estimator, in which the migratory distance of a country's prehistorically native population from East Africa is employed as an instrumental variable for the diversity of its contemporary national population. It rests on the identifying assumption that the migratory distance of a country's prehistorically native population from East Africa is exogenous to the risk of intrastate conflict faced by the country's overall population in the last half-century.

The empirical analysis at the country level establishes that, accounting for the potentially confounding effects of geographical and institutional characteristics, ethnolinguistic fragmentation, outcomes of economic development, and continent fixed effects, an increase in national population diversity that corresponds to the movement from the 10th to the 90th percentile of its global cross-country distribution (i.e., a movement from the diversity level of the Republic of Korea to that of the Democratic Republic of Congo) is associated with 2.3 new civil conflict outbreaks during the 1960–2017 time horizon (relative to a sample mean of 1.2 and a standard deviation of 1.7 new civil conflict outbreaks). In addition, this increase in diversity is also associated with (i) an increase in the likelihood of observing the incidence of civil conflict in any given 5-year interval during the 1960–2017 period from 18 percent to 34 percent; (ii) an increase in the likelihood of observing the onset of a new civil conflict in any given year during the 1960–2017 time horizon from 1 percent to 4 percent; (iii) an increase in the likelihood of observing the incidence of one or more intra-group factional conflict events in any given year during the 1985–2006 time horizon from 6 percent to 60 percent; and (iv) an increase in the intensity of social unrest by either 26 percent or 38 percent of a standard deviation of the observed distribution of intrastate conflict severity across countries in the post-1960 time period (depending on the employed measure of intrastate conflict severity).

Similarly, the analysis at the ethnic-homeland level establishes that, accounting for the potentially confounding influence of a wide range of geographical and historical factors, outcomes of economic development, and regional fixed effects, an increase in observed population diversity

of an ethnic group from the 10th percentile (e.g., the Mamusi people of Oceania) to the 90th percentile (e.g., the Pare people of Eastern Africa) of its global distribution is associated with an increase in the prevalence of conflicts within the territory of a homeland over the years 1989–2008 by 0.43 (relative to a sample mean of 0.14 and a standard deviation of 0.27). Further, this change in ethnic population diversity is also associated with an increase of about 57 conflict events, 9,731 conflict-related deaths, and 924 deaths per conflict during the same time period.

2 Related Literature

This study is related to several well-established lines of inquiry. Specifically, the paper contributes to the vast literature on the determinants of civil conflict. The determinants of civil conflict have been the focus of intensive research over the past two decades, highlighting the role of social, political, and economic grievances, along with the capability of the state to subdue armed opposition groups, the conduciveness of geographical characteristics towards rebel insurgencies, and the opportunity cost of engaging in rebellions, among other contributing factors (Sambanis, 2002; Fearon and Laitin, 2003; Collier and Hoeffler, 2007; Blattman and Miguel, 2010). The present study advances the understanding of the nature of grievance-related mechanisms in civil conflict, emphasizing the role of interpersonal population diversity and its deep determinants on the emergence of intra-group as well as inter-group social divisions.

The role of *fractionalization* was initially at the forefront of empirical analyses of the underlying determinants of civil conflict, in light of the conventional wisdom that inter-group competition over ownership of productive resources and political power, along with conflicting preferences for public goods and redistributive policies, are more difficult to reconcile in societies that are fragmented ethnolinguistically. Nevertheless, early evidence regarding the influence of ethnic, linguistic, and religious fractionalization on the risk of civil conflict in society had been largely inconclusive (Fearon and Laitin, 2003; Collier and Hoeffler, 2007), arguably due in part to conceptual limitations associated with fractionalization indices. The introduction of *polarization* indices to the analyses of civil conflict has led to more affirmative findings demonstrating that inter-group grievances are indeed contributors to the risk of civil conflict in society (Montalvo and Reynal-Querol, 2005; Esteban et al., 2012).[2]

Nevertheless, while measures of ethnolinguistic fragmentation are unable to account for the potentially critical role of intra-group heterogeneity in augmenting the risk of conflict in society at large, a central virtue of the proposed measure of population diversity is that it captures the impact of diversity across individuals *within* ethnic groups. Furthermore, even as a proxy for *interethnic* divisions, the proposed measure generates substantial insights relative to existing proxies that are based on fractionalization and polarization indices. Specifically, the commonly used measures of ethnolinguistic fragmentation typically do not exploit information beyond the proportional representations of ethnolinguistically differentiated groups in the national population – namely, they implicitly assume that these ethnic groups are internally homogenous and culturally "equidistant" from one another.[3] In contrast, the proposed measure of national population diversity

[2]However, in network-based models of conflict involving multiple groups (e.g., König et al., 2017), greater inter-group divergence could mitigate conflict propensity by reducing the strength of inter-group network alliances within one side or another of such conflicts.

[3]More sophisticated measures of ethnolinguistic fragmentation – such as (i) the Greenberg index of "cultural diversity," as measured by Fearon (2003) and Desmet et al. (2009), or (ii) the ethnolinguistic polarization index, as measured by Desmet et al. (2009) and by Esteban et al. (2012) – incorporate information on pairwise linguistic distances, wherein pairwise linguistic proximity monotonically increases in the number of shared branches between any two languages in a hierarchical linguistic tree. This information, however, is constrained by the nature of a hierarchical

incorporates information on pairwise inter-group genetic distances, as well as the genetic diversity within each ethnic group, as determined predominantly over the course of the "out of Africa" demic diffusion of humans to the rest of the globe tens of thousands of years ago.[4]

Moreover, the use of conventional measures of ethnolinguistic fragmentation in the exploration of the impact of fragmentation on conflict is unsatisfactory due to plausible concerns about reverse causality and measurement error. Due to the association of conflict with atrocities as well as voluntary and forced migrations, the degree of ethnolinguistic fractionalization is likely to be affected by past and potential conflicts. Although the proposed measure of population diversity exploits information on the population shares of subnational groups possessing ethnically differentiated ancestries, because the endowment of population diversity in a given location was overwhelmingly determined during the prehistoric "out of Africa" expansion of humans, the analysis is able to exploit a plausibly exogenous source of the contemporary cross-country variation in this measure, thereby mitigating the biases associated with measurement and endogeneity issues that plague the widely used proxies of ethnolinguistic fragmentation. Furthermore, in contrast to the plausibly exogenous component of population diversity, the degree of ethnolinguistic fragmentation may be systematically mismeasured in more conflict-prone societies, due to (i) the political economy of national census categorizations of subnational groups, and (ii) the endogenous constructivism of individual self-identification with an ethnic group (Eifert et al., 2010; Caselli and Coleman, 2013; Besley and Reynal-Querol, 2014).

The present study also contributes to a vast literature that explores the impact of ethnolinguistic fragmentation and interethnic economic inequality on other societal outcomes, including the rate of economic growth, the quality of national institutions, the extent of financial development, the efficiency in the provision of public goods, and the level of social capital (Easterly and Levine, 1997; Alesina and La Ferrara, 2005; Alesina et al., 2016). In particular, since population diversity encompasses the degree of heterogeneity within each ethnic group as well as the pairwise distances amongst them, the current analysis is uniquely positioned to capture the contribution of these additional dimensions of diversity to social dissonance and aggregate inefficiency.

Furthermore, in light of the view that the contemporary variation in population diversity across the globe predominantly reflects the human expansion out of Africa tens of thousands of years ago, the paper contributes to the exploration of the role of deeply rooted human characteristics in comparative economic development. In particular, the study contributes to the understanding of the importance of inter-personal population diversity for social outcomes in the course of human history (e.g., population density, urbanization, and income) as explored by Ashraf and Galor (2013a).[5]

Finally, the study is consistent with the primordialist theories of conflict, maintaining that ethnic conflict springs from differences in ethnic identity, as well as with the instrumentalist theories, suggesting that ethnic conflict may emerge for pragmatic reasons (e.g., inequality, security, and competition).[6] In particular, since the initial endowment of interpersonal population diversity at a

linguistic tree, where languages residing at the same level of branching of the tree are necessarily equidistant from one another.

[4]The genetic distance between any two ethnic groups in a contemporary national population predominantly reflects the prehistoric migratory distance between their respective ancestral populations (from the precolonial era), and as follows from the continuity of geographical distances, the proposed population diversity measure captures *continuous* inter-group distances. Spolaore and Wacziarg (2016) documents a negative relationship between genetic distance and interstate warfare. They argue that if genetic relatedness proxies for unobserved similarity in preferences over rival and excludable goods, then conflict over the control of such resources would be more likely to arise between nations that are genetically closer to one another.

[5]The importance of prehistorically determined human characteristics is further explored by Spolaore and Wacziarg (2013) and Ashraf and Galor (2013b, 2018).

[6]In addition, the modernist viewpoint (Bates, 1983; Gellner, 1983; Wimmer, 2002) stresses that interethnic conflict arises from increased competition over scarce resources, especially when previously marginalized groups that were

given location may have facilitated the endogenous formation of groups, whose collective identities diverged over time under the forces of cultural drift, a reduced-form link between the prehistorically determined diversity and the contemporary risk of interethnic conflict may well be apparent in the data, regardless of whether these groups are mobilized into conflict by primordial or instrumentalist reasons.

3 Population Diversity and Conflict at the Country Level

3.1 Empirical Framework and Strategy

This section describes the various layers of the country-level analyses of the influence of population diversity on intrastate conflicts, the key variables employed, and the strategies implemented to identify the impact of population diversity on conflict.

The analysis initially focuses on contemporary conflicts, exploiting variations in either cross-country or repeated cross-country data. It explores the explanatory power of interpersonal population diversity for (i) the average frequency of new conflict outbreaks, (ii) the persistence of conflicts, as captured by the likelihood of conflict prevalence, and (iii) the likelihood of conflict outbreak. It then analyzes the impact of interpersonal diversity on *intra-group* factional conflicts within a national population. Finally, it explores the influence of interpersonal diversity on conflicts in the distant past.

Following the convention in the civil conflict literature, the contemporary analysis is confined to the post-1960 time period, when most of the European colonies in Sub-Saharan Africa, the Middle East, and South and Southeast Asia had already gained independence. This time horizon thus permits an assessment of the correlates of civil conflict at the national level, independently of their interactions with the contemporaneous influence of the colonial powers. The baseline sample for the contemporary analysis contains information on 150 countries for the 1960–2017 time period, of which 123 are in the Old World.

3.1.1 Main Outcome Variables: Frequency, Incidence, and Onset of Civil Conflict

The main outcome variable in the cross-country regressions is the average number of new civil conflicts per annum during the 1960–2017 time period. It is based on conflict events listed in version 18.1 of the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002; Pettersson and Eck, 2018). In this data set, a civil conflict is defined as an armed conflict between the government of a state and internal opposition groups over a given incompatibility. Recurrent episodes of the same conflict between state actors and armed opposition groups are not treated as new conflicts. The study employs the most comprehensive armed conflict coding (PRIO25), encompassing all civil conflict events that resulted in at least 25 battle-related deaths in a given year.

The country-level analysis additionally exploits the temporal dimension of armed conflict events, examining the *incidence* of PRIO25 civil conflicts in a repeated cross-section of countries. In this analysis, the outcome variable is an indicator, coded 1 for each country-period (a period being a 5-year time interval) in which at least one active PRIO25 civil conflict is observed, and 0 otherwise. The study also examines the predictive power of population diversity for the *onset* of new PRIO25 civil conflicts in annually repeated cross-country data. This variable is coded 1 for each year in which at least one new PRIO25 civil conflict had erupted, and 0 otherwise. Moreover, outbreaks of subsequent episodes of the same conflict are not considered new conflict onsets.

excluded from the nation-building process experience socioeconomic modernization and, thus, begin to challenge the status quo.

3.1.2 Population Diversity: Measurement and Identification Strategy

The interpersonal population diversity of each country is captured by the measure of predicted genetic diversity developed by Ashraf and Galor (2013a). It is based on (i) the proportional representation of each of the ancestral populations of a contemporary nation, (ii) the genetic diversity of each of these ancestral populations, as predicted by its migratory distance from Africa, and (iii) the pairwise genetic distances between each pair of these ancestral populations, as predicted by their migratory distances from one another.

Observed genetic diversity at the ethnic group level is measured by an index referred to by population geneticists as expected heterozygosity. This index reflects the probability that two individuals, selected at random from the relevant population, are different from one another with respect to a given spectrum of genetic traits. The index is constructed by population geneticists using data on allelic frequencies (i.e., the frequency with which a gene variant or allele occurs in a given population).[7] Expected heterozygosity, H_{exp}, takes the form:

$$H_{exp} = 1 - \frac{1}{m}\sum_{l=1}^{m}\sum_{i=1}^{k_l} p_i^2,$$

where m is the number of genes or DNA loci in the sample, k_l is observed variants or alleles of gene l, and p_i denotes the frequency of occurrence of the ith allele.

Population geneticists have computed this index of expected heterozygosity, along with pairwise genetic distances, for a sample of 53 globally representative ethnic groups from the Human Genome Diversity Cell Line Panel.[8] These ethnic groups have been not only prehistorically native to their current geographical locations but also largely isolated from genetic flows from other ethnic groups. The index is constructed using data on allelic frequencies for a particular class of DNA loci called microsattelites, residing in non-protein-coding or "neutral" regions of the human genome – i.e., regions that do not directly result in phenotypic expression. Thus, this measure of observed genetic diversity has the advantage of not being tainted by the differential forces of natural selection that may have operated on these populations since their prehistoric exodus from Africa.

Nevertheless, like measures of ethnolinguistic fragmentation based on fractionalization or polarization indices, *observed* genetic diversity might be endogenous to civil conflict, since it could be tainted by genetic admixtures resulting from the movement of populations across space, triggered by cross-regional differences in patterns of historical conflict potential, the nature of political institutions, and levels of economic prosperity. To circumvent this concern, the analysis is based on the measure of *predicted* genetic diversity introduced by Ashraf and Galor (2013a). Exploiting the explanatory power of a serial founder effect associated with the "out of Africa" migration process, the diversity of a country's prehistorically indigenous population is predicted by the coefficients obtained from an ethnic-group-level regression of expected heterozygosity on migratory distance from Addis Ababa in the aforementioned sample comprising 53 globally representative ethnic groups from the Human Genome Diversity Cell Line Panel. This measure captures the component of observed interpersonal diversity within a country's indigenous ethnic groups that is predicted by migratory distance from Addis Ababa to the country's modern-day capital city, along prehistoric land-connected human migration routes.[9]

[7] See Ashraf and Galor (2018).

[8] The Human Genome Diversity Cell Line is compiled by the Human Genome Diversity Project (HGDP) in collaboration with the Centre d'Etudes du Polymorphisme Humain (CEPH).

[9] Consistent with the serial founder effects associated with the prehistoric "out of Africa" migration process, expected heterozygosity in microsattelites declines with migratory distance from East Africa across ethnic groups.

In the absence of systematic and large-scale population movements across geographically (and, thus, genetically) distant regions, as had been largely true during the precolonial era, the interpersonal diversity of the prehistorically native population in a given location serves as a good proxy for the contemporary population diversity of that location. While this continues to remain true to a large extent for nations in the Old World (i.e., Africa, Europe, and Asia), post-1500 population flows from the Old World to the New World have had a considerable impact on the ethnic composition and, thus, the contemporary interpersonal diversity of national populations in the Americas and Oceania. Thus, instead of employing the interpersonal diversity of prehistorically native populations (i.e., precolonial diversity) at the expense of limiting the analysis to the Old World, the measure of ancestry-adjusted genetic diversity from Ashraf and Galor (2013a) is employed as the main proxy for contemporary population diversity. Using the shares of different groups in a country's modern-day population, this measure accounts for (i) the diversity within the ethnic groups that can trace own ancestry around year 1500 to their current homelands, (ii) the diversity of those descended from immigrant settlers over the past half-millennium, and (iii) the additional component of population diversity at the national level that arises from the pairwise genetic distances amongst these different subnational groups.[10]

However, ancestry-adjusted population diversity may still be afflicted by endogeneity bias because it accounts for the impact of cross-country migrations in the post-1500 era on the diversity of contemporary national populations. In particular, these migrations may have been spurred by historically persistent spatial patterns of conflict. Two alternative strategies are implemented to address this issue. The first strategy is to exploit variations across countries that only belong to the Old World, where as discussed previously, the interpersonal diversity of contemporary national populations overwhelmingly reflects the diversity within populations that have been native to their current locations since well before the colonial era. This strategy is based on the view that the great human migrations of the post-1500 era had systematically differential impacts on the genetic composition of national populations in the Old World versus the New World. Specifically, although post-1500 population flows had a dramatic effect on the interpersonal diversity of national populations in the Americas and Oceania, the diversity of populations in Africa, Europe, and Asia remained largely unaltered, primarily because native populations in the Old World were not subjected to substantial inflows of migrant that were descended from genetically distant ancestral populations. By confining the analysis to the Old World, this strategy effectively exploits the spatial variation in contemporary population diversity that largely coincides with the variation in diversity of prehistorically indigenous populations, as determined overwhelmingly by an ancient serial founder effect associated with the "out of Africa" migration process.

The second strategy employs the migratory distance of the prehistorically native populations in each country from East Africa as an instrument for the country's contemporary population diversity. This strategy utilizes the observation that the mark of ancient population bottlenecks that occurred during the prehistoric "out of Africa" demic diffusion of humans across the globe

Mounting evidence from the fields of physical and cognitive anthropology, surveyed in Ashraf and Galor (2018), additionally reflect the influence of serial founder effects on various forms of intra-group phenotypic and cognitive diversity, including phonemic diversity and interpersonal diversity in skeletal features pertaining to cranial characteristics, dental attributes, and pelvic traits. Thus, the association of heterozygosity in neutral genetic markers with socioeconomic outcomes may plausibly reflect the influence of diversity in various observed and unobserved phenotypic characteristics.

[10]The data on the population shares of these different subnational groups at the country level are obtained from the World Migration Matrix, 1500–2000 of Putterman and Weil (2010), who compile for each country in their data set, the share of the country's population in 2000 that is descended from the population of every other country in 1500. For an in-depth discussion of the methodology underlying the construction of the ancestry-adjusted measure of genetic diversity, the reader is referred to the data appendix of Ashraf and Galor (2013a).

continues to be seen in the worldwide pattern of genetic diversity across contemporary national populations, as reflected by the sizable correlation of 0.75 between the proxies for precolonial and contemporary population diversity in a global sample of countries. This strategy rests on the identifying assumption that the migratory distance of a country's prehistorically indigenous population from East Africa has no direct effect on the potential for civil conflict faced by its modern national population, conditional on a large set of controls for the geographical and institutional determinants of conflict as well as the correlates of economic development.

3.1.3 Confounding Characteristics

The vast empirical literature on civil conflict has considered a large number of contributing factors. Drawing on this literature, a wide range of control variables are included in the baseline specifications. The discussion below describes these potential confounders. Additional control variables used in robustness checks are discussed in corresponding Appendices.[11]

Geographical Characteristics The study accounts for a wide range of geographical attributes that may be correlated with prehistoric migratory distance from East Africa and can influence conflict risk through channels unrelated to population diversity. Absolute latitude and distance to the nearest waterway, for instance, can exert an influence on economic development and, thus, on conflict potential through climatological, institutional, and trade-related mechanisms.

Rugged terrains can provide safe havens for rebels and enable them to sustain continued resistance by protecting them from superior government forces (Fearon and Laitin, 2003). Moreover, in regions with rough terrains, subgroups of a regional population may be geographically more isolated. Such isolation may strengthen the forces of "cultural drift" and ethnic differentiation among these groups (Michalopoulos, 2012), thus increasing the potential for inter-group conflict. Further, in light of evidence that conditional on their respective country-level means, greater intracountry dispersion in agricultural land suitability and elevation can contribute to ethnolinguistic diversity (Michalopoulos, 2012), these natural attributes could also generate an indirect influence on conflict propensity through the ethnolinguistic fragmentation of the population.[12] To account for these factors, the baseline analysis controls for terrain ruggedness, as well as the mean and range of both agricultural land suitability and elevation.

The baseline specifications also include a dummy for island nations. Due to their greater isolation in space, islands nations possibly followed different historical trajectories than nations that are connected by land to one another. For example, the settlement process that took place in island nations and their relative immunity from cross-border spillovers may influence both population diversity and conflict potential. Finally, the baseline specifications additionally account for a complete set of continent fixed effects to ensure that the estimated reduced-form impact of population diversity on conflict potential is not simply reflecting the latent influence of unobserved time-invariant cultural, institutional, and geographical factors at the continent level.

Institutional Factors Colonial legacies may have significantly shaped the political economy of interethnic cleavages in newly independent states (Posner, 2003). More generally, the heritage of colonial rule and the identity of the former colonizers may have important ramifications for the nature and stability of contemporary political institutions at the national level, thereby influencing

[11] The definitions and data sources of all variables employed by the analysis at the country level are listed in Section A.4 of the Supplemental Material.

[12] Although these measures of ethnolinguistic fragmentation are directly accounted for, their exogenous geographical determinants may still explain some unobserved component of intrapopulation heterogeneity in ethnic and cultural traits, thereby exerting some influence on the potential for conflict in society.

the potential for conflict in society. Two different sets of covariates are included in the baseline specifications to account for the impact of colonial legacies. Depending on the unit of analysis, the first set comprises either binary indicators for the historical prevalence of colonial rule (as is the case in the cross-country regressions) or time-varying measures of the lagged prevalence of colonial rule (as is the case in the regressions using repeated cross-country data). In either case, a distinction is made between colonial rule by the U.K., France, and any other major colonizing power. The second set of covariates comprises time-invariant binary indicators for British and French legal origins, included to account for any latent influence of legal codes and institutions that may not necessarily be captured by colonial experience.

The baseline specifications additionally include three control variables, all based on yearly data at the country level from the Polity IV Project, in order to account for the direct influence of contemporary political institutions on the risk of civil conflict. The first variable is based on an ordinal index that reflects the degree of executive constraints in any given year, whereas the other two variables are based on binary indicators for the type of political regime, reflecting the prevalence of either democracy (when the polity score is above 5) or autocracy (when the polity score is below -5) in a given year.[13]

Ethnolinguistic Fragmentation Previous empirical findings regarding the role of ethnic fragmentation in civil conflicts have been somewhat mixed, exhibiting substantial sensitivity to model specifications and conflict codings (Fearon and Laitin, 2003). Moreover, theoretical work on the link between the ethnic composition of a society and the risk of civil conflict suggests that ethnic fractionalization by itself may be insufficient to fully capture the conflict potential that can be attributed to broader ethnolinguistic configurations of the population (Esteban and Ray, 2011a). In light of their well-grounded structural foundations, indices of polarization have gained popularity as a substitute for – or in addition to – the fractionalization measures commonly considered by empirical analyses of civil conflict. Indeed, many empirical studies find that ethnic polarization is a stronger predictor of the likelihood of civil conflict (e.g., Montalvo and Reynal-Querol, 2005; Esteban et al., 2012).

Two time-invariant controls are thus included in the baseline specifications to capture the influence of the ethnolinguistic composition of national populations on the potential for civil conflict. The first proxy is the well-known ethnic fractionalization index of Alesina et al. (2003), reflecting the probability that two individuals, randomly selected from a country's population, will belong to different ethnic groups. The second proxy for this channel is an index of ethnolinguistic polarization, obtained from the data set of Desmet et al. (2012). The authors provide measures of several such polarization indices, constructed at different levels of aggregation of linguistic groups in a country's population (based on hierarchical linguistic trees). The specific polarization measure employed here corresponds to the most disaggregated level of the linguistic tree and reflects the extent of polarization across subnational groups classified according to modern-day languages.[14]

Natural Resources and Development Outcomes Natural resources can foster the risk of civil conflict by weakening political institutions and facilitating state capture, easing the financial constraints on rebel organizations (e.g., Fearon and Laitin, 2003; Dube and Vargas, 2013; Collier and Hoeffler, 2007), increasing the vulnerability of political elites to terms-of-trade shocks (e.g.,

[13]The prevalence of anocracy, occurring when the polity score is between -5 and 5, therefore serves as the omitted political regime category.

[14]The choice of Desmet et al. (2012) as the data source for ethnolinguistic polarization is primarily due to the more comprehensive geographical coverage of their data set, relative to other potential data sources such as Montalvo and Reynal-Querol (2005) or Esteban et al. (2012).

Humphreys, 2005), and raising the return to regional secession (e.g., Ross, 2006). The baseline specifications thus include an indicator for the presence of oil or gas reserves.

Average living standards can influence civil conflict potential in a country through several channels. One argument, due to Grossman (1991) and Hirshleifer (1995), is that higher per-capita incomes raise the opportunity cost for potential rebels to engage in insurrections, thus predicting an inverse relationship between the level or growth rate of income, on the one hand, and the risk of civil conflict, on the other (Miguel et al., 2004; Collier and Hoeffler, 2007). Another argument, due to Hirshleifer (1991) and Grossman (1999), is that by raising the return to predation, higher per-capita incomes can contribute to the risk of rapacious activities over society's resources, consistently with empirical findings from some of the aforementioned studies on the link between income from natural resources and conflict potential. Furthermore, to the extent that income per capita serves as a proxy for state capabilities (Fearon and Laitin, 2003), a higher level of per-capita income can reflect the notion of a state that is better able to prevent or defend itself against rebel insurgencies; an idea that has also found some recent empirical support (e.g., Bazzi and Blattman, 2014). Therefore, the baseline specifications control for GDP per capita, as reported by the World Bank's World Development Indicators (WDI). Importantly, because population diversity, as proxied by genetic diversity, has been shown to confer a hump-shaped influence on productivity at the country level (Ashraf and Galor, 2013a), the inclusion of GDP per capita accounts for the indirect effect of population diversity on conflict potential via the income channel.

Like income per capita, population size is also a standard covariate in empirical models of conflict. One reason is that operational definitions of civil conflict typically impose a death threshold, and violence-related casualties may be mechanically related to the size of population. In addition, a larger population may imply a greater recruitment pool for rebels (Fearon and Laitin, 2003). Further, to the extent that more populous countries exhibit greater intrapopulation heterogeneity, they could also harbor stronger motives for secessionist conflicts (Alesina and Spolaore, 2003; Desmet et al., 2011). The baseline specifications thus include controls for population size.

It should be noted that many of the aforementioned controls for institutional quality, ethnolinguistic fragmentation, and the correlates of economic development are endogenous in an empirical model of civil conflict, and as such, their estimated coefficients in the regressions do not permit a causal interpretation. Nonetheless, controlling for these factors is essential to minimize specification errors and assess the extent to which the reduced-form influence of population diversity on conflict potential can be attributed to more conventional explanations in the literature.

Appendix A.4 presents the summary statistics of all the main variables exploited by the baseline cross-country analysis of civil conflict frequency.

3.2 Empirical Results

This section presents the main findings from several country-level analyses, establishing a highly significant and robust reduced-form causal influence of population diversity on various intrastate conflict outcomes over the past half-century. The exposition commences with the results of the baseline cross-country regressions that explain the annual frequency of civil conflict outbreaks in the post-1960 time period. It then discusses the results from conflict incidence and onset regressions that exploit variations in repeated cross-country data, before presenting evidence that population diversity has also been a significant predictor of contemporary intra-group conflict outcomes. The section concludes with an analysis of conflicts during the 1400–1799 period, showing that population diversity has had a deep influence on the conflict potential of societies over many centuries. The analysis of each conflict outcome includes several robustness checks. Some of these are collected

and discussed in Appendix A.2 while others are relegated to Sections A.1–A.2 of the Supplemental Material.

3.2.1 Analysis of Civil Conflict Frequency in Cross-Country Data

The cross-country regressions attempt to explain the variation across countries in the annual frequency of new civil conflict onsets – i.e., the average number of new PRIO25 civil conflict eruptions per year – during the 1960–2017 time horizon. Specifically, the baseline empirical model for the cross-country analysis is as follows.

$$CF_i = \beta_0 + \beta_1 \widehat{DIV_i} + \beta_2' GEO_i + \beta_3' ETH_i + \beta_4' INS_i + \beta_5' DEV_i + \varepsilon_i, \tag{1}$$

where CF_i is the (log-transformed) average number of new PRIO25 civil conflict outbreaks per year in country i; $\widehat{DIV_i}$ is the ancestry-adjusted population diversity of the national population; GEO_i, ETH_i, INS_i, and DEV_i are the respective vectors of control variables for geographical characteristics (including continent dummies), ethnolinguistic fragmentation, institutional factors, and the correlates of economic development, as described in Section 3.1; and finally, ε_i is a country-specific disturbance term. All time-varying controls for institutional factors and development outcomes enter the model as their respective temporal means over the 1960–2017 time horizon.

Table I presents the results from the baseline cross-country analysis. The analysis begins with a bivariate regression in Column 1, showing that population diversity is indeed a positive and highly significant correlate of the annual frequency of new civil conflict eruptions. Specifically, the estimated coefficient suggests that a move from the 10th to the 90th percentile of the cross-country distribution of population diversity is associated with an increase in conflict frequency by 0.014 new civil conflict outbreaks per year, a relationship that is statistically significant at the 1 percent level. Bearing in mind that the sample mean of the dependent variable is 0.022 outbreaks per year, this association is also of sizable economic significance, reflecting 44 percent of a standard deviation across countries in the temporal frequency of new civil conflict onsets. Next, beginning with Column 2, the analysis progressively includes an expanding set of covariates to the specification. It first incorporates exogenous geographical characteristics and then additionally accounts for measures of ethnolinguistic fragmentation, before controlling for semi-endogenous institutional factors and more endogenous outcomes of economic development in the full empirical model in Column 8.

Upon accounting for the potentially confounding influence of geographical characteristics in Column 2, population diversity continues to remain statistically significant at the 1 percent level, but now, its coefficient is more than twice as large as the unconditioned estimate from Column 1. This increase appears to be largely driven by the inclusion of absolute latitude and the range of elevation and of land suitability as covariates to the model, as all three variables enter the regression significantly and with expected signs.[15] Based on the specification in Column 2, the scatter plots in Figure 2 depict the positive and statistically significant cross-country relationship between population diversity and the annual frequency of new civil conflict onsets, both in the full sample of countries and in a sample that omits apparently influential outliers.

As revealed by the regression in Column 3, the point estimate of the impact of population diversity on conflict becomes somewhat diminished once the specification is conditioned to only

[15] Specifically, countries located farther from the equator have seen fewer conflict outbreaks on average, while those with greater dispersion in their respective land endowments have experienced such outbreaks more frequently, a result that plausibly reflects the conflict-promoting role of ethnolinguistic fragmentation, following the rationale provided by the findings of Michalopoulos (2012).

TABLE I: Population Diversity and the Frequency of Civil Conflict Onset across Countries – The Baseline Analysis

Cross-country sample:	Global								Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) 2SLS	(12) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017											
Population diversity (ancestry adjusted)	0.209***	0.439***	0.306***	0.290**	0.326***	0.318***		0.309**	0.548***	0.597***	0.537***	0.602***
	(0.066)	(0.104)	(0.115)	(0.113)	(0.118)	(0.119)		(0.130)	(0.191)	(0.209)	(0.176)	(0.185)
Within-group population diversity							0.364***					
							(0.140)					
Between-group population diversity							0.284*					
							(0.166)					
Ethnic fractionalization				0.011		0.004	0.004	0.001		0.002		−0.005
				(0.012)		(0.013)	(0.013)	(0.010)		(0.012)		(0.010)
Ethnolinguistic polarization					0.016	0.014	0.014	0.012		0.016		0.020*
					(0.011)	(0.011)	(0.012)	(0.012)		(0.014)		(0.012)
Absolute latitude		−0.307**	−0.396*	−0.294	−0.435**	−0.392	−0.391	0.166	−0.319	0.289	−0.477**	−0.046
		(0.124)	(0.204)	(0.249)	(0.199)	(0.244)	(0.245)	(0.242)	(0.255)	(0.305)	(0.201)	(0.243)
Ruggedness		0.015	−0.005	−0.001	0.000	0.001	0.003	0.031	0.002	0.048	−0.001	0.028
		(0.030)	(0.035)	(0.036)	(0.035)	(0.036)	(0.036)	(0.036)	(0.040)	(0.041)	(0.034)	(0.033)
Mean elevation		−0.019**	−0.018*	−0.018*	−0.019*	−0.019*	−0.020*	−0.020**	−0.023**	−0.023**	−0.019**	−0.021**
		(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.012)	(0.011)	(0.009)	(0.009)
Range of elevation		0.011***	0.012***	0.011***	0.011***	0.011***	0.011***	0.004	0.014***	0.004	0.012***	0.005*
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)	(0.004)	(0.003)
Mean land suitability		0.014	0.020	0.023	0.024*	0.024*	0.025	0.001	0.018	0.000	0.021*	−0.000
		(0.012)	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)	(0.014)	(0.016)	(0.017)	(0.012)	(0.013)
Range of land suitability		0.014*	0.014	0.013	0.017*	0.017	0.017	0.008	0.017	0.007	0.017*	0.011
		(0.008)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.012)	(0.012)	(0.015)	(0.010)	(0.012)
Distance to nearest waterway		0.007	0.006	0.005	0.006	0.006	0.006	0.003	0.005	0.005	0.005	0.002
		(0.010)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)
Island nation dummy		−0.012	−0.015**	−0.015**	−0.015**	−0.015**	−0.015**	−0.021**	−0.008	−0.021*	−0.015**	−0.022***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.010)	(0.011)	(0.007)	(0.008)
Executive constraints, 1960–2017 average								−0.002		−0.003		−0.000
								(0.004)		(0.005)		(0.004)
Fraction of years under democracy, 1960–2017								0.017		0.023		0.013
								(0.018)		(0.019)		(0.017)
Fraction of years under autocracy, 1960–2017								−0.009		−0.010		−0.010
								(0.015)		(0.016)		(0.014)
Oil or gas reserve discovery								0.008*		0.007		0.007
								(0.005)		(0.005)		(0.005)
Log population, 1960–2017 average								0.005**		0.007**		0.005**
								(0.003)		(0.003)		(0.002)
Log GDP per capita, 1960–2017 average								−0.010***		−0.009***		−0.010***
								(0.002)		(0.003)		(0.002)
Continent dummies			×	×	×	×	×	×	×	×	×	×
Legal origin dummies								×		×		×
Colonial history dummies								×		×		×
Observations	150	150	150	150	150	150	150	147	123	121	150	147
Partial R^2 of population diversity		0.128	0.044	0.040	0.050	0.046		0.051	0.068	0.088		
Partial R^2 of within-group							0.042					
Partial R^2 of between-group							0.015					
Adjusted R^2	0.029	0.189	0.213	0.212	0.220	0.215	0.212	0.358	0.225	0.392		
Effect of 10th–90th %ile move in diversity	0.014***	0.029***	0.020***	0.019**	0.022***	0.021***		0.021**	0.026***	0.026***	0.036***	0.041***
	(0.004)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)		(0.009)	(0.009)	(0.009)	(0.012)	(0.013)
Effect of 10th–90th %ile move in within-group							0.037***					
							(0.014)					
Effect of 10th–90th %ile move in between-group							0.023*					
							(0.013)					
FIRST STAGE											Population diversity (ancestry adjusted)	
Migratory distance from East Africa (in 10,000 km)											−0.068***	−0.065***
											(0.005)	(0.007)
First-stage F statistic											153.543	92.693

Notes: This table exploits cross-country variations to establish a significant positive reduced-form impact of contemporary population diversity on the annual frequency of new PRIO25 civil conflict onsets during the 1960–2017 time period, conditional on ethnic diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. For regressions based on the global sample, the set of continent dummies includes five indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes two indicators for Africa and Asia. The set of legal origin dummies includes two indicators for British and French legal origins, and the set of colonial history dummies includes three indicators for experience as a colony of the U.K., France, and any other major colonizing power. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

exploit intra-continental cross-country variations. However, even after including a complete set of continent dummies, the coefficient of interest remains statistically significant at the 1 percent level and larger than the unconditioned estimate from Column 1. It suggests that a move from the 10th

to the 90th percentile of the cross-country distribution of population diversity is associated with an increase in conflict frequency by 0.020 civil conflict outbreaks per year, corresponding to 65 percent of a standard deviation of the cross-country conflict frequency distribution.

The regressions in Columns 4–6 indicate that when additionally subjected to controls for ethnic fractionalization and ethnolinguistic polarization, either individually or jointly, the point estimate of the coefficient on population diversity continues to remain largely stable in both magnitude and statistical precision.[16] In contrast, neither ethnic fractionalization nor ethnolinguistic polarization appears to possess any significant explanatory power for the cross-country variation in the temporal frequency of civil conflict outbreaks, conditional on population diversity and the baseline set of geographical covariates.[17]

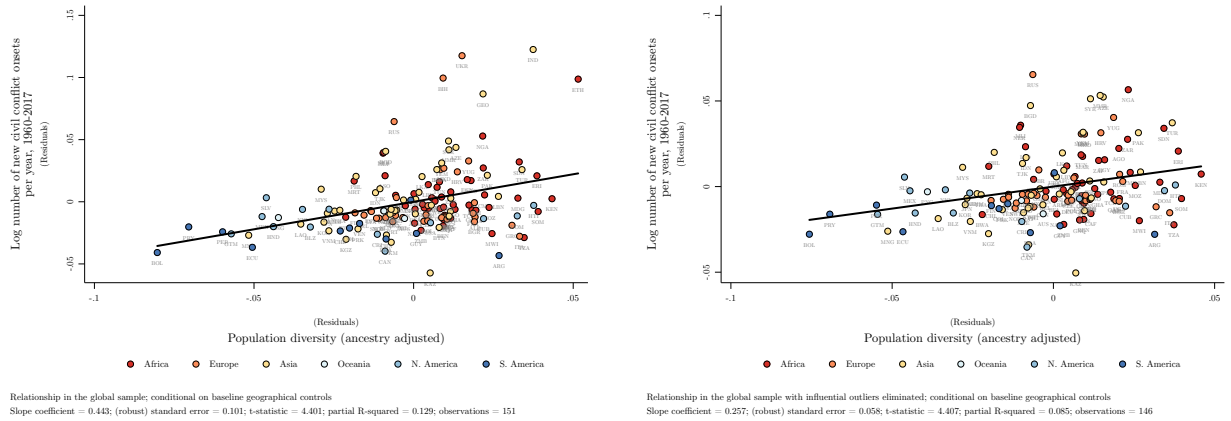
The analysis in Column 7 replicates the specification from Column 6 except that it decomposes the measure of overall interpersonal diversity of the national population into its two components and jointly examines their conditional associations with conflict. The two components of overall diversity capture the average interpersonal diversity *within* versus *between* groups in the contemporary national population, where the subnational groups are categorized by their ancestral origins prior to the great intercontinental migrations of the post-1500 era.[18] The results indicate that the within-group component of population diversity is economically and statistically more important for explaining civil conflict. Specifically, a move from the 10th to the 90th percentile of the cross-country distribution of within-group diversity is associated with an increase in conflict frequency by 0.037 civil conflict outbreaks per year, a relationship that is statistically significant at the 1 percent level. On the other hand, a similar move along the cross-country distribution of between-group diversity is associated with a less pronounced increase of 0.023 new civil conflict onsets per year. The estimated response in the latter case is also statistically less precise, reflecting statistical significance only at the 10 percent level. The greater importance of the within-group component of population diversity is additionally reflected by a corresponding partial R^2 statistic that is nearly 3 times as large as that associated with the between-group component.

The full specification in Column 8 augments the intermediate specification from Column 6 with controls for colonial legacy and contemporary institutional factors, as well as controls for the natural resource curse, population size, and GDP per capita. Reassuringly, regardless of the potential endogeneity of these additional covariates, the point estimate of the coefficient on population diversity remains remarkably stable in both magnitude and statistical significance in comparison to the estimates from previous columns. In particular, the coefficient of interest from this regression suggests that conditional on the complete set of controls for geographical characteristics, ethnolinguistic fragmentation, institutional factors, and outcomes of economic development, a move from the 10th to the 90th percentile of the cross-country distribution of population diversity is associated with an increase in conflict frequency by 0.021 new PRIO25 civil conflict outbreaks

[16] By restricting both fractionalization and polarization measures to enter the regressions linearly, the current approach follows Esteban et al. (2012). Nevertheless, a robustness check of the main finding to employing alternative specifications that allow for both a linear and a quadratic term in ethnic fractionalization yielded qualitatively similar results (not reported).

[17] The analysis in Table SA.IX in Section A.1 of the Supplemental Material shows that although the two measures of ethnolinguistic fragmentation do independently possess some explanatory power for the temporal frequency of conflict onsets after accounting for geographical confounders, these conditional relationships are not statistically robust to the inclusion of continent dummies to the specifications.

[18] Thus, for a given contemporary national population, the within-group component of overall diversity reflects the weighted average group-level interpersonal diversity, using the population shares of these subnational groups as weights, whereas the between-group component reflects the residual fraction of overall diversity that is unexplained by the within-group component. The latter component therefore corresponds to an aggregate measure of intergroup distances amongst all subnational groups in the national population.



(a) Full sample

(b) Outliers omitted

FIGURE 2: Population Diversity and the Frequency of Civil Conflict Onset across Countries

Notes: This figure depicts the global cross-country relationship between contemporary population diversity and the annual frequency of new PRIO25 civil conflict onsets during the 1960–2017 time period, conditional on the baseline geographical correlates of conflict, as considered by the specification in Column 2 of Table I. The relationship is depicted for either an unrestricted sample of countries (Panel (a)) or a sample that omits apparently influential outliers (Panel (b)). Each of the two panels presents an added-variable plot with a partial regression line. Given that the unrestricted sample employed by the left panel is not constrained by the availability of data on other covariates considered by the analysis in Table I, the regression coefficients reported in this panel are marginally different from those presented in Column 2 of Table I. The set of influential outliers omitted from the sample in Panel (b) includes Bosnia and Herzegovina (BIH), Ethiopia (ETH), Georgia (GEO), India (IND), and Ukraine (UKR).

per year, or 68 percent of a standard deviation of the cross-country conflict frequency distribution. Moreover, the adjusted R^2 statistic of the regression suggests that the full empirical model explains about 36 percent of the cross-country variation in conflict frequency, whereas the partial R^2 statistic associated with population diversity indicates that 5 percent of the residual cross-country variation in conflict frequency can be explained by the residual cross-country variation in population diversity.

Addressing Endogeneity The results thus far demonstrate a highly significant and robust cross-country association between population diversity and the temporal frequency of civil conflict onsets over the last half-century, even after conditioning the analysis on a sizable set of controls for geographical characteristics, ethnolinguistic fragmentation, institutional factors, and development outcomes. Nevertheless, this association could be marred by endogeneity bias, in light of the possibility that the large-scale human migrations of the post-1500 era – as captured by the ancestry-adjusted measure of interpersonal diversity for contemporary national populations – and the spatial pattern of conflicts in the modern era could be codetermined by common unobserved forces (e.g., the spatial pattern of *historical* conflicts) that may not be fully accounted for by covariates. Although the stability of the coefficient of interest across specifications suggests that selection on unobservables needs to be unreasonably strong to fully explain away the main finding, one cannot rely entirely on OLS point estimates to assess causality.[19] Thus, as discussed previously in Section 3.1, the analysis exploits two alternative identification strategies to address this issue. The specifications in Columns 9–10 implement the first approach to causal identification by simply restricting the OLS estimator to exploit variations in a subsample of countries that only belong to the Old World. Then, in Columns 11–12, the analysis conducts 2SLS regressions that employ the migratory distance of the prehistorically native population in each country from East Africa as

[19]For a more formal analysis of selection on observables and unobservables, see Appendix A.2.

an instrument for the country's contemporary population diversity. The identifying assumption is that migratory distance from East Africa is plausibly exogenous to the risk of civil conflict in the post-1960 time period, conditional on the sizable vector of control variables.

As is evident from the regressions in Columns 9–12, the two alternative identification strategies yield remarkably similar results, with the point estimate of the coefficient on population diversity being noticeably larger in magnitude, relative to its less well-identified counterpart in the global-sample OLS regressions (from either Column 3 or Column 8). In particular, the coefficient is highly statistically significant across the four better-identified specifications, and as estimated by the 2SLS regression in Column 12, it suggests that a move from the 10th to the 90th percentile of the global cross-country distribution of population diversity is associated with an increase in conflict frequency by 0.041 new PRIO25 civil conflict outbreaks per year, corresponding to 133 percent of a standard deviation of the global cross-country conflict frequency distribution.

There are plausibly three distinct rationales – perhaps operating in tandem – for why the better-identified point estimates of the coefficient on population diversity are larger than their less well-identified counterparts. First, the spatial pattern of social conflict may exhibit long-term persistence for reasons other than population diversity. If persistent conflict spurred emigrations and atrocities that gradually led to systematically more homogeneous populations (Fletcher and Iyigun, 2010) in conflict-prone areas, there should be a downward bias in the estimated coefficient on population diversity in an OLS regression that explains the global variation in civil conflict potential in the modern era.

A second plausible explanation is that the pattern of conflict risk in the modern era, especially across populations in the New World that experienced a substantial increase in diversity from migrations in the post-1500 era, has been influenced not so much by the higher population diversity of the immigrants but more so by the unobserved (or observed but noisily measured) human capital that European settlers brought with them, the colonization strategies that they pursued, and the socio-political institutions that they established. To the extent that these unobserved factors associated with European settlers in the New World served, in one way or another, to reduce the risk of social conflict in the modern national populations of the Americas and Oceania, they could also introduce a negative bias in the OLS estimates of the relationship between population diversity and conflict risk in a global sample of countries.

A third possible rationale is that in the end, population diversity explains the conflict propensity of a population mostly through its prehistorically determined component. This component may have contributed to the formation and ethnic differentiation of *native* groups in a given location and, thus, to more deeply rooted inter-ethnic divisions amongst these groups. As such, conditional on continent fixed effects that absorb any systematic differences in the pattern of post-1500 population flows into locations in the Old World versus the New World, the ancestry-adjusted measure of interpersonal diversity – which incorporates the diversity of both native and non-native groups in a contemporary national population – might be a noisy proxy for the "true" measure of prehistorically determined population diversity. Therefore, as a result of this "measurement error," the influence of the ancestry-adjusted measure of population diversity might be attenuated in an OLS regression that exploits worldwide variations.

Given that both of the identification strategies ultimately exploit the variation in population diversity across populations that have been prehistorically indigenous to their current locations, either by omitting the modern national populations of the New World from the estimation sample or by instrumenting contemporary population diversity in a globally representative sample of countries with the prehistoric migratory distance of a country's geographical location from East Africa, the better-identified estimates mitigate all the aforementioned sources of negative bias.

Robustness Checks The analysis in Appendix A.2 shows that population diversity possesses significant power for explaining the cross-country variation in the total count of new conflict onsets during the 1960–2017 time period (Table A.II). It also establishes the robustness of the baseline cross-country findings to accounting for: spatial dependence across observations by estimating spatial regressions (Table A.III); and the property of population diversity as a generated regressor by bootstrapping the standard errors (Table A.IV).

Further, Section A.1 of the Supplemental Material presents several robustness checks for the cross-country analysis of the influence of population diversity on the temporal frequency of civil conflict outbreaks in the post-1960 time horizon. It demonstrates that the main findings are qualitatively robust to **(1)** accounting for various ecological and climatic covariates, including the temporal means and volatilities of annual temperature and precipitation over the relevant sample period as well as time-invariant measures of ecological fractionalization and polarization (Table SA.I); **(2)** accounting for the timing of the Neolithic Revolution, state antiquity, the duration of human settlement, and distance from the regional technological frontier in 1500 (Table SA.II); **(3)** accounting for inequality across ethnic homelands as well as overall spatial inequality in nighttime luminosity within a country (Table SA.III); **(4)** accounting for linguistic rather than ethnic fractionalization as a baseline covariate (Table SA.IV); **(5)** accounting for alternative measures of ethnolinguistic fractionalization and polarization, based on the spatial distribution of language homelands and on gridded population data (Table SA.V); **(6)** accounting for the initial-year values of time-varying baseline covariates rather than their temporal means over the sample period (Table SA.VI); **(7)** accounting for spatial autocorrelation in unobserved heterogeneity (Table SA.VII); and **(8)** the elimination of world regions from the estimation sample that could have been statistically influential for generating the key empirical pattern (Table SA.VIII).

3.2.2 Analysis of Civil Conflict Incidence in Repeated Cross-Country Data

The analysis now proceeds to examine the temporal prevalence of civil conflict. Specifically, exploiting the time structure of quinquennially repeated cross-country data, it investigates the predictive power of population diversity for the likelihood of observing the incidence of one or more active conflict episodes in a given 5-year interval during the 1960–2017 time horizon. The following probit model is therefore estimated using maximum-likelihood estimation.

$$\begin{aligned} CP^*_{i,t} = \gamma_0 + \gamma_1 \widehat{DIV}_i + \gamma_2' GEO_i + \gamma_3' INS_{i,t-1} + \gamma_4' ETH_i + \gamma_5' DEV_{i,t-1} + \gamma_6 C_{i,t-1} \\ + \gamma_7' \delta_t + \eta_{i,t} \equiv \gamma' Z_{i,t} + \eta_{i,t}; \end{aligned} \tag{2}$$

$$C_{i,t} = \mathbf{1}\left[CP^*_{i,t} \geq D^*\right]; \tag{3}$$

$$Pr\left(C_{i,t} = 1 | Z_{i,t}\right) = Pr\left(CP^*_{i,t} \geq D^* | Z_{i,t}\right) = \Phi\left(\gamma' Z_{i,t} - D^*\right), \tag{4}$$

where $CP^*_{i,t}$ is a latent variable measuring the potential for an active conflict episode in country i during any given 5-year interval, t, and it is modeled as a linear function of explanatory variables. In particular, the time-invariant explanatory variables $\widehat{DIV}_i$, GEO_i, and ETH_i are all as previously defined, but now, the time-varying covariates included in $INS_{i,t-1}$ and $DEV_{i,t-1}$ enter as their respective temporal means over the previous 5-year interval. Further, δ_t is a vector of time-interval (5-year period) dummies, and $\eta_{i,t}$ is a country-period-specific disturbance term.[20] By

[20]The robustness of the current analysis of conflict incidence to exploiting variations in annually (rather than quinquennially) repeated cross-country data is confirmed in Appendix A.2. Naturally, in those regressions, the time-dependent covariates enter as their lagged annual values (instead of their lagged 5-year temporal means) and time fixed effects are captured by a set of year dummies.

specifying each of the time-varying controls to enter the model with a one-period lag, the analysis aims to mitigate the concern that the use of contemporaneous measures of these covariates may exacerbate reverse-causality bias in their estimated coefficients.[21] Finally, the model assumes that contemporary conflict potential additionally depends on the lagged incidence of civil conflict, $C_{i,t-1}$, which accounts for the possibility that countries with a conflict experience in the immediate past may exhibit a higher conflict potential in the current period, mainly because of the intertemporal spillovers that are common to most conflict processes – e.g., the self-reinforcing nature of past casualties on either side of a conflict.[22] Because the continuous variable reflecting conflict potential, $CP^{*}_{i,t}$, is unobserved, its level can only be inferred from the binary incidence variable, $C_{i,t}$, indicating whether the latent conflict potential was sufficiently intense for the annual battle-related death threshold of a civil conflict episode to have been surpassed during a given 5-year interval. As is evident from equations (3)-(4), D^{*} is the corresponding threshold for unobserved conflict potential, and it appears as an intercept in $\Phi(.)$, the cumulative distribution function for the disturbance term, $\eta_{i,t}$.

The main results for the temporal prevalence (or incidence) of PRIO25 civil conflict episodes are presented in Columns 1–4 of Table II. In the interest of brevity, the analysis exclusively reports the better-identified point estimates – namely, from probit regressions in a sample of countries belonging only to the Old World, and from IV probit regressions that exploit migratory distance from East Africa as an instrument for contemporary population diversity in a global sample of countries. For each of these two identification strategies, two distinct specifications are estimated; one that partials out the influence of only exogenous geographical covariates (including continent fixed effects), and another that conditions the analysis on the full set of control variables from the empirical model of conflict incidence.

As is evident from the results, interpersonal population diversity enters all four specifications with a positive and highly significant coefficient. To interpret the coefficient of interest, the IV probit regression presented in Column 4 suggests that conditional on the complete set of control variables, a 1 percentage point increase in population diversity leads to an increase in the quinquennial likelihood of a PRIO25 civil conflict incidence by 2.6 percentage points. Indeed, this sample-wide average marginal effect of population diversity is statistically significant at the 1 percent level. In addition, the economic significance of population diversity for conflict incidence is evident in the plots presented in Figure SA.1 in Section A.3 of the Supplemental Material. Based on the regressions in Columns 2 and 4, these plots illustrate how the *predicted* quinquennial likelihood of a civil conflict incidence varies as one moves along the cross-country distribution of population diversity in the relevant estimation sample. Specifically, a move from the 10th to the 90th percentile of the cross-country distribution of population diversity leads to an increase in the predicted quinquennial likelihood of civil conflict incidence from about 23 to 33 percent amongst countries in the Old World, and from about 18 to 34 percent in the global sample of countries.

[21] An alternative method to address the reverse-causality problem, in the context of quinquennially repeated cross-country data, would have been to control for time-dependent covariates as measured in the initial year of each 5-year interval. Although this method would retain the first-period observation for each country, which is dropped under the current specification, it leaves open the possibility that the presence or absence of an active conflict in the first year of each period may still exert a direct influence on the time-varying controls.

[22] In adopting this strategy, the current analysis of conflict incidence follows Esteban et al. (2012). It may also be noted here that because the measure of population diversity is time-invariant (as is the case with all known measures of ethnolinguistic fragmentation, based on fractionalization or polarization indices), the analysis is unable to either account for country fixed effects or exploit dynamic panel estimation methods, despite the time dimension of the repeated cross-country data. In all regressions exploiting such data, however, the robust standard errors of the estimated coefficients are always clustered at the country level.

TABLE II: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data

Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	13.366***	12.203***	14.304***	13.578***	6.172**	6.356**	7.066***	8.804***
	(3.700)	(3.787)	(3.652)	(4.210)	(2.576)	(2.645)	(2.594)	(3.170)
Ethnic fractionalization		−0.399		−0.519		−0.084		−0.322
		(0.353)		(0.332)		(0.252)		(0.280)
Ethnolinguistic polarization		0.049		0.322		0.172		0.334
		(0.344)		(0.340)		(0.248)		(0.254)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	1,270	1,045	1,583	1,311	5,452	4,377	6,996	5,757
Countries	123	121	150	147	123	121	150	147
Pseudo R^2	0.416	0.440			0.131	0.161		
Marginal effect of diversity	2.553***	2.261***	2.817***	2.595***	0.324**	0.332**	0.336**	0.421**
	(0.683)	(0.709)	(0.741)	(0.850)	(0.139)	(0.140)	(0.133)	(0.170)
FIRST STAGE			Population diversity (ancestry adjusted)				Population diversity (ancestry adjusted)	
Migratory distance from East Africa (in 10,000 km)			−0.068***	−0.066***			−0.068***	−0.066***
			(0.006)	(0.006)			(0.006)	(0.006)
First-stage F statistic			145.394	99.876			151.502	102.614

Notes: This table exploits variations in repeated cross-country data to establish a significant positive reduced-form impact of contemporary population diversity on the likelihood of observing (i) the incidence of a PRIO25 civil conflict in any given 5-year interval during the 1960–2017 time period (Columns 1–4); and (ii) the onset of a new PRIO25 civil conflict in any given year during the 1960–2017 time period (Columns 5–8), conditional on ethnic diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. The controls for geography include absolute latitude, ruggedness, distance to the nearest waterway, the mean and range of agricultural suitability, the mean and range of elevation, and an indicator for small island nations. The controls for institutions include a set of legal origin dummies, comprising two indicators for British and French legal origins, as well as six time-dependent covariates, comprising the degree of executive constraints, two indicators for the type of political regime (democracy and autocracy), and three indicators for experience as a colony of the U.K., France, and any other major colonizing power. The control for oil presence is a time-invariant indicator for the discovery of a petroleum (oil or gas) reserve by the year 2003. The controls for population and income are the time-dependent log-transformed values of total population and GDP per capita. In Columns 1–4, all time-dependent covariates assume their average annual values over the previous 5-year interval, whereas in Columns 5–8, they assume their annual values from the previous year. To account for duration dependence and temporal spillovers in conflict outcomes, all regressions control for the lagged incidence of conflict, and the regressions in Columns 5–8 additionally control for a set of cubic splines of the number of peace years. For regressions based on the global sample, the set of continent dummies includes five indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes two indicators for Africa and Asia. The IV probit regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. The estimated marginal effect of a 1 percentage point increase in population diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in either the quinquennial likelihood of a conflict incidence (Columns 1–4) or the annual likelihood of a conflict onset (Columns 5–8), both expressed in percentage points. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Robustness Checks The analysis in Appendix A.2 shows that the influence of population diversity on the incidence or prevalence of conflict is robust to: considering alternative definitions and types of intrastate conflict as the outcome variable, such as the prevalence of large-scale civil conflicts (i.e., "civil wars") and of intrastate conflicts involving only non-state actors (Table A.V); exploiting variations in annually rather than quinquennially repeated cross-country data (Table A.VI, Columns 1–4); and considering an alternative coding of conflict prevalence that captures the share of years with an active civil conflict in a given 5-year interval (Table A.VI, Columns 5–8).

Further, Section A.2 of the Supplemental Material demonstrates that the main findings for the impact of population diversity on civil conflict incidence are qualitatively insensitive to **(1)** accounting for time-invariant fractionalization and polarization indices of ecological diversity as well as time-varying climatic covariates, including the inter-annual means and volatilities of temperature and precipitation over the previous 5-year interval (Table SA.X, Columns 1–4); **(2)** accounting for various deep-rooted correlates of long-run economic development, such as the depth of state antiquity, the time elapsed since the Neolithic Revolution, the duration of human settlement, and distance to the year-1500 technological frontier (Table SA.XI, Columns 1–4); **(3)** accounting for inequality in nighttime luminosity across gridded space and across ethnic homelands within a country (Table SA.XII, Columns 1–4); **(4)** accounting for alternative distributional indices of intergroup diversity (Alesina et al., 2003; Fearon, 2003; Esteban et al., 2012) and for additional time-invariant geographical and historical correlates of conflict incidence potential, including the percentage of mountainous terrain, the presence of noncontiguous subnational territories, and the intensity of the disease environment (Table SA.XIII); **(5)** empirically modeling conflict prevalence using either classical logit or "rare events" logit (King and Zeng, 2001) estimators, in lieu of the standard probit estimator (Table SA.XIV, Columns 1–4); and **(6)** allowing for spatiotemporal dependence across country-time observations by exploiting two-dimensional clustering of standard errors (Table SA.XV, Columns 1–4).

Finally, akin to the current analysis of conflict prevalence, the analysis in Appendix A.1 exploits variations in quinquennially repeated cross-country data to establish interpersonal population diversity as a significant predictor of the *intensity* of social conflicts. In particular, it examines both ordinal and continuous measures that capture the "severity" of intrastate conflicts and of events related to general social unrest, including but not limited to armed conflict.

3.2.3 Analysis of Civil Conflict Onset in Repeated Cross-Country Data

This section examines the onset of civil conflict. Unlike the model of conflict incidence, the onset model focuses solely on explaining the outbreak of conflict events, classifying the subsequent years into which a given conflict persists as nonevent years (akin to civil peace), unless they coincide with the eruption of a "new" conflict.[23] Conceptually, this analysis assesses the extent to which population diversity at the national level influences socio-political instability by *triggering* conflicts, rather than contributing to their perpetuation over time. The probit model for the analysis of conflict onset is similar to the one for conflict incidence, as described by equations (2)-(4), but with two notable exceptions. Specifically, following the convention in the literature, the model (i) exploits variations in annually repeated cross-country data, with the binary outcome variable assuming a value of 1 if a country-year observation coincides with the first year of a new civil conflict, and 0 otherwise; and (ii) controls for a set of cubic splines in the number of preceding years of uninterrupted peace, along with year dummies, in order to account for temporal or duration dependence (Beck et al., 1998). To mitigate issues of causal identification of the influence of population diversity on conflict onset, the analysis implements the same two strategies followed by the preceding analyses of conflict frequency and conflict incidence.

The main results for the onset of new PRIO25 civil conflicts are presented in Columns 5–8 of Table II. Irrespective of the identification strategy employed, or the set of covariates considered by the specification, population diversity appears to confer a statistically significant and robust positive influence on the annual likelihood of new civil conflict outbreaks. To elucidate the economic

[23] A "new" civil conflict in a country is defined as one involving a previously unobserved set of actors and/or a previously unobserved set of contentious issues.

significance of this impact in the global sample of countries, the sample-wide average marginal effect estimated by the specification in Column 8 suggests that conditional on the complete set of control variables, a 1 percentage point increase in population diversity leads to an increase in the annual likelihood of a new PRIO25 civil conflict outbreak by 0.4 percentage points. Further, based on the regressions in Columns 6 and 8, the plots presented in Figure SA.2 in Section A.3 of the Supplemental Material depict how the *predicted* annual likelihood of a new conflict onset responds as one moves along the cross-country distribution of population diversity in the relevant estimation sample. For instance, in response to a move from the 10th to the 90th percentile of the cross-country distribution of population diversity, the predicted annual likelihood of a new conflict onset rises from about 1.9 to 3.4 percent in the Old-World sample of countries, and from about 1.1 to 3.6 percent amongst countries worldwide.

Robustness Checks Section A.2 of the Supplemental Material demonstrates that the main findings regarding the impact of population diversity on civil conflict onset remain qualitatively unaffected after **(1)** accounting for time-invariant fractionalization and polarization indices of ecological diversity as well as time-varying climatic covariates, including the lagged annual values of temperature and precipitation and their inter-annual volatilities over the previous 5 years (Table SA.X, Columns 5–8); **(2)** accounting for various deep-rooted correlates of long-run economic development, such as the depth of state antiquity, the time elapsed since the Neolithic Revolution, the duration of human settlement, and distance to the year-1500 technological frontier (Table SA.XI, Columns 5–8); **(3)** accounting for inequality in nighttime luminosity across gridded space and across ethnic homelands within a country (Table SA.XII, Columns 5–8); **(4)** empirically modeling conflict onset using either classical logit or "rare events" logit (King and Zeng, 2001) estimators, in lieu of the standard probit estimator (Table SA.XIV, Columns 5–8); **(5)** allowing for spatiotemporal dependence across country-year observations by exploiting two-dimensional clustering of standard errors (Table SA.XV, Columns 5–8); **(6)** accounting for the influence of additional correlates of conflict onset potential, including the time-invariant "ethnic dominance" indicator of Collier and Hoeffler (2004) and the time-varying "political instability" and "new state" indicators of Fearon and Laitin (2003) (Table SA.XVI); and **(7)** accounting for the contemporaneous and lagged values of annual price shocks to various export commodities, as studied by Bazzi and Blattman (2014) (Table SA.XVII).

3.2.4 Analyses of Intra-group Conflict Incidence in Cross-Country and Repeated Cross-Country Data

One crucial dimension in which the advanced measure of population diversity adds value beyond standard indices of ethnolinguistic fragmentation is that it incorporates information on interpersonal heterogeneity not only across group boundaries but within such boundaries as well. Thus, in contrast to standard measures of ethnolinguistic fragmentation at the national level, to the extent that interpersonal diversity can be expected to give rise to social, political, and economic grievances that culminate to violent contentions even *within* ethnically or linguistically homogeneous groups, the measure is naturally better-suited to empirically link population diversity with intra-group conflicts in society. The analysis in this section provides evidence that supports this important aspect of the advanced measure, exploiting cross-country variations to establish a positive link between population diversity and the incidence of intra-group conflict events during the 1985–2006 time period.

The primary source of the exploited data on the incidence of intra-group conflict events across the globe is the All Minorities at Risk (AMAR) Phase 1 Sample Data (Birnir et al., 2018). The AMAR Sample Data is a single integrated data set, combining information on 291 subnational

groups originally included in the Minorities at Risk (MAR) project with information on 74 new groups randomly selected from the AMAR Sample Frame of "socially relevant" subnational groups, in order to correct for potential selection issues in the original MAR data (Phases I–V). A "socially relevant" subnational group is defined as an ethnic group (majority or minority) that satisfies five criteria outlined and discussed in (Birnir et al., 2015).[24] For each subnational group in the AMAR Sample Data, the data set provides information on whether the group experienced one or more intra-group conflicts in each year during the 1985–2006 time horizon.

Table III presents two distinct analyses of intra-group conflict incidence. The outcome variable for the cross-country analysis (Panel A) is the share of group-years in a given country with at least one active intra-group conflict over the sample period. For the analysis based on annually repeated cross-country data (Panel B), the outcome is a binary variable that reflects whether any of the AMAR groups within a given country experienced one or more intra-group conflicts in a given year. Depending on the identification strategy from earlier sections (i.e., restricting the estimation sample to countries in the Old World versus exploiting migratory distance from East Africa as an excluded instrument for population diversity in a global sample of countries), the analysis employs either OLS or 2SLS estimators in Panel A, and either probit or IV probit estimators in Panel B. For each outcome variable, and for each of the two identification strategies, three alternative specifications are estimated. The first two of these follow from the methodology in previous sections, in that one conditions the analysis on only exogenous geographical covariates (including continent fixed effects), whereas the other partials out the influence of the full set of controls for geographical characteristics, ethnolinguistic fragmentation, institutional factors, and development outcomes. However, to account for the possibility that the AMAR groups in a given country may not be fully representative of all its subnational groups, the third specification augments the full model with additional controls for the total number and the total share of all AMAR groups in the national population. Finally, in line with the methodology from earlier sections, all time-varying controls for institutional factors and development outcomes enter the specifications in Panel A as their respective temporal means over the 1985–2006 time period, whereas in Panel B, these covariates assume their respective lagged annual values.

The results in Table III indicate that regardless of the outcome variable examined, the set of covariates considered, or the identification strategy employed, population diversity has contributed substantially to the risk of intra-group conflicts during the 1985–2006 time period. This impact is not only highly statistically significant but considerable in terms of economic significance as well. For instance, the regression in Column 5 of Panel A suggests that conditional on the full set of baseline controls, a move from the 10th to the 90th percentile of the global cross-country distribution of population diversity is associated with an increase of 38 percentage points in the likelihood that an AMAR group of a country will have experienced an intra-group conflict at some point during the 1985–2006 time horizon. Moreover, as estimated by the regression in Column 5 of Panel B, a 1 percentage point increase in population diversity leads to an increase in the likelihood of an intra-group conflict incidence in any given country-year during this time horizon by 10 percentage points. Based on the regressions in Columns 2 and 5 of Panel B, the plots presented in Figure SA.3 in Section A.3 of the Supplemental Material illustrate the *predicted* annual likelihood of an intra-

[24]These criteria are as follows: (1) Membership in the group is determined primarily by descent by both members and non-members; (2) Membership in the group is recognized and viewed as important by members and/or non-members, where importance may be psychological, normative, and/or strategic; (3) Members share some distinguishing cultural features, such as common language, religion, occupational niche, and customs with respect to other groups in the country; (4) One or more of these cultural features are either practiced by a majority of the group or preserved and studied by a set of members who are broadly respected by the wider membership for so doing; and (5) The group has at least 100,000 members or constitutes one percent of the national population.

Table III: Population Diversity and Intra-group Conflict

Cross-country sample:	Old World			Global		
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A	OLS	OLS	OLS	2SLS	2SLS	2SLS
	Share of AMAR group-years with intra-group conflict, 1985–2006					
Population diversity (ancestry adjusted)	4.456**	4.267**	3.580**	5.728***	5.606***	5.124***
	(1.692)	(1.711)	(1.694)	(1.761)	(1.879)	(1.894)
Continent dummies	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×
Controls for ethnic diversity		×	×		×	×
Controls for institutions		×	×		×	×
Controls for oil, population, and income		×	×		×	×
Controls for number/share of AMAR groups			×			×
Observations	91	91	91	115	115	115
Partial R^2 of population diversity	0.079	0.068	0.051			
Adjusted R^2	0.092	0.187	0.231			
Effect of 10th–90th %ile move in diversity	0.218***	0.209**	0.175**	0.392***	0.384***	0.351***
	(0.083)	(0.084)	(0.083)	(0.121)	(0.129)	(0.130)
FIRST STAGE				Population diversity (ancestry adjusted)		
Migratory distance from East Africa (in 10,000 km)				−0.061***	−0.057***	−0.057***
				(0.007)	(0.008)	(0.008)
First-stage F statistic				83.366	47.887	47.107
PANEL B	Probit	Probit	Probit	IV Probit	IV Probit	IV Probit
	Annual AMAR intra-group conflict incidence, 1985–2006					
Population diversity (ancestry adjusted)	25.350***	37.535***	31.687***	31.929***	40.579***	38.375***
	(9.336)	(9.792)	(10.547)	(7.335)	(7.261)	(7.973)
Controls as in same column of Panel A	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×
Observations	1,658	1,658	1,658	2,179	2,179	2,179
Countries	90	90	90	114	114	114
Pseudo R^2	0.207	0.338	0.390			
Marginal effect of diversity	7.378***	9.107***	7.067***	8.717***	10.318***	9.402***
	(2.528)	(2.301)	(2.428)	(1.992)	(2.008)	(2.212)
FIRST STAGE				Population diversity (ancestry adjusted)		
Migratory distance from East Africa (in 10,000 km)				−0.061***	−0.060***	−0.060***
				(0.007)	(0.007)	(0.007)
First-stage F statistic				74.527	66.911	66.939

Notes: This table exploits variations across countries and years to establish a significant positive reduced-form impact of contemporary population diversity on (i) the share of group-years of a country during the 1985–2006 time period in which an "all minorities at risk" (AMAR) ethnic group of the country experienced an intra-group conflict (Panel A); and (ii) the likelihood of observing the incidence of an intra-group conflict across a country's AMAR ethnic groups in any given year during the 1985–2006 time period (Panel B), conditional on ethnic diversity measures, the proximate geographical, institutional, and development-related correlates of conflict, and measures capturing the number and total share of AMAR groups in the national population. The controls for geography include absolute latitude, ruggedness, distance to the nearest waterway, the mean and range of agricultural suitability, the mean and range of elevation, and an indicator for small island nations. The controls for ethnic diversity include ethnic fractionalization and polarization. The controls for institutions include a set of legal origin dummies, comprising two indicators for British and French legal origins, as well as six time-dependent covariates, comprising the degree of executive constraints, two indicators for the type of political regime (democracy and autocracy), and three indicators for experience as a colony of the U.K., France, and any other major colonizing power. The control for oil presence is a time-invariant indicator for the discovery of a petroleum (oil or gas) reserve by the year 2003. The controls for population and income are the time-dependent log-transformed values of total population and GDP per capita. In Panel A, all time-dependent covariates assume their average annual values over the entire 1985–2006 time period, whereas in Panel B, they assume their annual values from the previous year. For regressions based on the global sample, the set of continent dummies includes five indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes two indicators for Africa and Asia. The 2SLS and IV probit regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. In Panel A, the estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the share of group-years of a country in which an intra-group conflict was experienced by an AMAR ethnic group. In Panel B, the estimated marginal effect of a 1 percentage point increase in population diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the percentage-point increase in the annual likelihood of an intra-group conflict incidence. Heteroskedasticity-robust standard errors (clustered at the country level in Panel B) are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

group conflict incidence as a function of the percentile of the cross-country distribution of population diversity in the relevant estimation sample. Specifically, a move from the 10th to the 90th percentile of this distribution is predicted to raise the annual likelihood of an intra-group conflict incidence from about 13 to 55 percent amongst countries in the Old World, and from about 9 to 62 percent in the global sample of countries.

3.2.5 Analysis of Historical Conflict Outcomes in Cross-Country Data

The analysis has thus far been confined to examining intrastate conflict events in the last half-century. This restriction permitted it to focus on the post-independence time period of the former European colonies, exploit better quality data and codings for intrastate conflict events, and employ time-varying controls for institutional and development outcomes, as is standard in civil conflict regressions. However, there is no a priori reason why the conflict-promoting role of population diversity should not extend to the distant past.

This section investigates whether population diversity predicts historical conflict events in a cross-section of countries. Specifically, the analysis exploits information on the locations of violent conflicts during the 1400–1799 time period, as compiled by Brecke (1999) and geocoded by Dincecco et al. (2015), employing the geocoding of conflict locations to map these historical conflicts to territories defined by contemporary national borders. The examined time period excludes the colonial wars of the 19th and early 20th centuries, many of which were associated with the Scramble for Africa. In particular, because these wars occurred as a consequence of local resistance to the European colonizers or were triggered by the conflicting interests of the different colonial powers, they are not expected to be related to local population diversity in a meaningful way.

The definition of a violent conflict in Brecke's data set is based on Cioffi-Revilla (1996):"An occurrence of purposive and lethal violence among 2+ social groups pursuing conflicting political goals that results in fatalities, with at least one belligerent group organized under the command of authoritative leadership. The state does not have to be an actor. Data can include massacres of unarmed civilians or territorial conflicts between warlords." The list is comprised of conflicts that resulted in at least 32 fatalities.[25] Further, although the data set does not systematically distinguish between intrastate and interstate conflicts, the latter appear to form the basis of the recorded conflicts. Finally, while the recorded conflicts do not necessarily represent the universe of conflict events during the sample period, the list contains almost all major conflicts that have been documented by historians.

In contrast to the analysis of modern conflicts, the explanatory variable of interest in the current analysis is the precolonial population diversity (predicted by migratory distance from East Africa) of a territory bounded by its contemporary national borders. By construction, this measure does not account for the impact of post-1500 migrations on population diversity. In addition, it is not clear at the outset if one should expect any systematic relationship between the *native* population diversity of a given territory and the outbreak of interstate – as opposed to internal – conflicts in that territory. However, given that the measure of precolonial population diversity is collinear in migratory distance from East Africa, if a conflict's location is relatively close to the native territories of the warring parties in the conflict, the measure should possess some explanatory power for the onset of such conflicts. Because the conflicts examined occurred during a time period when long-distance campaigns were uncommon, due to the constraints imposed by historical transportation and warfare technologies, precolonial population diversity could in principle explain a considerable part of the variation in interstate conflicts across the globe, especially in earlier periods of the 1400–1799 time horizon.

[25]This fatality level corresponds to a magnitude of 1.5 or higher on Richardsons (1960) base-10 log conflict scale.

TABLE IV: Precolonial Population Diversity and the Occurrence of Historical Conflicts across Countries

Historical period:	1400–1799	1400–1499	1500–1599	1600–1699	1700–1799	1400–1799	1400–1499	1500–1599	1600–1699	1700–1799
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) Probit	(7) Probit	(8) Probit	(9) Probit	(10) Probit
	Number of conflict onsets in historical period					Onset of any conflict in historical period				
Population diversity (precolonial)	16.336***	13.561***	10.919***	9.878***	6.456**	18.211***	35.761***	17.266***	17.622***	12.508**
	(4.264)	(3.425)	(3.603)	(3.127)	(2.801)	(5.799)	(6.754)	(6.241)	(5.745)	(5.297)
Region dummies	×	×	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×	×	×
Observations	155	155	155	155	155	155	155	155	155	155
Partial R^2 of population diversity	0.104	0.136	0.087	0.064	0.039					
Adjusted R^2	0.354	0.367	0.356	0.251	0.231					
Pseudo R^2						0.248	0.374	0.285	0.224	0.213
Effect of 10th–90th %ile move in diversity	31.725***	8.352***	7.603***	5.911***	2.826**	0.541***	0.631***	0.515***	0.560***	0.430***
	(8.281)	(2.109)	(2.508)	(1.871)	(1.226)	(0.098)	(0.045)	(0.097)	(0.085)	(0.118)

Notes: This table exploits cross-country variations to establish a significant positive reduced-form impact of indigenous (precolonial) population diversity on (i) the number of conflict onsets (Columns 1–5); and (ii) the likelihood of observing one or more conflict onsets (Columns 6–10), either during the entire 1400–1799 time period (Columns 1 and 6) or in each century therein (Columns 2–5 and 7–10), conditional on the baseline geographical correlates of conflict. The controls for geography include absolute latitude, ruggedness, distance to the nearest waterway, the mean and range of agricultural suitability, the mean and range of elevation, and an indicator for small island nations. The set of region dummies includes four indicators for Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, and South Asia. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of either the number of conflict onsets (Columns 1–5) or the percentage-point increase in the likelihood of a conflict onset (Columns 6–10) during the time period examined by the regression. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table IV presents the analysis of historical conflicts. For the specifications in Columns 1–5, the outcome variable captures the (log-transformed) total number of distinct conflict events in different time intervals during the 1400–1799 time horizon.[26] The specification in Column 1 examines conflicts during the entire time horizon of four centuries, whereas those in Columns 2–5 focus on the conflicts recorded for each individual century of the time horizon. Indeed, the data on conflicts that occurred prior to the discovery of the New World is expected to be less contaminated by information on interstate conflicts between warring parties whose combined population diversity is not representative of the population diversity of the locations in which these conflict occurred. The specifications in Columns 6–10 replicate the analysis from Columns 1–5, except that the outcome variable is an indicator for conflict onset that captures whether there was any recorded conflict event during the specified time interval. All specifications include the geographical controls from the earlier analysis of modern conflicts. In addition, regional dummies are included in all regressions to mitigate the concern that Brecke's conflict data could suffer from a regional bias in coverage, due to differences across world regions in the quality of primary sources and in the nature and scale of historical conflict events.[27]

The results indicate that pre-colonial population diversity had a statistically significant positive influence on both the number and the incidence of historical conflicts. This is true for conflicts that occurred both in the century prior to the discovery of the New World and in the centuries that followed. However, in line with the prior that the impact of native population diversity on conflicts ought to dissipate in time periods marred by mostly international or interregional conflicts (particularly, those involving ancestrally very distant warring populations like the European colonial

[26]The log transformation is applied to one plus the total number of conflicts in order to retain observations without any recorded conflict.

[27]For example, primary sources on historical warfare in Sub-Saharan Africa are relatively scarce (Reid, 2014), and unlike the large-scale campaigns common in European warfare, historical conflicts in Africa more often took the form raiding wars.

powers versus the natives), the association between population diversity and conflicts is noticeably weaker in the centuries following the advent of the colonial era.

The OLS estimate in Column 2 implies that a move from the 10th to the 90th percentile in the cross-country distribution of population diversity is associated with 8.4 more conflicts during the 15th century. This impact is somewhat larger than those implied by the comparable specifications for modern civil conflicts.[28] This could potentially reflect a waning, albeit significant, influence of population diversity in more contemporary time periods. However, it could also be a mechanical consequence of measurement issues associated with the fact that in contrast to the earlier analyses of modern civil conflicts, the current analysis of historical conflicts does not distinguish between purely intrastate conflicts and interstate conflicts involving ancestrally proximate warring populations. As for the economic significance of population diversity for historical conflict incidence, the probit regression in Column 7 implies that a move from the 10th to the 90th percentile in the cross-country distribution of population diversity is associated with a 63 percent increase in the likelihood of observing a conflict during the 15th century.

In sum, beyond providing temporal external validity to the main findings from the earlier analyses of civil conflict in the contemporary era, the findings in this section attest to a deep-rooted and persistent influence of population diversity on the risk of conflict in society – an impact that is indeed apparent across many centuries.

4 Population Diversity and Conflict at the Ethnicity Level

This section explores the contribution of interpersonal population diversity to the existing variation in the prevalence and severity of conflicts within ethnic homelands. The focus on ethnic homelands permits the analysis to disentangle the impact of population diversity within an ethnic group, rather than across groups, on conflict. The ethnic-level analysis mitigates potential concerns about the confounding effects of population diversity as well as conflict on national borders.

4.1 Data

The ethnic-level analysis is conducted using two novel geo-referenced datasets of ethnic homelands across the globe. The first dataset consists of homelands of indigenous ethnic groups (largely isolated and shielded from genetic admixture) whose levels of diversity is provided by the most comprehensive source on observed genetic diversity (Pemberton et al., 2013).[29] The geo-referenced dataset maps the genetic diversity of individuals within each ethnic homeland, as reported in Pemberton et al. (2013), to the geographical characteristics of this homeland.[30] The data consists of 207 ethnic homelands for which genetic diversity is observed.[31] The distribution of these ethnic groups across the globe is depicted in Figure 3. The level of observed genetic diversity ranges from

[28] For instance, in Column 3 of Table I, the estimated impact of the same move in the cross-country distribution of population diversity is 0.02 additional civil conflict outbreaks per year – i.e., 2 additional conflicts per century.

[29] Pemberton et al. (2013) combines eight human genetic diversity datasets based on the 645 loci that they share, including the HGDP-CEPH Human Genome Diversity Cell Line Panel used by Ashraf and Galor (2013a). Since ethnic groups have been largely native to their ethnic homelands, at least since the pre-colonial era, the measure of population diversity within the ethnic groups properly captures the degree of population diversity within the ethnic homelands.

[30] Further details on the construction of the dataset are presented in Section B.1 of the Supplemental Material.

[31] The analysis includes all ethnic groups in Pemberton et al. (2013) that can be mapped into an ethnic homeland, excluding the Surui of South America. Population geneticists view the Surui as an extreme outlier in terms of genetic diversity. In particular, Ramachandran et al. (2005) omit the Surui, as "an extreme outlier in a variety of previous analyses". Including this observation, nevertheless, does not affect the qualitative results.

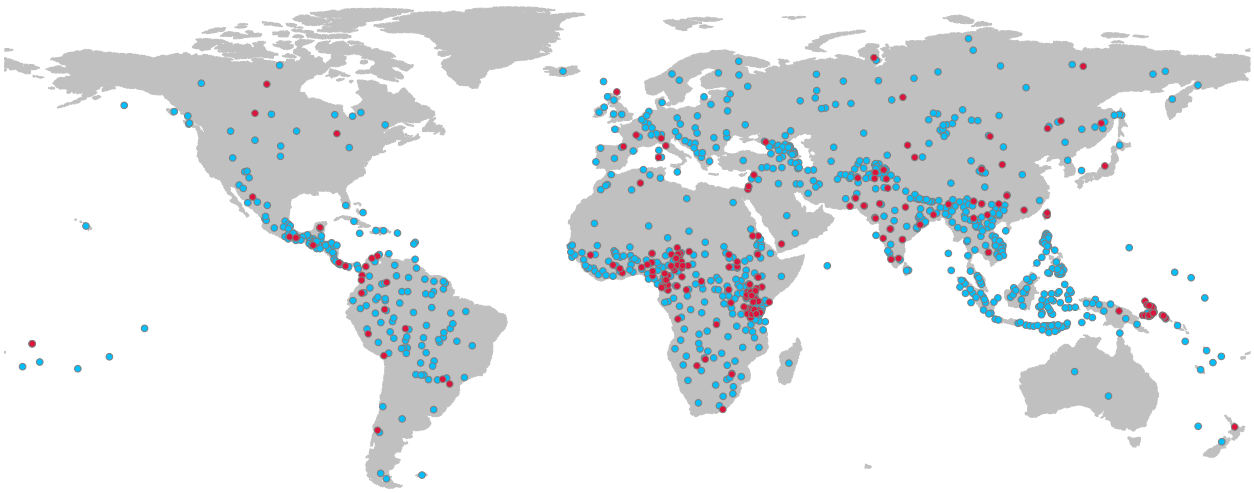


FIGURE 3: The Spatial Distribution of Ethnic Homelands

Notes: This map depicts the global spatial distribution of ethnic homelands in the sample. Each point represents the centroid of the historical homeland of an ethnic group. Red points depict homelands for which population diversity is observed, whereas blue points depict homelands for which population diversity is predicted.

0.55 among ethnic groups in South America to 0.77 among those in Africa. Appendix A.6 presents the summary statistics of the sample.

The second geo-referenced dataset consists of all homelands of ethnic groups, whose geographical territories are delineated by the GREG ("Geo-Referencing of Ethnic Groups"), as drawn from the classical Soviet Atlas Narodov Mira (Weidmann et al., 2010). Population diversity within these ethnic homelands is predicted based on prehistoric migratory distance from Addis Ababa, using the unconditional relationship between observed genetic diversity and prehistoric migratory distance from Addis Ababa derived from the 207-ethnic homelands sample.

While the historical homeland of each ethnic group captures the area of the globe in which the group is predominantly residing, the vast majority of ethnic homelands tend to be fractionalized, as indicated by the fact that they are inhabited by multiple linguistic groups. Hence, the analysis of the impact of interpersonal population diversity on conflict accounts for the potentially confounding effects of the degree of linguistic fractionalization and polarization within ethnic homelands on conflict.[32]

The main measure of conflict that is used at the ethnic-level analysis is derived from the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002). In particular, the analysis focuses on the average yearly share of the area of each ethnic homeland, over the period 1989–2008, that fell within the boundaries of internal armed conflict events (between the government of a state and internal opposition groups).[33] Furthermore, the analysis utilizes a second measure that accounts for the number of conflict events, the number of deaths, and the number of deaths per event, as recorded within each ethnic homeland in the UCDP Georeferenced Event Dataset (Sundberg et al., 2012; Croicu and Sundberg, 2015).

[32] As elaborated in Section B.2 of the Supplemental Material, the measures of the degree of ethnolinguistic fractionalization and polarization in ethnic homelands is based on the proportional representation of each linguistic group within the ethnic homeland.

[33] This measure is calculated using the gridded PRIO data (PRIO-GRID version 1.01) as reported by Tollefsen et al. (2012) based on the UCDP/PRIO Armed Conflict Dataset.

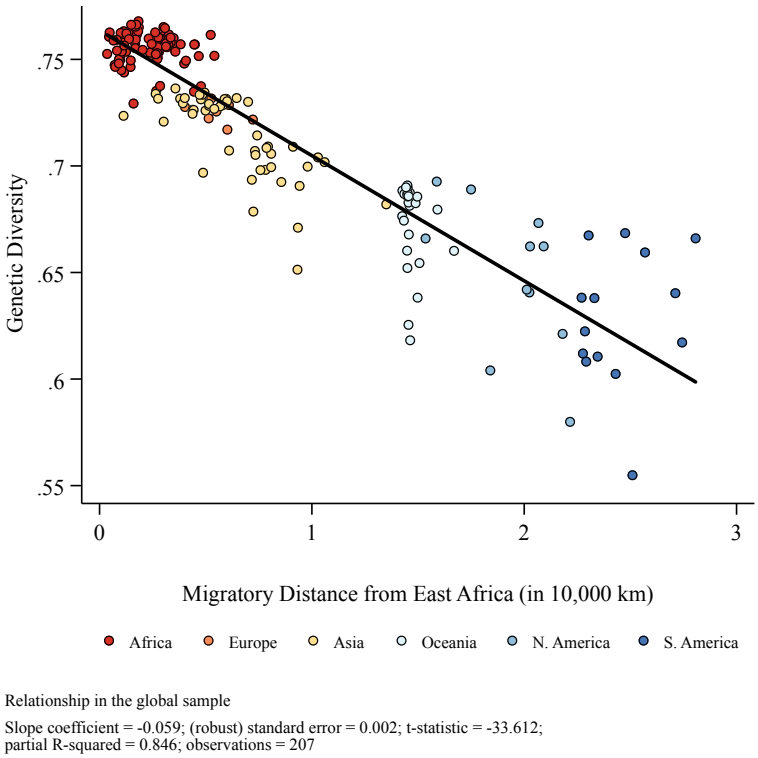


FIGURE 4: Migratory Distance from East Africa and Observed Genetic Diversity across Ethnic Homelands

Notes: This figure depicts the relationship between prehistoric migratory distance from East Africa and observed population diversity in a sample of 207 ethnic homelands. The negative relationship reflects the serial founder effect associated with expansion of humans from East Africa to the rest of the world.

4.2 Empirical Strategy

The analysis implements several empirical strategies to mitigate concerns about the potential role of reverse causality, omitted cultural, geographical and human characteristics, as well as sorting in the observed association between population diversity and civil conflicts within ethnic homelands. In particular, the positive associations between the extent of the observed population diversity within an ethnic homeland and civil conflict may reflect reverse causality from conflict to population diversity. It is not inconceivable that in the course of human history, conflicts within ethnic groups have operated towards homogenization of the population, reducing its observed levels of diversity. Hence, in order to mitigate concerns about reverse causality, as well as concerns about sample limitations, the ethnic-level analysis further exploits *predicted* population diversity, rather than observed diversity, to explore the effect of diversity on civil conflict. In particular, as caused by the serial founder effect (e.g., Harpending and Rogers, 2000; Ramachandran et al., 2005; Ashraf and Galor, 2013a) and depicted in Figure 4, observed population diversity within geographically indigenous contemporary ethnic groups decreases with distance along ancient migratory paths from East Africa. Hence, migratory distance from Africa is exploited to predict population diversity for all ethnic groups in the GREG.

Furthermore, the associations between ethnic-level population diversity and civil conflicts may be governed by omitted cultural, geographical and human characteristics. Thus, in order to mitigate these concerns, the empirical analysis exploits two related strategies. In light of the serial

founder effect, the analysis exploits the migratory distance from Africa to each ethnic group as an instrumental variable for the observed level of population diversity, and as a predictor for its level of diversity. Nevertheless, there are several plausible scenarios that would weaken this identification strategy. First, selective migration out of Africa, or natural selection operating in different ways along the migratory paths, could have affected human traits and therefore conflict independently of the effect of migratory distance from Africa on the degree of *diversity* in human traits. Second, migratory distance from Africa could be correlated with distances from focal historical locations (e.g., technological frontiers) and could therefore capture the effect of these distances on the process of development and the emergence of conflicts, rather than the effect of these migratory distances via population diversity.

These potential concerns are mitigated, however, by the following observations. First, while migratory distance from Africa has a significant negative association with the degree of genetic diversity, it has no apparent association with the mean *level* of human traits (Ashraf and Galor, 2013a), conditional on the distance from the equator. Second, conditional on migratory distance from East Africa, migratory distances from historical technological frontiers in the years 1, 1000, and 1500 do not affect the impact of population diversity on conflict, reinforcing the justification for the reliance on the "out of Africa" hypothesis and the serial founder effect.

Moreover, an unlikely threat to the identification strategy would emerge if the actual migration path out of Africa would have been correlated with geographical characteristics that are directly conducive to conflicts (e.g., soil quality, ruggedness, climatic conditions, and propensity to trade). This, however, would have implausibly involved that the conduciveness of these geographical characteristics to conflict would be aligned along the main root of the migratory path out of Africa, as well as along each of the main forks that emerge from this primary path. In particular, in several important forks in the course of this migratory path (e.g., the Fertile Crescent and the associated eastward migration towards East Asia and western migration towards Europe), the geographical characteristics that are conducive to conflicts would have to diminish symmetrically along these diverging migratory routes. Nevertheless, in order to further mitigate this unlikely concern, the analysis establishes that the results are unaffected qualitatively if it accounts for the potentially confounding effects of a wide range of geographical factors in the homeland of each ethnic group. In addition, in order to further mitigate concerns regarding the role of omitted variables, the analysis accounts for spatial auto-correlation as well as regional fixed effects, capturing time-invariant unobserved heterogeneity in each region and hence identifying the association between interpersonal diversity and conflict, within, rather than across, regions. Furthermore, it establishes that selection on unobservables is not a concern.

The observed associations between population diversity and the extent of conflicts may further reflect the sorting of less diverse populations into geographical niches characterized by lower conflict. While this sorting would not affect the existence of a positive association between population diversity and the extent of conflict, it could weaken the proposed mechanism. However, in view of the serial founder effect and the tight negative association between migratory distance from Africa and population diversity, sorting would necessitate that the ex-ante spatial distribution of conflict would have to be negatively correlated with migratory distance from Africa. As argued above, this would have implausibly involved that the conduciveness of geographical characteristics to conflict would be negatively aligned with the primary migratory path out of Africa, as well as with each of its diverging forks, diminishing symmetrically along these diverging migratory routes. Nevertheless, to further mitigate this unlikely scenario, the empirical analysis accounts for the potentially confounding effects of a wide range of geographical characteristics, as well as regional fixed effects.

4.3 Empirical Results

This subsection establishes a highly significant and robust reduced-form impact of observed and predicted diversity within an ethnic homeland on intra-societal conflicts within this homeland. The analyzes explores the effect of population diversity within ethnic groups on the prevalence of conflict, as well as on its intensity at the ethnic level. The empirical specifications in the ethnic-level analysis follows rather closely the specifications in the country-level analysis, assuring the comparability of the findings.

Tables V and VI present the results of the baseline analysis of the influence of interpersonal population diversity within an ethnic homeland on log conflict prevalence over the period 1989–2008. Table V conducts the analysis for the observed-diversity sample. In particular, Column 1 establishes a highly significant association between observed diversity across the 207 ethnic homelands and conflict prevalence, conditional on world-region fixed effects.[34] Column 2 demonstrates that — as depicted in Figure 5 — the association remains highly significant and even increases slightly in magnitude if one accounts for the potentially confounding effects of some exogenous geographical factors (i.e., absolute latitude, ruggedness, mean and range of elevation, mean and range of land suitability, distance from waterway, and an island dummy). Column 3 establishes that accounting for additional exogenous climatic factors which have been shown to be relevant for conflict (i.e., temperature and precipitation), the association between observed diversity and conflict remains highly significant. The coefficient estimate suggests that an increase in population diversity from the 10th percentile (e.g., the Mamusi people of Oceania) to the 90th percentile (e.g., the Pare people of Eastern Africa) corresponds to an average increase of 0.43 in the prevalence of conflicts within the territory of a homeland over the years 1989–2008 (compared to a sample mean of 0.14 and a standard deviation of 0.27).[35] Columns 4 and 5 establishe that the qualitative results are unaffected by accounting for the potentially confounding effects of linguistic fractionalization and polarization. Finally, Columns 6 and 7 demonstrate that the estimates remain highly significant and stable if one accounts for a set of potentially endogenous confounders (i.e., log luminosity, malaria endemicity, and time since settlement).

In light of the potential endogeneity of observed population diversity, Table VI presents the effect of predicted population diversity, based on prehistoric migratory distance from East Africa, on the prevalence of conflict in a sample of 901 ethnic homelands.[36] In particular, Column 1 establishes a highly significant effect of predicted diversity on log conflict prevalence, conditional on world-region fixed effects. Column 2 demonstrates that, as depicted in Figure 6, the effect remains highly significant and stable if one accounts for the potentially confounding effects of some exogenous geographical factors (i.e., absolute latitude, ruggedness, mean and range of elevation, mean and range of land suitability, distance from waterway, and an island dummy) as well as additional exogenous climatic factors which have been shown to be relevant for conflict (i.e., temperature and precipitation). In particular, the coefficient estimate suggests that an increase in predicted diversity from the 10th percentile (e.g., the Boruca people of Central America) to the 90th percentile (e.g.,

[34]The observed sample of 207 ethnic homelands disproportionately represents sub-Saharan Africa. Moreover, while the prevalence of conflict in ethnic homelands in Africa is significantly above the worldwide average, in the observed sample the prevalence of conflict is below the world average, introducing undesirable biases in the estimation and necessitating the use of regional-fixed effects, and in particular a Sub-Saharan dummy variable, to account for these regional anomalies. In contrast, in the representative predicted sample, considered in Table VI, the positive association between population diversity and conflict, within as well as between continents, can be identified.

[35]See Appendix A.6.

[36]The larger coefficient estimates for the impact of diversity on conflict in the predicted sample (relative to the observed sample) plausibly reflects the more representative spatial coverage of conflicts across the globe. Further, these larger estimates for predicted diversity are in line with the fact that the 2SLS estimates of instrumented observed diversity are also larger than their OLS counterparts.

TABLE V: Observed Population Diversity and Conflict across Ethnic Homelands

	Log conflict prevalence						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS
Observed population diversity	28.740***	33.896***	27.559***	27.998***	27.619***	29.020***	28.550***
	(9.638)	(10.161)	(9.634)	(9.455)	(9.511)	(10.662)	(10.727)
Ethnolinguistic fractionalization				1.291**		1.088*	
				(0.626)		(0.642)	
Ethnolinguistic polarization					0.811		0.733
					(0.523)		(0.527)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	No	No	Yes	Yes	Yes	Yes	Yes
Development outcomes	No	No	No	No	No	Yes	Yes
Disease environment controls	No	No	No	No	No	Yes	Yes
Sample	Observed	Observed	Observed	Observed	Observed	Observed	Observed
Observations	207	207	207	207	207	207	207
Effect of 10th-90th %ile move in diversity	0.449***	0.530***	0.431***	0.438***	0.432***	0.454***	0.446***
	(0.151)	(0.159)	(0.151)	(0.148)	(0.149)	(0.167)	(0.168)
Adjusted R^2	0.107	0.168	0.298	0.310	0.303	0.316	0.312
β^*		37.750	26.984	27.645	27.080	29.149	28.462

Notes: This table exploits variations across ethnic homelands to establish a significant positive impact of observed population diversity on the log prevalence of conflict during the 1989–2008 period, conditional on the potentially confounding effects of geographic, climatic, and development-related characteristics, as well as the disease environment. World-region fixed effects include Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Geographical controls are absolute latitude, ruggedness, mean and range of elevation, and mean and range of land suitability, distance from waterway, and an island dummy. Climatic controls are the mean levels of temperature and precipitation. Development outcomes are time since settlement, presence of oil and gas, and log luminosity. The disease environment control is malaria endemicity. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the prevalence of conflicts within the territory of a homeland over the years 1989–2008. The β^* statistic is the estimated effect of population diversity, if selection on observables and unobservables are of equal proportions, and the maximal R^2 is equal to 1.3 times the observed R^2 (Oster, 2019). Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

the Wafipa people of East Africa) corresponds to an average increase of 1.71 in the prevalence of conflicts within the territory of a homeland over the years 1989–2008 (compared to a sample mean of 0.19 and a standard deviation of 0.32).

Columns 3 and 4 of Table 6 establish that the qualitative results are unaffected by the potentially confounding effects of linguistic fractionalization and polarization, accounting for a set of potentially endogenous confounders (i.e., log luminosity, malaria endemicity, and time since settlement). Importantly, restricting the analysis to a sample of 697 ethnic homelands in the Old World, that are arguably less sensitive to the mass-migration in the post-1500 period, Columns 5 and 6 suggest that the effect of predicted diversity on conflict remains highly significant and larger, plausibly due to smaller measurement errors.

Finally, using prehistoric migratory distance from Africa as an instrumental variable for observed population diversity, the 2SLS regression analysis reported in Column 7, suggests that there exists a highly significant reduced-form impact of population diversity on conflict, accounting for world-region fixed effects, geographical, and climatic characteristics.[37] Furthermore, the results

[37]The first-stage F-statistic indicates that prehistoric migratory distance is a strong instrument for the level of observed population diversity. The large 2SLS coefficient on observed diversity, relative to its OLS counterpart from Column 3 of Table V, may be explained by the following two reasons. First, the OLS estimates may be afflicted by attenuation bias due to the possibility that observed diversity in neutral genetic markers is merely a *noisy* proxy of interpersonal diversity in unobserved traits that are relevant for socioeconomic interactions. Second, in line with the interpretation of a local average treatment effect (LATE), there could be certain ethnic groups in the observed sample

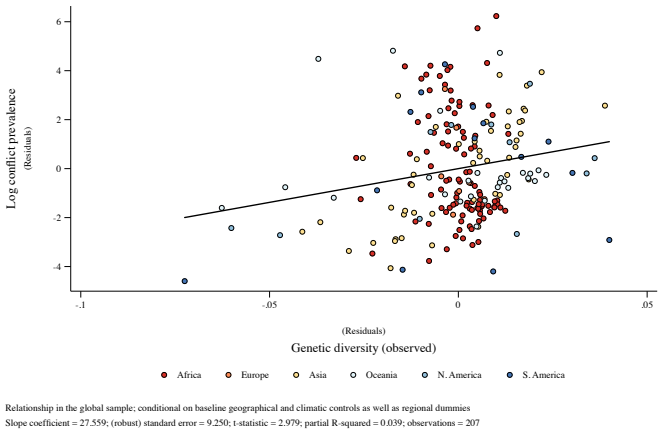


FIGURE 5: Observed Population Diversity and Conflict across Ethnic Homelands

Notes: This figure depicts the relationship between observed population diversity and conflict prevalence during the period 1989–2008 across 207 ethnic homelands, conditional on world-region fixed effects, and potential geographic and climatic confounders, as reported in Column 3 of Table V.

remain highly significant if one accounts for the potentially confounding effects of linguistic fractionalization and polarization in the ethnic homelands as well as development outcomes and the disease environment (results available upon request). In line with the results based on predicted diversity, once the potential change in diversity of ethnic groups due to conflict is accounted for, the estimated coefficient of interest in Column 7 suggests that an increase in population diversity from the 10th percentile (e.g., the Mamusi people of Oceania) to the 90th percentile (e.g., the Pare people of Eastern Africa) corresponds to an average increase of 2.03 in the prevalence of conflicts within the territory of a homeland over the years 1989–2008 (compared to a sample mean of 0.14 and a standard deviation of 0.27).

Table A.VIII in Appendix A.5 establishes that population diversity is a significant contributor to the total number of conflict events within an ethnic homeland during the 1989–2008 time period. Table A.IX establishes the significant impact of both observed population diversity and predicted population diversity on the number of conflicts, the number of deaths, and the number of deaths per conflict, accounting for world-region fixed effects, geographic and climatic characteristics, as well as linguistic fractionalization and polarization. Further, the baseline results with respect to the prevalence of conflicts across ethnic homelands are shown to be robust to accounting for: (i) spatial dependence across observations (Tables A.X and A.XI), and (ii) the use of predicted population diversity as a generated regressor (Table A.XII).

Finally, as established in Section B.3 of the Supplemental Material, the baseline results are qualitatively insensitive to accounting for: (i) migratory distances from historical technological frontiers (Table SB.I), and (ii) ecological diversity and ecological polarization (Tables SB.II and SB.III).

that are not perfect compliers of the "migratory distance" treatment, in the sense that their population diversities improperly reflect the legacy of the serial founder effect (due to some degree of admixture from non-native populations in the era of European colonization).

TABLE VI: Predicted Population Diversity and Conflict across Ethnic Homelands

	Log conflict prevalence						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Predicted population diversity	77.710***	77.031***	74.010***	73.581***	81.354***	80.889***	
	(6.279)	(7.282)	(7.396)	(7.418)	(9.623)	(9.735)	
Observed population diversity							129.610***
							(32.407)
Ethnolinguistic fractionalization			0.347		0.200		
			(0.299)		(0.356)		
Ethnolinguistic polarization				0.457*		0.629**	
				(0.263)		(0.311)	
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Development outcomes	No	No	Yes	Yes	Yes	Yes	No
Disease environment controls	No	No	Yes	Yes	Yes	Yes	No
Sample	Predicted	Predicted	Predicted	Predicted	Old World	Old World	Observed
Observations	901	901	901	901	697	697	207
Effect of 10th-90th %ile move in diversity	1.725***	1.710***	1.643***	1.633***	1.019***	1.013***	2.027***
	(0.139)	(0.162)	(0.164)	(0.165)	(0.120)	(0.122)	(0.507)
Adjusted R^2	0.211	0.362	0.378	0.379	0.401	0.404	
β^*		76.546	71.535	70.829	73.903	73.187	
Migratory distance from East Africa (in 10,000 km)							-0.044***
							(0.009)
First-stage F-statistic							26.185

Notes: This table exploits variations across ethnic homelands to establish a significant positive impact of predicted population diversity, based on prehistoric migratory distance from East Africa, on the log prevalence of conflict during the 1989–2008 period, conditional on the potentially confounding effects of geographic, climatic, and development-related characteristics, as well as the disease environment. World-region fixed effects include Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Geographical controls are absolute latitude, ruggedness, mean and range of elevation, and mean and range of land suitability, distance from waterway, and an island dummy. Climatic controls are the mean levels of temperature and precipitation. Development outcomes are time since settlement, presence of oil and gas, and log luminosity. The disease environment control is malaria endemicity. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the prevalence of conflicts within the territory of a homeland over the years 1989–2008. The 2SLS regression exploits prehistoric migratory distance from East Africa to each ethnic homeland as an excluded instrument for the observed population diversity of the ethnic group. The β^* statistic is the estimated effect of population diversity, if selection on observables and unobservables are of equal proportions, and the maximal R^2 is equal to 1.3 times the observed R^2 (Oster, 2019). Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

5 Potential Mediating Channels

What are the proximate factors that could explain the adverse reduced-form influence of interpersonal population diversity on different forms and dimensions of social conflict? This section explores some potential mediating channels at the national and subnational levels.

5.1 Ethnic Diversity, Interpersonal Trust, and Dispersion in Political Preferences at the Country Level

This subsection examines some hypothesized proximate mechanisms that can potentially mediate the positive reduced-form cross-country relationship between population diversity and the risk of intrastate conflict, as reflected by the annual frequency of new PRIO25 civil conflict outbreaks during the 1960–2017 time period. Specifically, it provides evidence that the main cross-country empirical finding may partly be a ramification of (i) the contribution of interpersonal population diversity to the degree of ethnolinguistic fragmentation at the country level, measured by the total

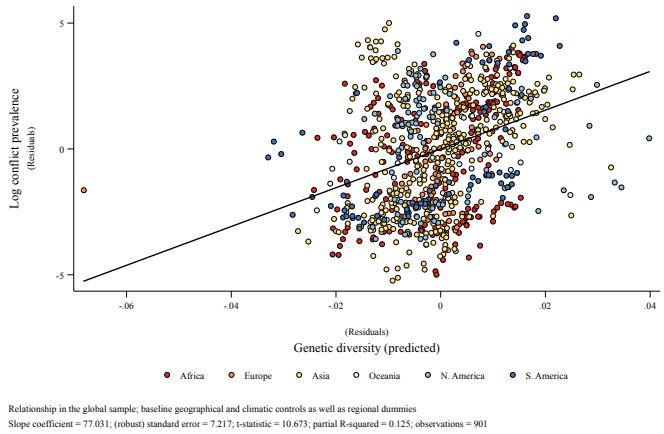


FIGURE 6: Predicted Population Diversity and Conflict across Ethnic Homelands

Notes: This figure depicts the relationship between predicted population diversity and conflict prevalence during the period 1989–2008 across 901 ethnic homelands, conditional on world-region fixed effects, and potential geographic and climatic confounders, as reported Column 2 of Table VI.

number of ethnic groups in a national population Fearon (2003);[38] (ii) the adverse influence of population diversity on social capital, based on data from the World Values Survey (2006, 2009) (henceforth, WVS) on the prevalence of generalized interpersonal trust in a country's population;[39] and (iii) the association between population diversity and heterogeneity in preferences for public goods and redistributive policies at the national level, as captured by the intra-country dispersion in self-reported individual political positions on a politically "left"–"right" categorical scale, based on data from the WVS.[40]

Table VII reports the findings from an empirical examination of the aforementioned three potential mechanisms through which population diversity could partly contribute to the risk of intrastate conflict in society. For each posited channel, the analysis presents the results from estimating three different OLS regressions, exploiting worldwide variations in a common sample of countries, conditioned primarily by the availability of data on the mediating variable in question. In addition, all examined specifications partial out the influence of only the baseline set of geographical covariates (including continent or regional fixed effects). Specifically, the analysis does not include potentially endogenous control variables, many of which (like GDP per capita) may well be afflicted

[38]Unlike measures of ethnolinguistic fragmentation that are based on fractionalization or polarization indices, the number of ethnic groups in the national population is potentially less endogenous in an empirical model of the risk of civil conflict, in light of the fact that this measure is not additionally tainted by the incorporation of information on the endogenous shares of the different subnational groups.

[39]In particular, this well-known measure of social capital reflects the proportion in a given country of all respondents (from across five different waves of the WVS, conducted over the 1981–2009 time horizon) that opted for the answer "Most people can be trusted" (as opposed to "Can't be too careful") when responding to the survey question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"

[40]Specifically, this country-level measure of heterogeneity in political attitudes reflects the intra-country standard deviation across all respondents (sampled over five different waves of the WVS during the 1981–2009 time horizon) of their self-reported positions on a categorical scale from 1 (politically "left") to 10 (politically "right") when answering the survey question "In political matters, people talk of 'the left' and 'the right.' How would you place your views on this scale, generally speaking?" Given that this variable's unit of measurement does not possess a natural interpretation, the cross-country distribution of this variable is standardized prior to conducting the regressions.

TABLE VII: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Mediating Channels

Mediating channel:	Cultural fragmentation			Interpersonal trust			Preference heterogeneity		
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS
	Log number of ethnic groups	Annual frequency of new civil conflict onsets, 1960–2017		Prevalence of interpersonal trust	Annual frequency of new civil conflict onsets, 1960–2017		Variation in political attitudes	Annual frequency of new civil conflict onsets, 1960–2017	
Population diversity (ancestry adjusted)	5.187***	0.316***	0.294***	−1.817**	0.488**	0.447*	14.344**	0.451**	0.375
	(1.887)	(0.114)	(0.109)	(0.848)	(0.221)	(0.236)	(6.675)	(0.219)	(0.254)
Log number of ethnic groups			0.004						
			(0.005)						
Prevalence of interpersonal trust						−0.023			
						(0.026)			
Variation in political attitudes									0.005
									(0.006)
Continent/region dummies	×	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×	×
Observations	147	147	147	84	84	84	81	81	81
Partial R^2 of population diversity	0.049	0.047	0.039	0.075	0.062	0.049	0.082	0.050	0.033
Adjusted R^2	0.342	0.203	0.201	0.441	0.232	0.226	0.397	0.247	0.249
Effect of 10th–90th %ile move in diversity	2.136***	0.021***	0.020***	−0.104**	0.029**	0.026*	0.824**	0.027**	0.022
	(0.777)	(0.008)	(0.007)	(0.049)	(0.013)	(0.014)	(0.383)	(0.013)	(0.015)

Notes: This table exploits cross-country variations to demonstrate that the significant positive reduced-form influence of contemporary population diversity on the annual frequency of new PRIO25 civil conflict onsets during the 1960–2017 time period, conditional on the baseline geographical correlates of conflict, is at least partly mediated by each of three potentially conflict-augmenting proximate channels that capture the contribution of population diversity to (i) the degree of cultural fragmentation, as reflected by the number of ethnic groups in the national population (Columns 1–3); (ii) the diminished prevalence of generalized interpersonal trust at the country level (Columns 4–6); and (iii) the extent of heterogeneity in preferences for redistribution and public-goods provision, as reflected by the intra-country dispersion in individual political attitudes on a politically "left"–"right" categorical scale (Columns 7–9). For each of the three mediating channels examined, the first regression documents the impact of population diversity on the proximate variable in the channel, the second presents the reduced-form influence of population diversity on conflict, and the third runs a "horse race" between population diversity and the proximate variable to establish reductions in the magnitude and explanatory power of the reduced-form influence of population diversity on conflict. All three regressions for each channel are conducted using a common cross-country sample, conditioned by the availability of data on the relevant variables employed by the analysis of the channel in question. The controls for geography include absolute latitude, ruggedness, distance to the nearest waterway, the mean and range of agricultural suitability, the mean and range of elevation, and an indicator for small island nations. The regressions for the "cultural fragmentation" channel control for the full set of continent dummies (i.e., five indicators for Africa, Asia, North America, South America, and Oceania), whereas for the "trust" and "preference heterogeneity" channels, given the smaller degrees of freedom afforded by the more limited sample of countries, the regressions control for a more modest set of region dummies, including an indicator for Sub-Saharan Africa and another for Latin America and the Caribbean. Given that the unit of measurement for the variable reflecting the degree of intra-country dispersion in political attitudes has no natural interpretation, its cross-country distribution is standardized prior to conducting the relevant regressions. The estimated effect associated with increasing diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of (i) the actual number of ethnic groups in the national population in Column 1; (ii) the fraction of individuals in a country who "think that most people can be trusted" in Column 4; (iii) the number of standard deviations of the cross-country distribution of the national-level dispersion in political attitudes in Column 7; and (iv) the number of new conflict onsets per year in all the remaining columns. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

by reverse causality from the temporal frequency of civil conflict onsets and may also be partly determined by both population diversity and the mediating variable.

The analysis of each mechanism proceeds by first regressing the mediating variable on population diversity. These regressions are presented in Columns 1, 4, and 7. All coefficients on the mediating variables are statistically significant at the 5 percent level or below. They suggest that conditional on exogenous geographical factors, a move from the 10th to the 90th percentile of the cross-country diversity distribution in the relevant sample is associated with (i) an increase in the total number of ethnic groups in a national population by 2.1 groups; (ii) a decrease in the prevalence of generalized interpersonal trust at the country level by 10.4 percent; and (iii) an

increase in the intra-country dispersion in individual political attitudes by 82.4 percent of a standard deviation from the cross-country distribution of this particular measure.[41]

The latter two regressions in the analysis of each hypothesized channel establish that the quantitative importance of population diversity as a predictor of the risk of civil conflict becomes diminished in both magnitude and explanatory power once the reduced-form influence of population diversity on the temporal frequency of civil conflict outbreaks is conditioned on the mediating variable of interest. Specifically, a comparison of the regressions in Columns 2 versus 3 indicates that, when conditioned on the total number of ethnic groups in the national population, the influence of population diversity on conflict frequency, in terms of the response associated with a move from the 10th to the 90th percentile of the cross-country diversity distribution, is reduced in magnitude by about 5 percent (from 0.021 to 0.020 new PRIO25 civil conflict onsets per year). Moreover, the explanatory power of population diversity for conflict frequency, as reflected by the partial R^2 statistic, diminishes by 17 percent. The corresponding results obtained for each of the other two posited mechanisms are qualitatively similar, and if anything, even more pronounced. In particular, when conditioned on either the prevalence of generalized interpersonal trust in the national population or the intra-country dispersion in political attitudes, the magnitude of the response in conflict frequency that is associated with a move from the 10th to the 90th percentile of the cross-country diversity distribution decreases by either 10.3 percent (Columns 5 versus 6) or 18.5 percent (Columns 8 versus 9), with the explanatory power of population diversity for conflict frequency declining by either 21 percent or 34 percent. Further, as shown in Column 9, the reduced-form influence of population diversity on the frequency of conflict outbreaks becomes statistically indistinguishable from zero when conditioned on the intra-country dispersion in political attitudes.

One important caveat regarding the interpretation of the findings in Table VII is that the mediating variables considered here may themselves be endogenous in a model of conflict risk (Rohner et al., 2013a). Indeed, as corroborated by empirical evidence from recent studies (e.g., Fletcher and Iyigun, 2010; Rohner et al., 2013b; Cassar et al., 2013; Besley and Reynal-Querol, 2014), the unobserved historical cross-regional pattern of conflict risk may have partly contributed to the contemporary variations observed across countries in the degree of ethnolinguistic fragmentation, the prevalence of interpersonal trust, and the intra-country dispersion in revealed political preferences. In particular, past conflicts plausibly triggered movements of ethnic groups across space and reinforced extant inter-ethnic cleavages along with the social, political, and economic grievances associated with such divisions. Thus, one ought to be cautious when interpreting the findings from the current analysis as conclusive evidence of the role of these factors as mediators. In order to assess these hypothesized mechanisms more conclusively, one would need to exploit an independent exogenous source of variation for each of these proximate factors, a task that remains open for future exploration.

5.2 Interpersonal Trust at the Individual Level

The proposed hypothesis suggests that interpersonal population diversity is conducive to conflict partly due to its adverse effect on trust and social cohesiveness. This section sheds light on this suggested mechanism, exploring the relationship between interpersonal population diversity

[41] The three scatter plots presented in Figure A.1 of Appendix A.3 depict these statistically significant cross-country relationships, conditional on the baseline set of geographical covariates (including continent or region fixed effects). Specifically, they show the relationship between population diversity and (i) the total number of ethnic groups in a national population (Panel (a)); (ii) the prevalence of generalized interpersonal trust at the country level (Panel (b)); and (iii) the intra-country dispersion in political attitudes (Panel (c)).

and interpersonal trust, using individual data.[42] The analysis establishes that a higher degree of population diversity is indeed associated with a lower level of interpersonal trust, suggesting that the impact of diversity on the prevalence of conflict could plausibly operate through the adverse effect of diversity on trust.

5.2.1 Population Diversity and Trust: Individuals in Africa

The analysis establishes a negative association between observed population diversity in ethnic homelands in Africa and the level of trust of individuals (surveyed by the Afrobarometer) who are originated from these homelands and are either residing in their ethnic homelands or in other regions of Africa. This negative association is robust if one accounts for (i) host-country fixed effects, (ii) individual-level characteristics (i.e., age, gender, education, occupation, living condition, and religion), (iii) exposure to slave exports, (iv) indicators of host district characteristics (i.e., presence of school, electricity, piped water, sewage, health clinic, and urban status), and (v) ancestral country fixed effects.[43] Moreover, the analysis accounts for the degree of fragmentation in the ethnic homeland as well as in the host district. Fragmentation in ethnic homelands is captured by linguistic fractionalization and polarization in these ethnic homelands, whereas fragmentation in the host district is captured by ethnic fractionalization in the district as well as the proportion of the respondent's group in the district population.

Table VIII presents the regression analysis of trust towards other individuals within the ethnic group on the level of interpersonal population diversity in the group.[44] The coefficient suggests that an increase in observed population diversity within an ethnic group from the 10th percentile of the distribution (e.g., individuals belonging to the Ashanti people) to the 90th percentile (e.g., individuals belonging to the Sukuma people) corresponds to a 0.29–0.59 point decrease in intra-group trust (compared to a sample mean of 1.52 and a standard deviation of 1.00). The analysis further suggests that ethnolinguistic fractionalization and polarization in the ethnic homeland has an adverse effect on intra-group trust.

5.2.2 Population Diversity and Trust: Second-Generation Migrants (U.S.)

This subsection explores the effect of population diversity in the ancestral country of second-generation migrants in the United States on their level of trust (as reported in the General Social Survey, GSS). The focus on a single country permits the analysis to account for time-invariant unobserved heterogeneity in the host country (e.g., geographical, cultural, and institutional characteristics).[45] Moreover, the analysis accounts for a range of individual controls, as well as geographical characteristics, regional fixed effects, and the degree of ethnolinguistic fractionalization and polarization, all in the ancestral country of origin.[46]

[42] Summary statistics for the trust analysis samples can be found in Section B.4 of the Supplemental Material.

[43] Since a third of the observations in the sample are individuals who are currently residing in Nigeria, and since Nigeria has the lowest level of trust among the 9 countries in the sample, possibly due to omitted variables (e.g., corruption), and since the level of genetic diversity in Nigeria is not among the highest in the sample, the actual relationship between diversity and trust may be masked in the absence of country dummies. Thus, all columns of the table account for host country fixed effects.

[44] The classification of individuals and their association with various ethnic homelands is based on Nunn and Wantchekon (2011).

[45] In addition, the focus on second-generation rather than first-generation migrants allows the analysis to exploit the individual-level variation in trust that plausibly mostly reflects the trust attitudes transmitted intergenerationally from parents rather than from society at large.

[46] Since the sample of second-generation migrants consists of 76% immigrants from Europe, 3% immigrants from Asia and 21% immigrants from the Americas, and since individuals from Europe has the highest level of trust among

TABLE VIII: Ethnic-Homeland Population Diversity and Individual-Level Trust in Africa

	Intra-group trust									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Observed population diversity (ancestral)	-39.496**	-38.335**	-45.303***	-34.849**	-37.840**	-38.467**	-45.567***	-35.190**	-64.122***	-70.334***
	(17.304)	(15.859)	(12.489)	(17.698)	(17.240)	(15.572)	(10.702)	(15.368)	(16.540)	(20.333)
Ethnolinguistic fractionalization (ancestral)			-0.443				-0.447		-0.934***	
			(0.314)				(0.306)		(0.228)	
Ethnolinguistic polarization (ancestral)				-0.973***				-0.959***		-1.264***
				(0.160)				(0.211)		(0.415)
District-level ethnic fractionalization					-0.057		0.006	0.019	0.030	0.027
					(0.052)		(0.185)	(0.201)	(0.213)	(0.225)
Proportion of ethnic group in district						0.076	0.087	0.071	0.037	0.029
						(0.108)	(0.264)	(0.258)	(0.213)	(0.210)
Host country dummies	×	×	×	×	×	×	×	×	×	×
Baseline individual controls		×	×	×	×	×	×	×	×	×
Education dummies									×	×
Occupation dummies									×	×
Living conditions dummies									×	×
Religion dummies									×	×
Slave export control									×	×
Host district characteristics dummies									×	×
Ancestral country dummies									×	×
Urban dummy									×	×
Number of Observations	3212	3212	3212	3212	3212	3212	3212	3212	2916	2916
Adjusted R2	0.218	0.225	0.230	0.234	0.225	0.226	0.230	0.234	0.289	0.287
Effect of 10th-90th %ile move in diversity	-0.331**	-0.321**	-0.379***	-0.292**	-0.317**	-0.322**	-0.382***	-0.295**	-0.537***	-0.589***
	(0.145)	(0.133)	(0.105)	(0.148)	(0.144)	(0.130)	(0.090)	(0.129)	(0.138)	(0.170)

Notes: This table presents the results of an individual-level OLS regression analysis of interpersonal trust towards individuals of the same ethnicity (as reported in Nunn and Wantchekon (2011)) on observed population diversity in the ancestral ethnicity of these individuals, controlling for a range of individual characteristics (i.e., age, gender, living conditions, education, religion), the presence of a school, electricity, piped water, sewage, a health clinic, in the local area, whether the local area is urban, and the intensity of Atlantic and Indian slave exports. In addition, the analysis accounts for host country fixed effects as well as fixed effects associated with the ancestral country. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the trust variable. Heteroskedasticity-robust standard errors, clustered multi-dimensionally at both the ancestral ethnic group and the host country, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table IX explores the association between the trust of second-generation migrants and the degree of population diversity in their parental country of origin. Column 1 establishes a negative and highly significant association between population diversity in the parental country of origin and trust of second-generation migrants, accounting for regional fixed effects associated with the parental country of origin.[47] This highly significant negative association remains largely stable if one accounts for interview-year fixed effects (Column 2), and the fixed effects associated with the respondent's age, sex, income, education, religion, and region within the United States (i.e., where the interview was conducted) in addition to the ethnic fractionalization or polarization of the homeland (Columns 3 and 4). Moreover, the results are robust to controlling for geographical characteristics of the parental country of origin (Columns 5 and 6).[48] The coefficient of interest in Column 4 suggests that an increase in population diversity in the parental country of origin from

these three groups, possibly due to omitted variables (e.g., income), and since the level of genetic diversity in Europe is highest among the three groups, an artificially positive relationship between trust and population diversity in the sample as a whole may appear in the absence of ancestral regional dummies. Thus, all columns of the table account for ancestral regional fixed effects. Since migrants from the Americas in the sample are originated from either Canada or Mexico, where Canada is significantly more diverse, due to a larger European population and significantly more trustful, possibly due to higher income, the use of a North America dummy only will affect the significance of the results. Hence, all columns of the table account for Latin American regional fixed effects.

[47]Since the sample is composed of individuals from European countries, Asian countries, and three countries in the Americas: Canada, Mexico, and Puerto Rico, the regional dummies distinguish between European, Asian, and Latin American countries.

[48]The inclusion of geographical characteristics of the ancestral homeland reduces the sample due to the absence of some of the relevant data for Puerto Rico.

TABLE IX: Country-of-Origin Population Diversity and Individual-Level Trust among Second-Generation U.S. Immigrants

	Trust					
	(1)	(2)	(3)	(4)	(5)	(6)
Population diversity (ancestral)	-14.670***	-15.036***	-10.175**	-9.820**	-12.343***	-12.358***
	(4.234)	(3.736)	(4.483)	(4.546)	(2.368)	(1.714)
Ethnic fractionalization (ancestral)			0.014		0.004	
			(0.182)		(0.202)	
Ethnolinguistic polarization (ancestral)				-0.028		-0.012
				(0.094)		(0.122)
Regional dummies (ancestral)	×	×	×	×	×	×
GSS year		×	×	×	×	×
Baseline individual controls			×	×	×	×
Income dummies			×	×	×	×
Education dummies			×	×	×	×
Religion dummies			×	×	×	×
Region of interview dummies			×	×	×	×
Geographical controls (ancestral)					×	×
Number of Observations	2294	2294	1785	1785	1785	1785
Adjusted R2	0.029	0.036	0.096	0.096	0.096	0.096
Effect of 10th-90th %ile move in diversity	-1.032***	-1.058***	-0.716**	-0.691**	-0.868***	-0.869***
	(0.298)	(0.263)	(0.315)	(0.320)	(0.167)	(0.121)

Notes: This table presents the results of an individual-level OLS regression analysis of interpersonal trust among second-generation migrants in the US on population diversity in their parental country of origin (as captured by ancestry-adjusted predicted diversity; Ashraf and Galor (2013a)), accounting for a range of individual-level socioeconomic characteristics (i.e., age, gender, income, religion, education), as well as time period fixed effects, parental region fixed effects, and the US host region fixed-effect. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the trust variable. Heteroskedasticity-robust standard errors, clustered multi-dimensionally at both the ancestral country and the US region of interview, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

the 10th percentile of the predicted contemporary level of diversity (e.g., individuals of Mexican descent) to the 90th percentile (e.g., individuals of Austrian descent) corresponds to a decrease in trust by 0.69 points (compared to a sample mean of 1.88 and standard deviation of 0.97). The analysis further suggests that ethnolinguistic fractionalization and polarization in the parental country of origin have no significant association with trust.

6 Concluding Remarks

This research explores one of the deepest roots of the prevailing variations in the emergence, prevalence, recurrence, and severity of intrasocietal conflicts, molded during the dawn of the dispersion of anatomically modern humans across the globe. It advances the hypothesis and establishes empirically that interpersonal population diversity, as determined predominantly during the exodus of humans from Africa tens of thousands of years ago, has been pivotal to historical and contemporary civil conflicts. The findings arguably reflect the contribution of population diversity to the non-cohesiveness of society, as reflected partly in the prevalence of mistrust, the divergence in preferences for public goods and redistributive policies, and the degree of fractionalization and polarization across ethnic, linguistic, and religious groups. Future research ought to focus on a deeper exploration of these and other possible mechanisms in order to better inform policies geared towards the implementation of optimal educational and sociopolitical institutions that could address the contribution of diversity to the non-cohesiveness of society.

Appendix

A.1 Analysis of Intrastate Conflict Severity in Repeated Cross-Country Data

The findings in Section 3.2 indicate that population diversity is a robust and significant reduced-form contributor to the contemporary risk of conflict in society, as manifested by the frequency, prevalence, and emergence of civil conflict events in the post-1960 time period. However, the outcome variables employed by those regressions are based on binary measures that are subject to a predefined threshold of annual battle-related casualties, which needs to be surpassed for a civil conflict event to be identified as such. Therefore, broadly speaking, the earlier findings reflect the influence of interpersonal population diversity on the *extensive* margin of conflict. This appendix section explores the influence of population diversity on the *intensive* margin of conflict. In particular, it employs both ordinal and continuous measures that capture the "severity" of intrastate conflicts and of events related to general social unrest, including but not limited to armed conflict.

The first measure of conflict intensity exploits information on the apparent "magnitude scores" associated with "major episodes" of intrastate armed conflict, as reported by the Major Episodes of Political Violence (MEPV) data set (Marshall, 2017).[49] According to this data set, a "major episode" of armed conflict involves both (i) a minimum of 500 directly related fatalities in total; and (ii) systematic violence at a sustained rate of at least 100 directly related casualties per year. Importantly, for each such episode of conflict, the MEPV data set provides a "magnitude score" —namely, an ordinal measure on a scale of 1 to 10 of the episode's destructive impact on the directly affected society, incorporating information on multiple dimensions of conflict severity, including the capabilities of the state, the interactive intensity (means and goals) of the oppositional actors, the area and scope of death and destruction, the extent of population displacement, and the duration of the episode. The specific outcome variable from the MEPV data set employed by the current analysis reflects the aggregated magnitude score across all conflict episodes that are classified as one of four types of intrastate conflict —namely, civil war, civil violence, ethnic war, and ethnic violence.[50] In particular, this variable is reported by the MEPV data set at the country-year level, with nonevent years for a country being coded as 0.

The second measure of conflict intensity is based on annual time-series data on a continuous index of social conflict at the country level, as reported by the Cross-National Time-Series (CNTS) Data Archive (Banks and Wilson, 2018). Rather than adopting an ad hoc fatality-related threshold for the identification of conflict events, this index provides an aggregate summary of the general level of social dissonance in any given country-year, by way of measuring a weighted average across all observed occurrences of eight different types of sociopolitical unrest, including assassinations, general strikes, guerrilla warfare, major government crises, political purges, riots, revolutions, and anti-government demonstrations.[51]

[49] The version of the MEPV data set employed provides annual information for a total of 179 countries over the 1946–2017 time period. See http://www.systemicpeace.org/inscr/MEPVcodebook2016.pdf for further details on the measure of conflict intensity from the MEPV data set.

[50] Specifically, all episodes of intrastate conflict in the MEPV data set are categorized along two dimensions. With respect to the first dimension, an episode may be considered either (i) one of "civil" conflict, involving rival political groups; or (ii) one of "ethnic" conflict, involving the state agent and a distinct ethnic group. In terms of the second dimension, however, an episode may be either (i) one of "violence," involving the use of instrumental force, without necessarily possessing any exclusive goals; or (ii) one of "war," involving violent activities between distinct groups, with the intent to impose a unilateral result to the contention.

[51] The specific weights (reported in parentheses) assigned to the different types of sociopolitical unrest considered by the index are as follows: assassinations (25), general strikes (20), guerrilla warfare (100), major government crises (20), political purges (20), riots (25), revolutions (150), and anti-government demonstrations (10). This

TABLE A.I: Population Diversity and the Severity of Civil Conflict in Repeated Cross-Country Data

Cross-country sample:	Old World		Global		Old World		Global	
	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS	(5) OLS	(6) OLS	(7) 2SLS	(8) 2SLS
	Quinquennial MEPV civil conflict severity, 1960–2017				Quinquennial CNTS social conflict index, 1960–2014			
Population diversity (ancestry adjusted)	4.241***	4.089**	4.159***	3.981**	5.306**	5.619***	5.679**	6.106***
	(1.452)	(1.803)	(1.531)	(1.987)	(2.350)	(1.982)	(2.599)	(2.289)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	1,270	1,045	1,576	1,311	1,144	924	1,430	1,165
Countries	123	121	149	147	123	120	150	146
Partial R^2 of population diversity	0.009	0.006			0.006	0.005		
Adjusted R^2	0.630	0.614			0.082	0.104		
Effect of 10th–90th %ile move in diversity	0.199***	0.183**	0.276***	0.264**	0.249**	0.264***	0.370**	0.405***
	(0.068)	(0.081)	(0.102)	(0.132)	(0.110)	(0.093)	(0.169)	(0.152)
First-stage F statistic			150.323	101.923			147.137	93.983

Notes: This table exploits variations in repeated cross-country data to establish a significant positive reduced-form impact of contemporary population diversity on the severity of conflict, as reflected by (i) the maximum value of an annual ordinal index of conflict intensity (from the MEPV data set) across all years in any given 5-year interval during the 1960–2017 time period; and (ii) the maximum value of an annual continuous index of the degree of social unrest (from the CNTS data set) across all years in any given 5-year interval during the 1960–2014 time period, conditional on other well-known diversity measures as well as the proximate geographical, institutional, and development-related correlates of conflict. Given that both measures of conflict severity are expressed in units that have no natural interpretation, their intertemporal cross-country distributions are standardized prior to conducting the regression analysis. The controls for geography include absolute latitude, ruggedness, distance to the nearest waterway, the mean and range of agricultural suitability, the mean and range of elevation, and an indicator for small island nations. The controls for ethnic diversity include ethnic fractionalization and polarization. The controls for institutions include a set of legal origin dummies, comprising two indicators for British and French legal origins, as well as six time-dependent covariates that capture the average annual values over the previous 5-year interval of the degree of executive constraints, two indicators for the type of political regime (democracy and autocracy), and three indicators for experience as a colony of the U.K., France, and any other major colonizing power. The control for oil presence is a time-invariant indicator for the discovery of a petroleum (oil or gas) reserve by the year 2003. The controls for population and income are the time-dependent log-transformed average annual values over the previous 5-year interval of total population and GDP per capita. To account for temporal dependence in conflict outcomes, all regressions control for the value of the outcome variable from the previous 5-year interval. For regressions based on the global sample, the set of continent dummies includes five indicators for Africa, Asia, North America, South America, and Oceania, whereas for regressions based on the Old-World sample, the set includes two indicators for Africa and Asia. The 2SLS regressions exploit prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of standard deviations of the intertemporal cross-country distribution of the outcome variable. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Given that the current analysis of conflict severity follows Esteban et al. (2012), in terms of exploiting variations in quinquennially repeated cross-country data, for each country, the annual data on either measure of conflict intensity is collapsed to a quinquennial time series, by assigning to any given 5-year interval in the post-1960 sample period, the maximum level of conflict intensity reflected by that measure across all years in the 5-year interval. As in earlier analyses of civil conflict incidence and onset, the examination focuses on better-identified specifications that either (i) exploit variations in a sample of countries belonging only to the Old World, or (ii) exploit

weighting methodology is based on Rummel (1963). For further details, the reader is referred to the codebook of the CNTS data archive, available at http://www.cntsdata.com/.

migratory distance from East Africa as an instrument for contemporary population diversity in a global sample of countries. All regressions account for temporal dependence in conflict severity by allowing both the lagged observation of the outcome variable and a full set of time-interval (5-year period) dummies to enter the specification. Further, whenever time-varying covariates are allowed to enter the specification, they do so with a one-period lag. Finally, because the units in which the proxies of conflict intensity are measured in the data have no natural interpretation, the outcome variables are standardized prior to running the regressions.

Table A.I presents the results from the analysis of the influence of interpersonal diversity on intrastate conflict severity, as reflected by either the MEPV aggregate magnitude score of conflict intensity (Columns 1–4) or the CNTS index of social conflict (Columns 5–8).[52] Regardless of the measure for conflict intensity examined, the identification strategy exploited, or the set of covariates considered by the specification, the results from the analysis of conflict severity in Table A.I establish population diversity as a qualitatively robust and significant reduced-form contributor to the intensive margin of intrastate conflict. Specifically, a move from the 10th to the 90th percentile of the cross-country distribution of population diversity in the relevant sample is associated with an increase in conflict severity by 18 to 28 percent of a standard deviation from the observed distribution of the MEPV magnitude score of conflict intensity, and with an an increase in general social unrest by 25 to 41 percent of a standard deviation from the observed distribution of the CNTS index of social conflict.

A.2 Robustness Checks for the Country-Level Analyses

Selection on Observables and Unobservables Following the methodology of Altonji et al. (2005), the current analysis exploits the idea that the amount of selection bias due to the unobserved variables in a model can be inferred from the reduction in selection bias from the inclusion of additional observed variables, thus permitting an assessment of how much larger the bias from unobserved heterogeneity needs to be, relative to the bias from observables, in order to fully explain away the coefficient on the explanatory variable of interest.[53] Specifically, the analysis compares the estimated coefficient, $\hat{\beta}_1^R$, on population diversity from a restricted model (conditioned on a subset of controls) with its estimated coefficient, $\hat{\beta}_1^F$, from an augmented model (conditioned on the full set of controls), examining the Altonji et al. (2005) ratio, $AET = \hat{\beta}_1^F/(\hat{\beta}_1^R - \hat{\beta}_1^F)$. Intuitively, a higher absolute value for AET suggests that the additional control variables included in the augmented model, relative to the restricted one, are not sufficient to explain away the estimated coefficient on population diversity in the full specification, and as such, this coefficient cannot be completely attributed to omitted-variable bias unless the amount of selection on unobservables is much larger than that on observables.

The analysis additionally considers the δ and β^* statistics suggested by Oster (2019). The δ statistic reflects how strongly correlated the unobservables need to be with population diversity,

[52] Despite the fact that the measure of conflict intensity from the MEPV data set is ordinal rather than continuous in nature, the analysis pursues least-squares (as opposed to maximum-likelihood) estimation methods when examining this particular outcome variable, primarily because this permits the implementation of both of the key identification strategies. Specifically, although the main findings from Columns 1–2 can be qualitatively replicated using ordered probit rather than OLS regressions (results not shown), the absence of a readily available IV counterpart of the ordered probit regression model precludes conducting a similar robustness check on the main findings from Columns 3–4.

[53] Altonji et al. (2005) develop this method for the case where the explanatory variable of interest is binary in nature, while Bellows and Miguel (2009) consider the case of a continuous explanatory variable. Roughly speaking, the assumption in assessments of this type is that the covariation of the outcome variable with observables, on the one hand, and its covariation with unobservables, on the other, are identically related to the explanatory variable of interest. Altonji et al. (2005) provide some sufficient conditions for such an assumption to hold.

TABLE A.II: Population Diversity and the Count of Civil Conflict Onsets across Countries

Cross-country sample:	Global					Old World		Global	
	(1) Negative Binomial	(2) Negative Binomial	(3) Negative Binomial	(4) Negative Binomial	(5) Negative Binomial	(6) Negative Binomial	(7) Negative Binomial	(8) Poisson	(9) Poisson
	Total count of new PRIO25 civil conflict onsets, 1960–2017								
Population diversity (ancestry adjusted)	10.032***	19.339***	13.092**	14.180***	12.884***	17.968***	18.025***	13.592**	12.884***
	(3.878)	(3.559)	(5.238)	(5.232)	(4.674)	(6.045)	(5.358)	(5.512)	(4.674)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	150	150	150	150	147	123	121	150	147
Pseudo R^2	0.013	0.128	0.153	0.158	0.257	0.149	0.276	0.219	0.317
Marginal effect of diversity	0.114**	0.220***	0.149**	0.162**	0.147**	0.231***	0.231***	0.155**	0.147**
	(0.046)	(0.051)	(0.064)	(0.065)	(0.058)	(0.086)	(0.075)	(0.068)	(0.058)

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to considering the *total count* rather than the annual frequency of civil conflict onsets over the post-1960 time period as the outcome variable. In line with the standard for analyzing over-dispersed count data, the regressions are estimated using the negative-binomial as opposed to a least-squares estimator. Given the absence of a negative-binomial estimator that permits instrumentation, however, the current analysis is unable to implement the strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. Thus, in lieu of implementing the instrument-based identification strategy in the global sample of countries, Columns 8–9 examine robustness to employing the Poisson rather than the negative-binomial estimator for estimating the specifications from Columns 6–7, respectively. The specifications examined in this table are otherwise identical to corresponding OLS specifications reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis. The estimated marginal effect of a 1 percentage point increase in population diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in the total number of new conflict onsets over the post-1960 time period. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

relative to observables, in order to account for the full size of the coefficient on population diversity. It differs from AET by accounting for the empirical *relevance* of the observables in explaining the variation in the outcome variable, based on the idea that including observables that do not move the R^2 statistic of the regression very much leaves more room for unobservables that are correlated with the variable of interest. The β^* statistic reflects the estimated value of the coefficient on population diversity if unobservables were as correlated with population diversity as the observables. Oster (2019) shows that if zero does not belong to the interval between the estimated coefficient on population diversity and β^*, then one can reject the null hypothesis that the coefficient of interest is exclusively driven by unobservables.

The analysis treats the specification from Column 3 of Table I as the restricted model. This specification includes, besides population diversity, the baseline geographical controls and continent fixed effects. Coefficient stability is assessed relative to the augmented specification presented in Column 8 that includes the full set of control variables. The resulting AET ratio is -10.3, and it suggests that selection on unobservables would have to be at least ten times larger than the selection on observables to account for the full size of the estimated coefficient for population diversity.[54] On the other hand, Oster's δ statistic is 1.93, indicating that the correlation of unobservables with population diversity needs to be almost twice as large as the correlation of population diversity with observables in order to drive the estimate down to zero. Assuming that the unobservables are equally correlated with population diversity as are the observables, and that these correlations have

[54]The negative sign indicates that selection on unobservables needs to move the coefficient estimate in the opposite direction, compared to selection on observables.

TABLE A.III: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Spatial Dependence

Cross-country sample:	Global					Old World		Global	
	(1) SARAR OLS	(2) SARAR OLS	(3) SARAR OLS	(4) SARAR OLS	(5) SARAR OLS	(6) SARAR OLS	(7) SARAR OLS	(8) SARAR 2SLS	(9) SARAR 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.253**	0.447***	0.320***	0.329***	0.288**	0.717***	0.643***	0.602***	0.457***
	(0.099)	(0.109)	(0.120)	(0.121)	(0.130)	(0.251)	(0.223)	(0.219)	(0.175)
Spatial lag AR(1) of conflict (λ)	−0.633	−0.164	−0.226	−0.214	0.362	−1.123	−0.199	−0.851	0.317
	(1.078)	(0.750)	(0.750)	(0.729)	(0.761)	(0.833)	(0.772)	(0.849)	(0.748)
Spatial lag AR(1) of error (ρ)	0.177	0.579	0.629	0.328	0.470	1.103	0.963	1.115	0.346
	(0.814)	(0.846)	(0.840)	(0.842)	(0.798)	(0.817)	(0.669)	(0.821)	(0.744)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	150	150	150	150	147	123	121	150	147
Effect of 10th–90th %ile move in diversity	0.017**	0.030***	0.021***	0.022***	0.020**	0.035***	0.028***	0.040***	0.031***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.009)	(0.012)	(0.010)	(0.015)	(0.012)

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to accounting for spatial dependence across observations by estimating spatial-autoregressive models with spatial-autoregressive disturbances (SARAR(1,1)) using a generalized spatial two-stage least-squares (GS2SLS) estimator (e.g., Drukker et al., 2013). To perform this robustness check, which involves the estimation of the AR(1) coefficients, λ and ρ, respectively associated with the spatial lags in the outcome variable and the error term, the estimator exploits an inverse-distance spatial weighting matrix for the regression sample, based on the great-circle distances between the geodesic centroids of country pairs. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions in the last two columns. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

the same sign, the estimated coefficient for diversity, if one were able to control for all unobservables, would be $\beta^* = 1.15$. Thus, the interval between the actual coefficient estimate from the full specification (0.309) and β^* excludes zero.[55] It is therefore rather unlikely that the main results could be explained away by omitted variables.

Robustness to Examining the Count of Civil Conflict Onset across Countries Given that the baseline cross-country regressions employ least-squares estimation, a log transformation is applied to the outcome variable in order to partly address the issue that its cross-country distribution is positively skewed with excess zeros, arising from the fact that new civil conflict onsets are generally rare events in cross-sectional data. An alternative approach to this issue, however, is to employ an estimation method that is tailored to the analysis of over-dispersed count data. The analysis in Table A.II considers the *total count* rather than the annual frequency of civil conflict onsets over the 1960–2017 time period as the outcome variable. The regressions in Columns 1–7 are estimated using the negative-binomial (as opposed to a least-squares) estimator to account for over-dispersion. Given the absence of a negative-binomial estimator that permits instrumentation, in lieu of implementing the instrument-based identification strategy in the global sample of countries, Columns 8–9 examine robustness to employing the Poisson rather than the negative binomial-

[55]The reported Oster statistics are computed under the most conservative assumption that $R^2_{max} = 1$; i.e., that the entire cross-country variation in conflict frequency would be explained by the estimated model if one could include all unobservables correlated with population diversity to the model.

TABLE A.IV: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Population Diversity as a Generated Regressor

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.209***	0.439***	0.306***	0.318***	0.309**	0.548***	0.597***	0.537***	0.602***
	(0.066)	(0.103)	(0.118)	(0.123)	(0.138)	(0.189)	(0.227)	(0.184)	(0.223)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	150	150	150	150	147	123	121	150	147
Adjusted R^2	0.029	0.189	0.213	0.215	0.358	0.225	0.392		
Effect of 10th–90th %ile move in diversity	0.014***	0.029***	0.020**	0.021**	0.021**	0.026**	0.026**	0.036***	0.041**
	(0.005)	(0.007)	(0.008)	(0.009)	(0.010)	(0.011)	(0.012)	(0.013)	(0.016)

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness of the standard-error estimates to accounting for the fact that the country-level measure of contemporary population diversity is a generated regressor in the empirical specifications, because it is projected from implicit zeroth-stage relationships (a) between prehistoric migratory distance from East Africa and expected heterozygosity in the HGDP-CEPH sample of 53 ethnic groups, and (b) between pairwise migratory distance and pairwise F_{ST} genetic distance across all pairs of ethnic groups in this sample. To perform this robustness check, the current analysis adopts the two-step bootstrapping technique implemented by Ashraf and Galor (2013a) for computing the standard-error estimates, so the reader is referred to that work for additional details on the technique. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions in the last two columns. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Bootstrapped standard errors, accounting for the use of a generated regressor, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

estimator in the global sample of countries. To interpret the influence of population diversity, the estimate in Column 7 suggests that conditional on the full set of control variables, a 5 percentage point increase in population diversity translates roughly into an additional civil conflict amongst countries in the Old World during the 1960-2017 time period.

Robustness to Accounting for Spatial Dependence To account for spatial dependence across country observations, the analysis in Table A.III replicates the key specifications from Table I using spatial-autoregressive models with spatial-autoregressive disturbances (SARAR(1,1)), estimated by a generalized spatial two-stage least-squares (GS2SLS) estimator (e.g., Drukker et al., 2013). These spatial regressions involve the estimation of AR(1) coefficients, λ and ρ, that are respectively associated with the spatial lags in the outcome variable and the error term. To perform this robustness check, the estimator exploits an inverse-distance spatial weighting matrix for the regression sample, based on the great-circle distances between the geodesic centroids of country pairs. Reassuringly, all of the main cross-country findings remain qualitatively intact, indicating that spatial dependence across country observations is not a confounding issue.

Robustness to Accounting for Population Diversity as a Generated Regressor The measure of contemporary population diversity is a generated regressor in the main specifications, because it is projected from implicit zeroth-stage relationships (i) between prehistoric migratory distance from East Africa and expected heterozygosity in the HGDP-CEPH sample of 53 ethnic groups, and (ii) between pairwise migratory distance and pairwise F_{ST} genetic distance across all pairs of ethnic groups in this sample. Table A.IV therefore checks the robustness of the standard-error estimates to accounting for potential bias due to the use of a generated regressor. To perform

TABLE A.V: Population Diversity and the Incidence of Civil Conflict in Repeated Cross-Country Data – Robustness to Examining Alternative Measures of Conflict Incidence

Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	Quinquennial PRIO1000 civil war incidence, 1960–2017				Quinquennial UCDP nonstate conflict incidence, 1989–2017			
Population diversity (ancestry adjusted)	16.221***	11.251**	17.090***	16.327***	24.499***	25.186***	22.511***	24.662***
	(4.285)	(5.482)	(4.256)	(5.808)	(5.399)	(6.408)	(4.992)	(5.563)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	1,270	1,026	1,551	1,262	717	670	879	824
Countries	123	121	147	144	123	121	150	147
Pseudo R^2	0.392	0.390			0.436	0.459		
Marginal effect of diversity	1.850***	1.212**	2.005***	1.786**	3.835***	3.568***	3.790***	3.839***
	(0.540)	(0.617)	(0.631)	(0.777)	(0.837)	(0.927)	(0.911)	(1.013)
First-stage F statistic			168.723	113.194			148.632	120.800

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on the quinquennial incidence of intrastate conflict in repeated cross-country data, as shown in Columns 1–4 of Table II. Specifically, it establishes robustness to considering the temporal incidence of alternative forms of intrastate conflict as the outcome variable, including the incidence of (i) a high-intensity PRIO1000 civil war in any given 5-year interval during the 1960–2017 time period (Columns 1–4); and (ii) a low-intensity conflict involving nonstate actors in any given 5-year interval during the 1989–2017 time period (Columns 5–8). The specifications examined in this table are otherwise identical to corresponding ones reported in Columns 1–4 of Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the incidence of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

this robustness check, the analysis replicates the key specifications from Table I, adopting the two-step bootstrapping technique implemented by Ashraf and Galor (2013a) for estimating the standard errors. The reader is referred to that work for additional details on the technique. As expected, the bootstrapped standard errors are indeed somewhat larger than their robust counterparts from Table I, but reassuringly, the statistical significance of the coefficients on population diversity remain unaffected.

Robustness to Examining Alternative Measures of Conflict Incidence As shown in Columns 1–4 of Table II, population diversity is positively and significantly associated with the quinquennial incidence of a PRIO25 civil conflict (with at least 25 battle-related deaths in a year) in the post-1960 time period. The analysis in Table A.V examines whether the same result holds when considering the temporal incidence of alternative forms of intrastate conflict as the outcome variable, including the incidence in any given 5-year interval of (i) a high-intensity PRIO1000 civil war (with at least 1000 battle-related deaths in a year) during the 1960–2017 time period (Columns 1–4); and (ii) a low-intensity conflict (with at least 25 battle-related deaths in a year) involving only nonstate actors during the 1989–2017 time period (Columns 5–8). The findings indicate that regardless of the covariates included in the specification or the identification strategy exploited, population diversity exerts a positive and significant influence on the quinquennial incidence of either of the aforementioned types of intrastate conflict. To interpret the coefficient of interest, the IV probit

TABLE A.VI: Population Diversity and the Incidence of Civil Conflict in Repeated Cross-Country Data – Robustness to Examining the Annual Incidence or Quinquennial Prevalence of Conflict

Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) OLS	(6) OLS	(7) 2SLS	(8) 2SLS
	Annual PRIO25 civil conflict incidence, 1960–2017				Quinquennial PRIO25 civil conflict prevalence, 1960-2017			
Population diversity (ancestry adjusted)	9.301***	9.763***	10.762***	12.848***	1.710***	1.737***	1.773***	1.988***
	(3.015)	(3.203)	(3.121)	(3.914)	(0.558)	(0.637)	(0.565)	(0.716)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	6,280	5,221	7,801	6,569	1,270	1,045	1,583	1,311
Countries	123	121	150	147	123	121	150	147
Pseudo R^2	0.597	0.602						
Adjusted R^2					0.621	0.598		
Marginal effect of diversity	0.976***	0.973***	1.125***	1.297***				
	(0.329)	(0.339)	(0.367)	(0.463)				
Effect of 10th–90th %ile move in diversity					0.080***	0.078***	0.115***	0.132***
					(0.026)	(0.028)	(0.037)	(0.047)
First-stage F statistic			155.509	103.745			151.471	104.807

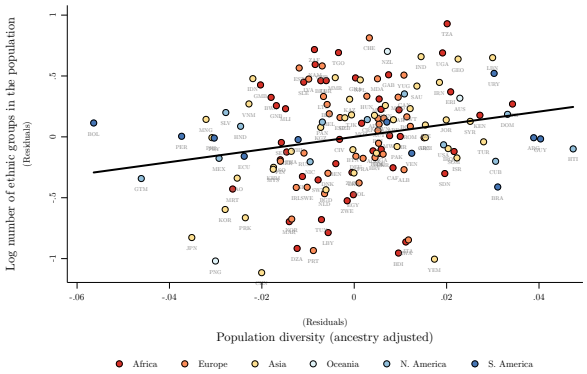
Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on the temporal incidence or prevalence of civil conflict in repeated cross-country data, as shown in Columns 1–4 of Table II. Specifically, it establishes robustness to considering (i) the annual incidence of conflict, by examining annual rather than quinquennial repetitions of the cross-section (Columns 1–4); and (ii) the quinquennial prevalence of conflict, by examining the share of years with an active civil conflict in any given 5-year interval (Columns 5–8). The specifications examined in this table are essentially identical to corresponding ones reported in Columns 1–4 of Table II, with the exception that in Columns 1–4 of the current analysis, the time-dependent baseline controls for institutions (i.e., executive constraints, indicators for the type of political regime, and indicators for colonial experience by identity of the colonizing power), total population, GDP per capita, and temporal spillovers are all appropriately adjusted to assume their respective lagged annual values, rather than their values corresponding to the previous 5-year interval. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the IV probit or 2SLS regressions. In Columns 1–4, the estimated marginal effect of a 1 percentage point increase in population diversity is the average marginal effect across the entire cross-section of observed diversity values, and it reflects the increase in the annual likelihood of a conflict incidence, expressed in percentage points. In Columns 5–8, the estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the share of years with an active conflict in any given 5-year interval. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

regressions presented in Columns 4 and 8 suggest that conditional on the full set of control variables, a 1 percentage point increase in population diversity increases the quinquennial likelihoods of conflict incidence by 1.8 percentage points for PRIO1000 civil wars and by 3.8 percentage points for internal conflicts involving nonstate actors.

Robustness to Examining the Annual Incidence or Quinquennial Prevalence of Civil Conflict The analysis in Table A.VI checks the robustness of the baseline results for the incidence of civil conflict, as shown in Columns 1–4 of Table II, to considering alternative outcomes of conflict incidence or prevalence, including (i) the annual incidence of conflict, by examining annual rather than quinquennial repetitions of the cross-section (Columns 1–4); and (ii) the quinquennial prevalence of conflict, by examining the share of years with an active civil conflict in any given 5-year interval (Columns 5–8). The specifications examined in this table are essentially identical to corresponding ones reported in Columns 1–4 of Table II, with the exception that in Columns 1–

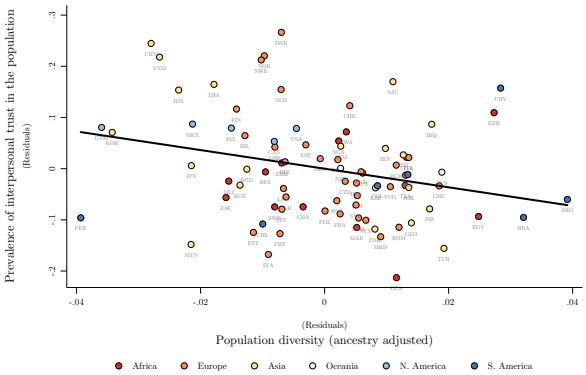
4 of the current analysis, the time-dependent baseline controls for institutions (i.e., executive constraints, indicators for the type of political regime, and indicators for colonial experience by identity of the colonizing power), total population, GDP per capita, and temporal spillovers are all appropriately adjusted to assume their respective lagged annual values, rather than their values corresponding to the previous 5-year interval. As is evident from the results in Table A.VI, regardless of the identification strategy exploited or the covariates included in the specification, population diversity contributes positively and significantly to both the annual incidence and the quinquennial prevalence of civil conflict during the 1960–2017 time period. Specifically, the global average marginal effect estimated by the specification in Column 4 suggests that conditional on the full set of control variables, a 1 percentage point increase in population diversity increases the annual likelihood of a conflict incidence by 1.3 percentage points. Further, the specification in Column 8 suggests that conditional on all covariates, a move from the 10th to the 90th percentile of the global cross-country distribution of population diversity is associated with an increase of 13 percentage points in the fraction of years with an active conflict in any given 5-year interval.

A.3 Appendix Figures for the Country-Level Analyses



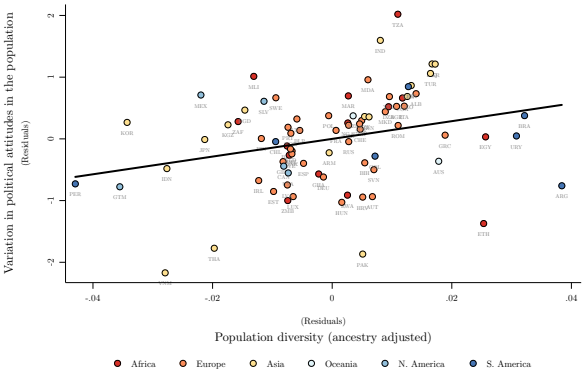
Relationship in the global sample; conditional on baseline geographical controls and continent fixed effects
Slope coefficient = 5.187; (robust) standard error = 1.801; t-statistic = 2.881; partial R-squared = 0.049; observations = 147

(a) Number of ethnic groups



Relationship in the global sample; conditional on baseline geographical controls and region fixed effects
Slope coefficient = -1.817; (robust) standard error = 0.795; t-statistic = -2.286; partial R-squared = 0.075; observations = 84

(b) Prevalence of interpersonal trust



Relationship in the global sample; conditional on baseline geographical controls and region fixed effects
Slope coefficient = 14.344; (robust) standard error = 6.238; t-statistic = 2.299; partial R-squared = 0.082; observations = 81

(c) Variation in political attitudes

FIGURE A.1: Population Diversity and Proximate Determinants of the Frequency of Civil Conflict Onset across Countries

Notes: This figure depicts the global cross-country relationship between contemporary population diversity and each of three potentially conflict-augmenting proximate channels, including (i) the degree of cultural fragmentation, as reflected by the number of ethnic groups in the national population (Panel (a)); (ii) the prevalence of generalized interpersonal trust at the country level (Panel (b)); and (iii) the extent of heterogeneity in preferences for redistribution and public-goods provision, as reflected by the intra-country dispersion in individual political attitudes on a politically "left"–"right" categorical scale (Panel (c)), conditional on the baseline geographical correlates of conflict, as considered by the analysis in Table VII. Each of Panels (a), (b), and (c) presents an added-variable plot with a partial regression line, corresponding to the estimated coefficient associated with population diversity in Columns 1, 4, and 7, respectively, of Table VII.

A.4 Descriptive Statistics at the Country Level

TABLE A.VII: Summary Statistics of Variables from the Baseline Cross-Country Analysis

			Percentile	
	Mean	SD	10th	90th
PANEL A	Old World sample ($N = 121$)			
New civil conflict onsets per year, 1960–2017	0.025	0.033	0.000	0.069
Population diversity (ancestry adjusted)	0.735	0.018	0.712	0.754
Migratory distance from East Africa (in 10,000 km)	0.515	0.244	0.262	0.831
Absolute latitude	0.029	0.017	0.007	0.052
Ruggedness	0.124	0.134	0.016	0.286
Mean elevation	0.610	0.584	0.106	1.265
Range of elevation	1.550	1.322	0.281	3.043
Mean land suitability	0.359	0.234	0.035	0.669
Range of land suitability	0.701	0.259	0.345	0.974
Distance to nearest waterway	0.383	0.483	0.039	1.036
Island nation dummy	0.033	0.180	0.000	0.000
Ethnic fractionalization	0.476	0.264	0.115	0.812
Ethnolinguistic polarization	0.491	0.220	0.181	0.747
Ever a U.K. colony dummy	0.264	0.443	0.000	1.000
Ever a French colony dummy	0.207	0.407	0.000	1.000
Ever a non-U.K./non-French colony dummy	0.198	0.400	0.000	1.000
British legal origin dummy	0.256	0.438	0.000	1.000
French legal origin dummy	0.405	0.493	0.000	1.000
Executive constraints, 1960–2017 average	3.983	1.875	1.684	7.000
Fraction of years under democracy, 1960–2017	0.367	0.381	0.000	1.000
Fraction of years under autocracy, 1960–2017	0.390	0.327	0.000	0.900
Oil or gas reserve discovery	0.669	0.472	0.000	1.000
Log population, 1960–2017 average	16.072	1.459	14.385	17.873
Log GDP per capita, 1960–2017 average	7.638	1.567	5.649	9.940
PANEL B	Global sample ($N = 147$)			
New civil conflict onsets per year, 1960–2017	0.022	0.031	0.000	0.064
Population diversity (ancestry adjusted)	0.728	0.027	0.685	0.752
Migratory distance from East Africa (in 10,000 km)	0.806	0.679	0.295	2.088
Absolute latitude	0.027	0.017	0.006	0.051
Ruggedness	0.125	0.126	0.018	0.278
Mean elevation	0.594	0.552	0.104	1.250
Range of elevation	1.701	1.389	0.283	3.752
Mean land suitability	0.386	0.246	0.046	0.718
Range of land suitability	0.715	0.264	0.317	0.994
Distance to nearest waterway	0.353	0.458	0.036	1.010
Island nation dummy	0.048	0.214	0.000	0.000
Ethnic fractionalization	0.467	0.254	0.115	0.792
Ethnolinguistic polarization	0.452	0.241	0.097	0.747
Ever a U.K. colony dummy	0.259	0.439	0.000	1.000
Ever a French colony dummy	0.190	0.394	0.000	1.000
Ever a non-U.K./non-French colony dummy	0.320	0.468	0.000	1.000
British legal origin dummy	0.252	0.435	0.000	1.000
French legal origin dummy	0.463	0.500	0.000	1.000
Executive constraints, 1960–2017 average	4.145	1.827	1.839	7.000
Fraction of years under democracy, 1960–2017	0.408	0.377	0.000	1.000
Fraction of years under autocracy, 1960–2017	0.352	0.323	0.000	0.879
Oil or gas reserve discovery	0.673	0.471	0.000	1.000
Log population, 1960–2017 average	16.087	1.431	14.423	17.877
Log GDP per capita, 1960–2017 average	7.703	1.489	5.705	9.937

A.5 Robustness Checks for the Ethnicity-Level Analyses

TABLE A.VIII: Population Diversity and the Number of Conflicts across Ethnic Homelands

	Number of conflict events							
	(1) Poisson	(2) Poisson	(3) Poisson	(4) Poisson	(5) Poisson	(6) Poisson	(7) Poisson	(8) Poisson
Observed population diversity	58.949***	54.959***	63.413***	61.608***				
	(18.286)	(18.020)	(16.750)	(16.685)				
Predicted population diversity					53.230***	43.748***	49.242***	51.593***
					(9.061)	(6.670)	(7.403)	(7.799)
Ethnolinguistic fractionalization			0.236				-0.784**	
			(0.498)				(0.356)	
Ethnolinguistic polarization				-0.190				-0.895**
				(0.430)				(0.441)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Climatic controls	No	No	Yes	Yes	No	No	Yes	Yes
Development outcomes	No	No	Yes	Yes	No	No	Yes	Yes
Disease environment controls	No	No	Yes	Yes	No	No	Yes	Yes
Sample	Observed	Observed	Observed	Observed	Predicted	Predicted	Predicted	Predicted
Observations	207	207	207	207	901	901	901	901
PseudoR^2	0.250	0.327	0.463	0.463	0.215	0.451	0.519	0.520
Effect of 10th-90th %ile move in diversity	61.225***	57.081***	65.859***	63.985***	60.379***	49.623***	55.855***	58.521***
	(22.925)	(21.677)	(21.558)	(21.168)	(15.693)	(10.515)	(11.546)	(12.213)

Notes: This table exploits variations across ethnic homelands to establish a significant positive reduced-form impact of contemporary population diversity on the number of conflict events during the 1989–2008 period, conditional on the baseline control variables (i.e., proximate geographical and development-related correlates of conflict). The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the number of conflict events. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE A.IX: Population Diversity and Alternative Conflict Outcomes across Ethnic Homelands

	Log number of conflicts		Log number of deaths		Log number of deaths per conflict	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Observed population diversity	6.037***		26.119***		20.082**	
	(2.284)		(9.789)		(7.792)	
Predicted population diversity		9.173***		40.406***		31.233***
		(1.918)		(8.581)		(6.932)
Ethnolinguistic fractionalization	0.552*	0.094	3.152**	0.576	2.600**	0.482
	(0.317)	(0.113)	(1.421)	(0.492)	(1.162)	(0.398)
Ethnolinguistic polarization	-0.439*	-0.171*	-2.489**	-0.758*	-2.050**	-0.587*
	(0.255)	(0.092)	(1.221)	(0.397)	(1.006)	(0.318)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Observed	Predicted	Observed	Predicted	Observed	Predicted
Observations	207	901	207	901	207	901
Effect of 10th-90th %ile move in diversity	1.319***	2.232***	9969.713***	10211.867***	948.445***	1008.051***
	(0.499)	(0.467)	(3736.546)	(2168.784)	(368.027)	(223.744)
Adjusted R^2	0.201	0.300	0.241	0.275	0.241	0.253

Notes: This table exploits variations across ethnic homelands to establish a significant positive impact of contemporary population diversity, predicted by prehistoric migratory distance from East Africa on the log number of UCDP/GED conflicts, the log number of UCDP/GED deaths, and the log number of UCDP/GED deaths per conflict, during the 1989–2008 period, accounting for geographical and development-related correlates of conflict. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The estimated effects associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution are expressed in terms of the non-logged levels of the respective outcome variables. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE A.X: Observed Population Diversity and Conflict across Ethnic Homelands – Robustness to Accounting for Spatial Dependence

	Log conflict prevalence						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS
Observed population diversity	31.788***	41.070***	37.111***	37.333***	37.148***	41.745***	41.403***
	(8.819)	(8.392)	(8.261)	(8.203)	(8.222)	(8.428)	(8.439)
Ethnolinguistic fractionalization				0.881*		0.804	
				(0.504)		(0.497)	
Ethnolinguistic polarization					0.593		0.562
					(0.426)		(0.417)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	No	No	Yes	Yes	Yes	Yes	Yes
Development outcomes	No	No	No	No	No	Yes	Yes
Disease environment controls	No	No	No	No	No	Yes	Yes
Sample	Observed	Observed	Observed	Observed	Observed	Observed	Observed
Direct impact of genetic diversity	32.803***	43.792***	38.509***	38.722***	38.550***	43.734***	43.391***
	(9.165)	(9.362)	(8.756)	(8.691)	(8.717)	(9.165)	(9.180)
Direct effect of 10th-90th %ile move in diversity	0.513***	0.685***	0.602***	0.605***	0.603***	0.684***	0.678***
	(0.143)	(0.146)	(0.137)	(0.136)	(0.136)	(0.143)	(0.144)
Observations	207	207	207	207	207	207	207

Notes: This table exploits variations across ethnic homelands to establish a significant positive reduced-form impact of contemporary population diversity on the log conflict prevalence during the 1989–2008 period, conditional on the baseline control variables (i.e., proximate geographical and development-related correlates of conflict) and accounting for spatial dependence using a spatial autoregressive (SARAR(1,1)) model, with a spectral-normalized inverse-distance weighting matrix, estimated with maximum-likelihood estimation, with a spatial lag of the dependent variable and a spatially lagged error. The model treat errors as heteroskedastic. Variables relating to observations associated with the same homeland polygon are averaged and a single observation is kept for each polygon. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the prevalence of conflicts within the territory of a homeland over the years 1989–2008. Standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE A.XI: Predicted Population Diversity and Conflict across Ethnic Homelands – Robustness to Accounting for Spatial Dependence

	Log conflict prevalence						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS
Predicted population diversity	57.609***	87.327***	88.759***	88.623***	86.281***	83.671***	84.245***
	(6.447)	(7.269)	(7.230)	(7.387)	(7.305)	(7.255)	(7.288)
Ethnolinguistic fractionalization				0.517**		0.332	
				(0.216)		(0.215)	
Ethnolinguistic polarization					0.007		-0.004
					(0.189)		(0.188)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	No	No	Yes	Yes	Yes	Yes	Yes
Development outcomes	No	No	No	No	No	Yes	Yes
Disease environment controls	No	No	No	No	No	Yes	Yes
Sample	Predicted	Predicted	Predicted	Predicted	Predicted	Predicted	Predicted
Direct Impact of Genetic Diversity	60.406***	87.543***	87.488***	79.433***	84.390***	81.962***	82.463***
	(6.930)	(7.509)	(7.510)	(8.977)	(7.690)	(7.597)	(7.643)
Effect of 10th-90th %ile move in diversity	1.341***	1.943***	1.942***	1.763***	1.873***	1.819***	1.830***
	(0.154)	(0.167)	(0.167)	(0.199)	(0.171)	(0.169)	(0.170)
Observations	901	901	901	901	901	901	901

Notes: This table exploits variations across ethnic homelands to establish a significant positive reduced-form impact of predicted population diversity on the log conflict prevalence during the 1989–2008 period, conditional on the baseline control variables (i.e., proximate geographical and development-related correlates of conflict) and accounting for spatial dependence using a spatial autoregressive (SARAR(1,1)) model, with a spectral-normalized inverse-distance weighting matrix, estimated with maximum-likelihood estimation, with a spatial lag of the dependent variable and a spatially lagged error. The model treats errors as heteroskedastic. Variables relating to observations associated with the same homeland polygon are averaged and a single observation is kept for each polygon. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the prevalence of conflicts within the territory of a homeland over the years 1989–2008. Standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE A.XII: Predicted Population Diversity and Conflict across Ethnic Homelands – Robustness to Accounting for Predicted Diversity as a Generated Regressor

	Log conflict prevalence			
	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Predicted population diversity	77.710***	77.031***	74.010***	73.581***
	(6.279)	(7.282)	(7.396)	(7.418)
Ethnolinguistic fractionalization			0.347	
			(0.299)	
Ethnolinguistic polarization				0.457*
				(0.263)
Regional dummies	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes
Climatic controls	No	Yes	Yes	Yes
Development outcomes	No	No	Yes	Yes
Disease environment controls	No	No	Yes	Yes
Sample	Predicted	Predicted	Predicted	Predicted
Observations	901	901	901	901
Effect of 10th-90th %ile move in diversity	1.725***	1.710***	1.643***	1.633***
	(0.139)	(0.162)	(0.164)	(0.165)
Adjusted R^2	0.211	0.362	0.378	0.379
Bootstrapped standard error	(7.128)***	(8.196)***	(8.244)***	(8.266)***

Notes: This table exploits variations across ethnic homelands to establish a significant positive impact of predicted population diversity on the log conflict prevalence during the 1989–2008 period, conditional on ecological diversity and ecological polarization as well as the baseline control variables. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. To perform this robustness check, the current analysis adopts the two-step bootstrapping technique implemented by Ashraf and Galor (2013a) for computing the standard-error estimates, so the reader is referred to that work for additional details on the technique. The specifications examined in this table are otherwise identical to corresponding ones reported in Table VI. The reader is therefore referred to Table VI and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions in the last two columns. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the prevalence of conflicts within the territory of a homeland over the years 1989–2008. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

A.6 Descriptive Statistics at the Ethnicity Level

TABLE A.XIII: Summary Statistics

			Percentile	
	Mean	SD	10th	90th
PANEL A	Observed population diversity sample ($N = 207$)			
Population diversity (observed)	0.72	0.05	0.65	0.76
Population diversity (predicted)	0.72	0.04	0.65	0.76
Conflict prevalence	0.14	0.27	0.00	0.63
Number of conflicts	1.04	2.78	0.00	3.00
Number of deaths (in thousands)	3.56	39.32	0.00	1.49
Ethnolinguistic fractionalization	0.26	0.30	0.00	0.74
Ethnolinguistic polarization	0.33	0.36	0.00	0.85
Absolute latitude	15.15	15.09	1.85	38.02
Ruggedness	133.37	144.14	14.69	299.79
Elevation	0.75	0.75	0.07	1.67
Range of elevation	1.60	1.25	0.31	3.36
Mean land suitability	8.50	3.50	3.69	12.38
Range of land suitability	5.09	4.42	0.36	11.76
Small island dummy	0.01	0.10	0.00	0.00
Distance to nearest waterway	56.45	60.96	0.00	140.47
Temperature	21.08	7.79	8.94	27.20
Precipitation	123.06	100.31	31.34	285.66
Years since settlement (centuries from present)	104.94	31.86	40.19	120.19
Malaria	0.16	0.19	0.00	0.49
Oil or gas discovery	0.27	0.45	0.00	1.00
Luminosity	1.20	2.95	0.00	3.70
PANEL B	Predicted population diversity sample ($N = 901$)			
Population diversity (predicted)	0.71	0.04	0.64	0.75
Conflict prevalence	0.19	0.32	0.00	0.76
Number of conflicts	1.13	4.30	0.00	3.00
Number of deaths (in thousands)	2.22	20.67	0.00	1.62
Ethnolinguistic fractionalization	0.49	0.28	0.02	0.83
Ethnolinguistic polarization	0.55	0.28	0.04	0.87
Absolute latitude	21.69	17.08	2.92	48.23
Ruggedness	172.23	176.69	16.32	403.90
Elevation	0.73	0.86	0.07	1.75
Range of elevation	1.84	1.37	0.34	3.69
Mean land suitability	8.24	3.61	2.09	12.21
Range of land suitability	5.56	4.64	0.55	13.25
Small island dummy	0.03	0.16	0.00	0.00
Distance to nearest waterway	43.72	56.33	0.00	94.88
Temperature	18.82	9.36	3.83	26.67
Precipitation	118.85	75.53	32.58	225.72
Years since settlement (centuries from present)	112.52	23.61	90.19	120.19
Malaria	0.10	0.15	0.00	0.37
Oil or gas discovery	0.35	0.48	0.00	1.00
Luminosity	1.47	3.69	0.00	3.76

References

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003): "Fractionalization," *Journal of Economic Growth*, 8, 155–194.

Alesina, A. and E. La Ferrara (2005): "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43, 762–800.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016): "Ethnic Inequality," *Journal of Political Economy*, 124, 428–488.

Alesina, A. and E. Spolaore (2003): *The Size of Nations*, Cambridge, MA: MIT Press.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151–184.

Ashraf, Q. and O. Galor (2013a): "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103, 1–48.

——— (2013b): "Genetic Diversity and the Origins of Cultural Fragmentation," *American Economic Review: Papers & Proceedings*, 103, 528–533.

Ashraf, Q. H. and O. Galor (2018): "The Macrogenoeconomics of Comparative Development," *Journal of Economic Literature*, 56, 1119–1155.

Banks, A. S. and K. A. Wilson (2018): "Cross-National Time-Series Data Archive [Data file]," Databanks International, Jerusalem, Israel. https://www.cntsdata.com/.

Bates, R. H. (1983): "Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa," in *State versus Ethnic Claims: African Policy Dilemmas*, ed. by D. Rothchild and V. A. Olorunsola, Boulder, CO: Westview Press, 152–171.

Bazzi, S. and C. Blattman (2014): "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, 6, 1–38.

Beck, N., J. N. Katz, and R. Tucker (1998): "Taking Time Seriously: Time-Series–Cross-Section Analysis with a Binary Dependent Variable," *American Journal of Political Science*, 42, 1260–1288.

Bellows, J. and E. Miguel (2009): "War and Local Collective Action in Sierra Leone," *Journal of Public Economics*, 93, 1144–1157.

Besley, T. and M. Reynal-Querol (2014): "The Legacy of Historical Conflict: Evidence from Africa," *American Political Science Review*, 108, 319–336.

Birnir, J. K., D. D. Laitin, J. Wilkenfeld, D. M. Waguespack, A. S. Hultquist, and T. R. Gurr (2018): "Introducing the AMAR (All Minorities at Risk) Data," *Journal of Conflict Resolution*, 62, 203–226.

Birnir, J. K., J. Wilkenfeld, J. D. Fearon, D. D. Laitin, T. R. Gurr, D. Brancati, S. M. Saideman, A. Pate, and A. S. Hultquist (2015): "Socially Relevant Ethnic Groups, Ethnic Structure, and AMAR," *Journal of Peace Research*, 52, 110–115.

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

Brecke, P. (1999): "Violent Conflicts 1400 A.D. to the Present in Different Regions of the World," Paper presented at the 1999 Annual Meeting of the Peace Science Society, October 8–10.

Caselli, F. and W. J. Coleman, II (2013): "On the Theory of Ethnic Conflict," *Journal of the European Economic Association*, 11, 161–192.

Cassar, A., P. Grosjean, and S. Whitt (2013): "Legacies of Violence: Trust and Market Development," *Journal of Economic Growth*, 18, 285–318.

Cioffi-Revilla, C. (1996): "Origins and Evolution of War and Politics," *International Studies Quarterly*, 40, 1–22.

Collier, P. and A. Hoeffler (2004): "Greed and Grievance in Civil War," *Oxford Economic Papers*, 56, 563–595.

——— (2007): "Civil War," in *Handbook of Defense Economics, Vol. 2: Defense in a Globalized World*, ed. by T. Sandler and K. Hartley, Amsterdam, The Netherlands: Elsevier, North-Holland, 711–740.

Croicu, M. and R. Sundberg (2015): "UCDP Georeferenced Event Dataset Codebook version 4.0," Department of Peace and Conflict Research, Uppsala University. http://ucdp.uu.se/downloads/ged/ucdp-ged-40-codebook.pdf.

Desmet, K., M. L. Breton, I. Ortuño-Ortín, and S. Weber (2011): "The Stability and Breakup of Nations: A Quantitative Analysis," *Journal of Economic Growth*, 16, 183–213.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012): "The Political Economy of Linguistic Cleavages," *Journal of Development Economics*, 97, 322–338.

Desmet, K., I. Ortuño-Ortín, and S. Weber (2009): "Linguistic Diversity and Redistribution," *Journal of the European Economic Association*, 7, 1291–1318.

Dincecco, M., J. Fenske, and M. G. Onorato (2015): "Is Africa Different? Historical Conflict and State Development," IMT Lucca EIC Working Paper No. 08/2015, IMT Institute for Advance Studies Lucca.

Drukker, D. M., I. R. Prucha, and R. Raciborski (2013): "Maximum Likelihood and Generalized Spatial Two-Stage Least-Squares Estimators for a Spatial-Autoregressive Model with Spatial-Autoregressive Disturbances," *Stata Journal*, 13, 221–241.

Dube, O. and J. F. Vargas (2013): "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," *Review of Economic Studies*, 80, 1384–1421.

Easterly, W. and R. Levine (1997): "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 112, 1203–1250.

Eifert, B., E. Miguel, and D. N. Posner (2010): "Political Competition and Ethnic Identification in Africa," *American Journal of Political Science*, 54, 494–510.

Esteban, J., L. Mayoral, and D. Ray (2012): "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, 102, 1310–1342.

Esteban, J. and D. Ray (2011a): "A Model of Ethnic Conflict," *Journal of the European Economic Association*, 9, 496–521.

Fearon, J. D. (2003): "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 8, 195–222.

Fearon, J. D. and D. D. Laitin (2003): "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 97, 75–90.

Fletcher, E. and M. Iyigun (2010): "The Clash of Civilizations: A Cliometric Investigation," http://www.colorado.edu/economics/courses/iyigun/fractionalization013109.pdf.

Gellner, E. (1983): *Nations and Nationalism*, Ithaca, NY: Cornell University Press.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002): "Armed Conflict 1946-2001: A New Dataset," *Journal of Peace Research*, 39, 615–637.

Grossman, H. I. (1991): "A General Equilibrium Model of Insurrections," *American Economic Review*, 81, 912–921.

——— (1999): "Kleptocracy and Revolutions," *Oxford Economic Papers*, 51, 267–283.

Harpending, H. and A. Rogers (2000): "Genetic Perspectives on Human Origins and Differentiation," *Annual Review of Genomics and Human Genetics*, 1, 361–385.

Hirshleifer, J. (1991): "The Technology of Conflict as an Economic Activity," *American Economic Review: Papers & Proceedings*, 81, 130–134.

——— (1995): "Anarchy and its Breakdown," *Journal of Political Economy*, 103, 26–52.

Humphreys, M. (2005): "Natural Resources, Conflict, and Conflict Resolution: Uncovering the Mechanisms," *Journal of Conflict Resolution*, 49, 508–537.

King, G. and L. Zeng (2001): "Logistic Regression in Rare Events Data," *Political Analysis*, 9, 137–163.

König, M. D., D. Rohner, M. Thoenig, and F. Zilibotti (2017): "Networks in Conflict: Theory and Evidence From the Great War of Africa," *Econometrica*, 85, 1093–1132.

Marshall, M. G. (2017): "Major Episodes of Political Violence (MEPV) and Conflict Regions, 1946–2017," Center for Systemic Peace, Vienna, VA. Data retrieved at http://www.systemicpeace.org/inscrdata.html.

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

Miguel, E., S. Satyanath, and E. Sergenti (2004): "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 112, 725–753.

Montalvo, J. G. and M. Reynal-Querol (2005): "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 95, 796–816.

Nunn, N. and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101, 3221–3252.

Oster, E. (2019): "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37, 187–204.

Pemberton, T. J., M. DeGiorgio, and N. A. Rosenberg (2013): "Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation," *G3: Genes, Genomes, and Genetics*, 3, 891–907.

Pettersson, T. and K. Eck (2018): "Organized Violence, 1989–2017," *Journal of Peace Research*, 55, 535–547.

Posner, D. N. (2003): "The Colonial Origins of Ethnic Cleavages: The Case of Linguistic Divisions in Zambia," *Comparative Politics*, 35, 127–146.

Putterman, L. and D. N. Weil (2010): "Post-1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 125, 1627–1682.

Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman, and L. L. Cavalli-Sforza (2005): "Support from the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa," *Proceedings of the National Academy of Sciences*, 102, 15942–15947.

Reid, R. (2014): "The Fragile Revolution: Rethinking War and Development in Africa's Violent Nineteenth Century," in *Africa's Development in Historical Perspective*, ed. by E. Akyeampong, R. H. Bates, N. Nunn, and J. A. Robinson, Cambridge, UK: Cambridge University Press, 393–423.

Rohner, D., M. Thoenig, and F. Zilibotti (2013a): "War Signals: A Theory of Trade, Trust, and Conflict," *Review of Economic Studies*, 80, 1114–1147.

——— (2013b): "Seeds of Distrust: Conflict in Uganda," *Journal of Economic Growth*, 18, 217–252.

Ross, M. (2006): "A Closer Look at Oil, Diamonds, and Civil War," *Annual Review of Political Science*, 9, 265–300.

Rummel, R. J. (1963): "Dimensions of Conflict Behavior Within and Between Nations," *General Systems Yearbook*, 8, 1–50.

Sambanis, N. (2002): "A Review of Recent Advances and Future Directions in the Quantitative Literature on Civil War," *Defence and Peace Economics*, 13, 215–243.

Spolaore, E. and R. Wacziarg (2013): "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature*, 51, 325–369.

——— (2016): "War and Relatedness," *Review of Economics and Statistics*, 98, 925–939.

Sundberg, R., K. Eck, and J. Kreutz (2012): "Introducing the UCDP Non-State Conflict Dataset," *Journal of Peace Research*, 49, 351–362.

Tollefsen, A. F., H. Strand, and H. Buhaug (2012): "PRIO-GRID: A Unified Spatial Data Structure," *Journal of Peace Research*, 49, 363–374.

Weidmann, N. B., J. K. Rød, and L.-E. Cederman (2010): "Representing Ethnic Groups in Space: A New Dataset," *Journal of Peace Research*, 47, 491–499.

Wimmer, A. (2002): *Nationalist Exclusion and Ethnic Conflict: Shadows of Modernity*, Cambridge, UK: Cambridge University Press.

World Values Survey (2006): "European and World Values Surveys, Four-Wave Integrated Data File, 1981–2004, version 20060423," The World Values Survey Association, Stockholm, Sweden. Data retrieved at http://www.worldvaluessurvey.org.

——— (2009): "World Values Survey, 1981–2008 Official Aggregate, version 20090914," The World Values Survey Association, Stockholm, Sweden. Data retrieved at http://www.worldvaluessurvey.org.

Supplement to Diversity and Conflict

Supplement A Supplement to the Country-Level Analyses

A.1 Robustness Checks for the Analysis of Civil Conflict in Cross-Country Data

In this appendix section, we present several robustness checks for our cross-country analysis of the influence of contemporary population diversity on the temporal frequency of civil conflict outbreaks in the post-1960 time horizon.

Robustness to Accounting for Ecological/Climatic Covariates A nascent interdisciplinary literature (e.g., Burke et al., 2009; Hsiang et al., 2013; Burke et al., 2015) has emphasized the role of climatic factors, like temperature and precipitation, as important correlates of the risk of civil conflict. Further, Fenske (2014) shows that ecological diversity facilitated state centralization in pre-colonial Africa. To prevent our main specifications from becoming too unwieldy, we chose to exclude the aforementioned climatic and ecological variables from our baseline set of covariates, especially because this set already included a sizable vector of geographical factors that are known to be correlated with the former. In Table SA.I, however, we establish that population diversity remains a significant predictor of civil conflicts when we augment our baseline set of covariates in Table I with controls for (i) time-invariant fractionalization and polarization measures of the ecological diversity of land (e.g., Fenske, 2014); and (ii) the temporal mean and volatility of climatic experience (e.g., Burke et al., 2015) with respect to annual temperature and annual precipitation over the post-1960 time period.

Robustness to Accounting for Deep-Rooted Determinants of Economic Development In Table SA.II, we establish the robustness of our baseline cross-country analysis of civil conflict to *additionally* accounting for the potentially confounding influence of other deep-rooted determinants of comparative economic development. Specifically, we augment the analysis in Table I with controls for (i) the time elapsed since the onset of the Neolithic Revolution (e.g., Ashraf and Galor, 2013a); (ii) an index of experience with institutionalized statehood since antiquity (e.g., Bockstette et al., 2002); (iii) the time elapsed since initial human settlement in prehistory (e.g., Ahlerup and Olsson, 2012); and (iv) the great-circle distance to the closest regional technological frontier in the year 1500 (e.g., Ashraf and Galor, 2013a). The results indicate that regardless of the estimation sample or the specification, contemporary population diversity remains a significant predictor of the annual frequency of civil conflict onsets.

Robustness to Accounting for Ethnic and Spatial Inequality In Table SA.III, we check the robustness of our findings from Table I to *additionally* accounting for intra-country economic inequality (e.g., Alesina et al., 2016), as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across either (i) the georeferenced homelands of ethnic groups (ethnic inequality); or (ii) 2.5×2.5-degree geospatial grid cells (spatial inequality). The two inequality measures enter these regressions with a positive coefficient, and in at least one case, the coefficient on ethnic inequality is statistically significant. Nonetheless, our results indicate that the positive and significant influence of population diversity on the annual frequency of civil conflicts cannot be attributed to the potentially confounding influence of these inequality measures.

Robustness to Using Alternative Measures of Ethnolinguistic Fragmentation Due to the sizable cross-country correlation between the ethnic and linguistic fractionalization measures of Alesina et al. (2003), rather than exploiting both variables simultaneously, we chose to employ the more widely used of the two indices – namely, ethnic fractionalization – as one of the many

covariates in our baseline analysis of the influence of population diversity on civil conflict frequency. In Table SA.IV, we examine the sensitivity of our baseline findings from Table I to employing the *linguistic* fractionalization index of Alesina et al. (2003) in lieu of our baseline control for the *ethnic* fractionalization index from the same source. Furthermore, in Table SA.V, we examine the robustness of our baseline findings to employing the country-level counterparts of our measures of linguistic fractionalization and polarization from our analysis of conflicts at the ethnic homelands level. Specifically, these measures are constructed using georeferenced information on the spatial distribution of language homelands (from the World Language Mapping System [WLMS]) in combination with gridded population data, and they enter our regressions in Table SA.V in liue of our baseline controls for ethnic fractionalization from Alesina et al. (2003) and ethnolinguistic polarization from Desmet et al. (2012). Reassuringly, the results in Tables SA.IV–SA.V confirm that all our baseline findings regarding the significant influence of population diversity on the temporal frequency of civil conflict onsets remain qualitatively intact under these alternative controls for ethnolinguistic fragmentation.

Robustness to Using Initial Values of Time-Varying Covariates In Table SA.VI, we exploit the initial or year-1960 values of the time-dependent baseline controls employed by our analysis in Table I (i.e., the degree of executive constraints, indicators for democracy and autocracy, total population, and GDP per capita), rather than their respective temporal averages over the 1960–2017 time period. This robustness check is intended to examine whether our baseline estimates of the influence of population diversity in Table I could be explained away by the fact that the temporal averages of our time-varying controls over the entire sample period are likely to be more endogenous to the frequency of civil conflict onsets over the same period. Reassuringly, population diversity continues to remain a significant predictor of conflict frequency in these alternative specifications.

Robustness to Accounting for Spatial Autocorrelation in Errors As with any analysis that exploits spatial variations in cross-sectional data, autocorrelation in disturbance terms across observations could be biasing our estimates of the standard errors in our baseline cross-country regressions of conflict frequency. Table SA.VII therefore reports, for our key specifications from Table I, standard errors that are corrected for cross-sectional spatial dependence, using the methodology proposed by Conley (1999). To perform this robustness check, the spatial distribution of observations is specified on the Euclidean plane using the full set of pairwise geodesic distances between country centroids, and the spatial autoregressive process across residuals is modeled as varying inversely with distance from each observation up to a maximum threshold of 25,000 kilometers, thus admitting the possibility of spatial dependence at a global scale. The GMM specifications in this table correspond to the 2SLS specifications from Table I. Reassuringly, depending on the specification, the corrected standard errors of the estimated coefficient on population diversity are either similar in magnitude or noticeably smaller when compared to their heteroskedasticity robust counterparts from our baseline analysis.

Robustness to the Elimination of Regions from the Estimation Sample Following the norm in cross-country empirical studies of civil conflict, we investigate whether our main findings are driven by potentially influential world regions. The analysis in Table SA.VIII checks the qualitative robustness of the results associated with our fully specified empirical models in Columns 8 and 12 of Table I, eliminating one-at-a-time the following world regions from our global sample of countries: Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Latin America and the Caribbean (LAC). Due to the lower degrees of freedom afforded by the regression samples with eliminated regions, the analysis omits continent dummies from the empirical models in order to preserve as much of the cross-country variation in conflict frequency

as possible. The findings reassuringly reveal that the significant influence of population diversity on conflict frequency is not qualitatively sensitive to the exclusion of any one of these potentially influential world region from our full estimation sample.

TABLE SA.I: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Ecological/Climatic Covariates

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.209***	0.409***	0.306**	0.313**	0.290**	0.558**	0.636**	0.577***	0.703***
	(0.066)	(0.104)	(0.119)	(0.126)	(0.132)	(0.247)	(0.248)	(0.206)	(0.217)
Ecological fractionalization		−0.004	−0.001	−0.003	−0.003	0.001	0.003	−0.004	−0.010
		(0.016)	(0.017)	(0.017)	(0.020)	(0.021)	(0.024)	(0.016)	(0.018)
Ecological polarization		0.028	0.027	0.028	0.005	0.028	−0.002	0.030*	0.007
		(0.017)	(0.018)	(0.018)	(0.020)	(0.021)	(0.023)	(0.017)	(0.017)
Annual temperature, 1960–2016 average		0.002*	0.001	0.001	0.000	0.002	−0.001	0.002*	0.000
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Annual precipitation, 1960–2016 average		0.010	0.006	0.005	−0.001	0.018**	0.006	0.011*	0.004
		(0.006)	(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.006)	(0.006)
Volatility of annual temperature, 1960–2016		0.029	0.016	0.010	−0.003	0.007	−0.019	0.012	−0.013
		(0.024)	(0.024)	(0.022)	(0.023)	(0.029)	(0.026)	(0.023)	(0.021)
Volatility of annual precipitation, 1960–2016		−0.081*	−0.057	−0.054	−0.021	−0.143*	−0.067	−0.053	−0.011
		(0.043)	(0.042)	(0.041)	(0.046)	(0.085)	(0.089)	(0.045)	(0.052)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	150	150	150	150	147	123	121	150	147
Partial R^2 of population diversity		0.090	0.038	0.039	0.038	0.049	0.062		
Adjusted R^2	0.029	0.208	0.213	0.210	0.327	0.221	0.360		
Effect of 10th–90th %ile move in diversity	0.014***	0.027***	0.020**	0.021**	0.020**	0.027**	0.027**	0.038***	0.048***
	(0.004)	(0.007)	(0.008)	(0.008)	(0.009)	(0.012)	(0.011)	(0.014)	(0.015)
First-stage F statistic								93.172	63.364

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of (i) time-invariant fractionalization and polarization measures of the ecological diversity of land (e.g., Fenske, 2014); and (ii) the temporal mean and volatility of climatic experience (e.g., Burke et al., 2015) with respect to annual temperature and annual precipitation over the post-1960 time period. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.II: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Deep-Rooted Determinants of Economic Development

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.228***	0.378***	0.315***	0.316***	0.325**	0.547**	0.664**	0.498***	0.603***
	(0.070)	(0.103)	(0.112)	(0.116)	(0.140)	(0.266)	(0.275)	(0.192)	(0.203)
Log years since Neolithic Revolution		0.008*	0.011**	0.010*	0.008	0.004	−0.001	0.010*	0.008
		(0.004)	(0.005)	(0.005)	(0.006)	(0.010)	(0.011)	(0.005)	(0.006)
Log index of state antiquity		0.007**	0.008**	0.008**	0.004	0.008*	0.001	0.008**	0.005
		(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.006)	(0.003)	(0.005)
Log duration of human settlement		0.005**	0.001	0.001	0.003	0.003	0.009*	0.000	0.002
		(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)
Log distance from regional frontier in 1500		0.002	0.002	0.002	0.001	0.003	0.002	0.002	0.001
		(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	136	136	136	136	135	110	109	136	135
Partial R^2 of population diversity		0.085	0.046	0.044	0.054	0.044	0.077		
Adjusted R^2	0.034	0.228	0.220	0.218	0.350	0.215	0.401		
Effect of 10th–90th %ile move in diversity	0.016***	0.026***	0.022***	0.022***	0.022**	0.026**	0.033**	0.034***	0.041***
	(0.005)	(0.007)	(0.008)	(0.008)	(0.010)	(0.013)	(0.014)	(0.013)	(0.014)
First-stage F statistic								69.283	52.108

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of other deep-rooted determinants of comparative economic development, including (i) the time elapsed since the onset of the Neolithic Revolution (e.g., Ashraf and Galor, 2013a); (ii) an index of experience with institutionalized statehood since antiquity (e.g., Bockstette et al., 2002); (iii) the time elapsed since initial human settlement in prehistory (e.g., Ahlerup and Olsson, 2012); and (iv) the great-circle distance to the closest regional technological frontier in the year 1500 (e.g., Ashraf and Galor, 2013a). The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.III: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Ethnic and Spatial Inequality

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.214***	0.443***	0.338***	0.353***	0.337**	0.665***	0.760***	0.674***	0.747***
	(0.066)	(0.108)	(0.123)	(0.127)	(0.132)	(0.211)	(0.213)	(0.197)	(0.188)
Ethnic inequality in luminosity		0.021	0.020	0.018	0.013	0.023	0.022	0.024*	0.018
		(0.014)	(0.014)	(0.015)	(0.017)	(0.017)	(0.018)	(0.014)	(0.015)
Spatial inequality in luminosity		0.004	0.014	0.015	0.013	0.021	0.019	0.018	0.014
		(0.017)	(0.017)	(0.018)	(0.015)	(0.021)	(0.018)	(0.016)	(0.014)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	147	147	147	147	145	120	119	147	145
Partial R^2 of population diversity		0.132	0.054	0.056	0.062	0.094	0.139		
Adjusted R^2	0.032	0.181	0.211	0.209	0.359	0.235	0.424		
Effect of 10th–90th %ile move in diversity	0.015***	0.030***	0.023***	0.024***	0.023**	0.028***	0.033***	0.046***	0.051***
	(0.004)	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.013)	(0.013)
First-stage F statistic								133.897	80.495

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of measures of intra-country economic inequality (e.g., Alesina et al., 2016), as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across either (i) the georeferenced homelands of ethnic groups (ethnic inequality); or (ii) 2.5×2.5-degree geospatial grid cells (spatial inequality). The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.IV: Population Diversity and the Frequency of Civil Conflict Onset across Countries – The Analysis under Linguistic Fractionalization

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.218***	0.470***	0.338***	0.357***	0.332**	0.545***	0.605***	0.554***	0.603***
	(0.069)	(0.109)	(0.125)	(0.125)	(0.136)	(0.193)	(0.211)	(0.182)	(0.190)
Linguistic fractionalization				0.011	0.005		0.010		0.005
				(0.012)	(0.009)		(0.011)		(0.009)
Ethnolinguistic polarization				0.014	0.012		0.013		0.016
				(0.013)	(0.012)		(0.014)		(0.012)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	146	146	146	146	143	122	120	146	143
Partial R^2 of population diversity		0.138	0.049	0.056	0.057	0.068	0.092		
Adjusted R^2	0.031	0.196	0.217	0.227	0.372	0.226	0.407		
Effect of 10th–90th %ile move in diversity	0.014***	0.031***	0.022***	0.023***	0.022**	0.025***	0.027***	0.036***	0.039***
	(0.004)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.012)	(0.012)
First-stage F statistic								163.933	100.133

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to accounting for the potentially confounding influence of linguistic rather than ethnic fractionalization (e.g., Alesina et al., 2003), as a baseline control for subnational intergroup cultural fragmentation. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the other baseline covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.V: Population Diversity and the Frequency of Civil Conflict Onset across Countries – The Analysis under Georeferenced Linguistic Fractionalization and Polarization

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.212***	0.443***	0.315***	0.326***	0.285**	0.556***	0.578***	0.543***	0.556***
	(0.066)	(0.103)	(0.115)	(0.118)	(0.123)	(0.191)	(0.210)	(0.176)	(0.182)
Linguistic fractionalization (georeferenced)				0.002	−0.006		−0.008		−0.002
				(0.011)	(0.010)		(0.012)		(0.010)
Linguistic polarization (georeferenced)				0.006	0.008		0.010		0.009
				(0.012)	(0.010)		(0.010)		(0.009)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	151	151	151	151	148	124	122	151	148
Partial R^2 of population diversity		0.129	0.047	0.049	0.046	0.070	0.083		
Adjusted R^2	0.030	0.188	0.214	0.206	0.359	0.226	0.389		
Effect of 10th–90th %ile move in diversity	0.014***	0.029***	0.021***	0.021***	0.019**	0.027***	0.025***	0.035***	0.038***
	(0.004)	(0.007)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)	(0.011)	(0.012)
First-stage F statistic								157.089	98.473

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to accounting for the potentially confounding influence of linguistic fractionalization and polarization, constructed using georeferenced information on the spatial distribution of language homelands (from the World Language Mapping System [WLMS]) in combination with gridded population data, rather than ethnic fractionalization (e.g., Alesina et al., 2003) and ethnolinguistic polarization (e.g., Desmet et al., 2012), as baseline controls for subnational intergroup cultural fragmentation. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the other baseline covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.VI: Population Diversity and the Frequency of Civil Conflict Onset across Countries – The Analysis under Initial Values of Time-Varying Covariates

Cross-country sample:	Global					Old World		Global	
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS	(9) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.209***	0.439***	0.306***	0.318***	0.366***	0.548***	0.734***	0.537***	0.693***
	(0.066)	(0.104)	(0.115)	(0.119)	(0.136)	(0.191)	(0.215)	(0.176)	(0.192)
Executive constraints in initial year					0.004		0.003		0.005**
					(0.002)		(0.003)		(0.002)
Democracy score in initial year					−0.002		−0.002		−0.003**
					(0.002)		(0.002)		(0.002)
Autocracy score in initial year					−0.001		−0.000		−0.001
					(0.001)		(0.002)		(0.001)
Log population in initial year					0.005*		0.007**		0.004*
					(0.003)		(0.003)		(0.002)
Log GDP per capita in initial year					−0.004*		−0.004*		−0.005**
					(0.002)		(0.002)		(0.002)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for legal origin and colonial history					×		×		×
Control for oil or gas reserve discovery					×		×		×
Observations	150	150	150	150	145	123	119	150	145
Partial R^2 of population diversity		0.128	0.044	0.046	0.063	0.068	0.118		
Adjusted R^2	0.029	0.189	0.213	0.215	0.276	0.225	0.339		
Effect of 10th–90th %ile move in diversity	0.014***	0.029***	0.020***	0.021***	0.025***	0.026***	0.031***	0.036***	0.047***
	(0.004)	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.012)	(0.013)
First-stage F statistic								153.543	81.221

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness to considering the initial or year-1960 values of the time-dependent baseline controls for institutions (i.e., the degree of executive constraints and indicators for democracy and autocracy), total population, and GDP per capita, rather than their respective temporal averages over the 1960–2017 time period. The methodology exploited by the current analysis aims to reduce any ex ante bias in the baseline estimates of the influence of population diversity, arising from the fact that the temporal averages of the aforementioned time-varying controls may well vary more endogenously across countries with the contemporaneous measure of civil conflict onsets. In order to maintain a cross-country sample that as consistent as possible with the baseline analysis, observations of the time-dependent covariates from the earliest available year after 1960 are used for the subset of countries with missing 1960 data. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the other baseline covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.VII: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to Accounting for Spatial Autocorrelation in Errors

Cross-country sample:	Global					Old World		Global	
	(1) Conley OLS	(2) Conley OLS	(3) Conley OLS	(4) Conley OLS	(5) Conley OLS	(6) Conley OLS	(7) Conley OLS	(8) Conley GMM	(9) Conley GMM
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Population diversity (ancestry adjusted)	0.209***	0.439***	0.306***	0.318***	0.309***	0.548***	0.597***	0.537***	0.602***
	(0.036)	(0.068)	(0.117)	(0.110)	(0.111)	(0.076)	(0.076)	(0.084)	(0.085)
Continent dummies			×	×	×	×	×	×	×
Controls for geography		×	×	×	×	×	×	×	×
Controls for ethnic diversity				×	×		×		×
Controls for institutions					×		×		×
Controls for oil, population, and income					×		×		×
Observations	150	150	150	150	147	123	121	150	147
Adjusted R^2	0.364	0.468	0.484	0.485	0.582	0.512	0.619		

Notes: This table conducts a robustness check on the results from the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Table I. Specifically, it establishes robustness of the standard-error estimates to accounting for spatial dependence across observations, following the methodology of Conley (1999). To perform this robustness check, the spatial distribution of observations is specified on the Euclidean plane using the full set of pairwise geodesic distances between country centroids, and the spatial autoregressive process across residuals is modeled as varying inversely with distance from each observation up to a maximum threshold of 25,000 kilometers, thus admitting the possibility of spatial dependence at a global scale. The GMM specifications in this table correspond to the 2SLS specifications from Table I, exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. The specifications examined in this table are otherwise identical to corresponding ones reported in Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis. Standard errors, corrected for spatial autocorrelation, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.VIII: Population Diversity and the Frequency of Civil Conflict Onset across Countries – Robustness to the Elimination of Regions from the Global Sample

Omitted region:	None		SSA		MENA		EAP		LAC	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS	(9) OLS	(10) 2SLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017									
Population diversity (ancestry adjusted)	0.344***	0.587***	0.411***	1.243***	0.368***	0.604***	0.310**	0.561***	0.385**	0.558***
	(0.115)	(0.178)	(0.139)	(0.379)	(0.128)	(0.187)	(0.124)	(0.193)	(0.161)	(0.204)
Controls for geography	×	×	×	×	×	×	×	×	×	×
Controls for ethnic diversity	×	×	×	×	×	×	×	×	×	×
Controls for institutions	×	×	×	×	×	×	×	×	×	×
Controls for oil, population, and income	×	×	×	×	×	×	×	×	×	×
Observations	147	147	105	105	131	131	132	132	126	126
Partial R^2 of population diversity	0.051		0.058		0.039		0.011		0.087	
Adjusted R^2	0.342		0.343		0.359		0.334		0.357	
Effect of 10th–90th %ile move in diversity	0.023***	0.040***	0.026***	0.077***	0.025***	0.041***	0.018**	0.033***	0.019**	0.027***
	(0.008)	(0.012)	(0.009)	(0.024)	(0.009)	(0.013)	(0.007)	(0.011)	(0.008)	(0.010)
First-stage F statistic		59.534		17.579		57.894		50.576		73.441

Notes: This table conducts a robustness check on the results associated with the fully specified empirical models in the baseline cross-country analysis of the reduced-form impact of contemporary population diversity on the annual frequency of civil conflict onsets, as shown in Columns 8 and 12 of Table I. Specifically, it establishes robustness to the one-at-a-time elimination of world regions from the global sample, including Sub-Saharan Africa (SSA), Middle East and North Africa (MENA), East Asia and Pacific (EAP), and Latin America and the Caribbean (LAC). Due to the lower degrees of freedom afforded by the regression samples with eliminated regions, the current analysis omits continent dummies from the empirical models in order to preserve as much of the cross-country variation in conflict as possible. The regressions in Columns 1–2 should therefore be viewed as the relevant baselines for assessing the robustness results presented in the remaining columns. The set of covariates, however, is otherwise identical to those reported in Columns 8 and 12 of Table I. The reader is therefore referred to Table I and the corresponding table notes for additional details on the set of covariates considered by the current analysis as well as the identification strategy employed by the 2SLS regressions. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the number of new conflict onsets per year. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.IX: Ethnic Fractionalization, Polarization, and the Frequency of Civil Conflict Onset across Countries

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS
	Log number of new PRIO25 civil conflict onsets per year, 1960–2017								
Ethnic fractionalization	0.024***	0.021*	0.016				0.022***	0.015	0.012
	(0.007)	(0.012)	(0.012)				(0.007)	(0.012)	(0.012)
Ethnolinguistic polarization				0.014	0.019*	0.012	0.007	0.014	0.008
				(0.008)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)
Continent dummies			×			×			×
Controls for geography		×	×		×	×		×	×
Observations	154	154	154	154	154	154	154	154	154
Adjusted R^2	0.037	0.095	0.182	0.006	0.096	0.180	0.034	0.098	0.179

Notes: This table examines the sensitivity of the association between ethnic fractionalization and ethnolinguistic polarization, on the one hand, and the annual frequency of new civil conflict onsets during the 1960–2017 time period, on the other, to controls for potentially confounding geographical characteristics and continent fixed effects. The controls for geography include absolute latitude, ruggedness, distance to the nearest waterway, the mean and range of agricultural suitability, the mean and range of elevation, and an indicator for small island nations. The set of continent dummies includes five indicators for Africa, Asia, North America, South America, and Oceania. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

A.2 Robustness Checks for the Analysis of Civil Conflict in Repeated Cross-Country Data

In this appendix section, we present several robustness checks for our analysis of the influence of contemporary population diversity on the quinquennial incidence or annual onset of civil conflict in repeated cross-country data for the post-1960 time horizon.

Robustness to Accounting for Ecological/Climatic Covariates A nascent interdisciplinary literature (e.g., Burke et al., 2009; Hsiang et al., 2013; Burke et al., 2015) has emphasized the role of climatic factors, like temperature and precipitation, as important correlates of the risk of civil conflict. Further, Fenske (2014) shows that ecological diversity facilitated state centralization in pre-colonial Africa. To prevent our main specifications from becoming too unwieldy, we chose to exclude the aforementioned climatic and ecological variables from our baseline set of covariates, especially because this set already included a sizable vector of geographical factors that are known to be correlated with the former. In Table SA.X, however, we establish that population diversity remains a significant predictor of both the quinquennial incidence (Columns 1–4) and the annual onset (Columns 5–8) of civil conflict when we augment our baseline set of covariates in Table II with controls for (i) time-invariant fractionalization and polarization measures of the ecological diversity of land (e.g., Fenske, 2014); and (ii) climatic experience in the recent past (e.g., Burke et al., 2015), as captured by either (a) the temporal mean and volatility of annual temperature and annual precipitation over the previous 5-year interval for the quinquennial incidence regressions; or (b) the lagged values of annual temperature and annual precipitation as well as their temporal volatility over the previous 5 years for the annual onset regressions.

Robustness to Accounting for Deep-Rooted Determinants of Economic Development The analysis in Table SA.XI establishes the robustness of our baseline results for the quinquennial incidence and annual onset of civil conflict in repeated cross-country data to *additionally* accounting for the potentially confounding influence of other deep-rooted determinants of comparative economic development. Specifically, we augment the analysis in Table II with controls for (i) the time elapsed since the onset of the Neolithic Revolution (e.g., Ashraf and Galor, 2013a); (ii) an index of experience with institutionalized statehood since antiquity (e.g., Bockstette et al., 2002); (iii) the time elapsed since initial human settlement in prehistory (e.g., Ahlerup and Olsson, 2012); and (iv) the great-circle distance to the closest regional technological frontier in the year 1500 (e.g., Ashraf and Galor, 2013a). The results indicate that regardless of the estimation sample or the specification, contemporary population diversity remains a significant predictor of both the quinquennial likelihood of a conflict incidence (Columns 1–4) and the annual likelihood of a conflict onset (Columns 5–8).

Robustness to Accounting for Ethnic and Spatial Inequality In Table SA.XII, we check the robustness of our findings from Table II to *additionally* accounting for intra-country economic inequality (e.g., Alesina et al., 2016), as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across either (i) the georeferenced homelands of ethnic groups (ethnic inequality); or (ii) 2.5×2.5-degree geospatial grid cells (spatial inequality). The two inequality measures enter these regressions with mostly positive but invariably insignificant coefficients. Thus, unsurprisingly, the positive and significant influence of population diversity on either the quinquennial incidence or the annual onset of civil conflict remains qualitatively unaffected.

Robustness to Accounting for Alternative Correlates of Conflict Incidence The analysis in Table SA.XIII checks the robustness of our baseline results for conflict incidence to controlling for the potentially confounding influence of *alternative* distributional indices of intergroup diversity

(e.g., Fearon, 2003; Alesina et al., 2003; Esteban et al., 2012) as well as *additional* geographical correlates of conflict (e.g., Fearon and Laitin, 2003; Cervellati et al., 2017). The specifications examined by this robustness analysis are identical to the fully specified baseline models reported in Columns 2 and 4 of Table II, with the exception that in Columns 1–3 and 6–8 of the current analysis, each of the reported control variables is employed in lieu of the baseline control for ethnic fractionalization (Alesina et al., 2003), whereas in Columns 4 and 9, the set of reported control variables replaces the baseline controls for both ethnic fractionalization and ethnolinguistic polarization (Desmet et al., 2012), in the interest of mitigating multicollinearity. Further, in Columns 5 and 10, the set of reported geographical controls augment our fully specified baseline models of conflict incidence. Among the additional controls considered, ethnolinguistic polarization (Esteban et al., 2012) and the geographical variables that capture the percentage of mountainous terrain and the presence of noncontiguous territories (Fearon and Laitin, 2003) enter the IV Probit regressions in the global sample of countries with positive and significant coefficients. Nevertheless, our baseline findings regarding the significant impact of population diversity on the quinquennial incidence of civil conflict remain qualitatively unaltered across all specifications.

Robustness to Employing the Classical Logit and Rare-Events Logit Estimators The analysis in Table SA.XIV establishes the robustness of our baseline results for the quinquennial incidence and annual onset of civil conflict in repeated cross-sectional data on countries from the Old World, as shown in Columns 1–2 and 5–6 of Table II, to employing the classical logit and rare-events logit (King and Zeng, 2001) estimators, rather than the standard probit estimator. Given the absence of readily available ordinary logit and rare-events logit estimators that permit instrumentation, the current analysis is unable to implement our global-sample identification strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. As expected, the rare-events logit estimates in Table SA.XIV are somewhat smaller in absolute value than their counterparts under the classical logit estimator, due to bias arising in the latter estimates from ignoring the fact that civil conflict events (involving at least 25 battle-related deaths in a year) are generally rare occurrences in repeated cross-country data. Nonetheless, the findings attest to the robustness of the reduced-form influence of population diversity on either the quinquennial incidence or the annual onset of civil conflict under these alternative estimators.

Robustness to Accounting for Spatiotemporal Dependence using Two-Way Clustering of Standard Errors In Table SA.XV, we check the robustness of the results from our baseline probit and logit analyses of the quinquennial incidence or annual onset of civil conflict in repeated cross-sectional data on countries from the Old World, as shown in Columns 1–2 and 5–6 of Table II and in odd-numbered columns of Table SA.XIV, to accounting for spatiotemporal dependence across country-time observations. Specifically, we probe the statistical precision of our coefficient estimates by implementing multi-dimensional clustering of standard errors, following the methodology of Cameron et al. (2011). To implement this robustness check, the standard errors across country-time observations are clustered in two dimensions: (i) the country level, which allows for temporal dependence within a country over time (i.e., across either 5-year intervals or years); and (ii) the time level, which allows for spatial dependence across countries within a given time period (i.e., either a 5-year interval or a year). Given the absence of readily available probit and logit estimators that not only allow for multi-dimensional clustering of standard errors but also permit instrumentation, the current analysis is unable to implement the global-sample identification strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. Reassuringly, the bi-dimensionally clustered standard errors of our coefficient of interest are either similar to

or, in the specifications for conflict incidence, noticeably smaller in magnitude than their classically estimated counterparts in Tables II and SA.XIV that do not admit spatiotemporal dependence across country-time observations.

Robustness to Accounting for Alternative Correlates of Conflict Onset In Table SA.XVI, we check the robustness of the results from our baseline analysis of the annual onset of civil conflict in repeated cross-country data, as shown in Columns 5–8 of Table II, to accounting for the potentially confounding influence of an *additional* time-invariant distributional index of intergroup diversity, capturing the degree of "ethnic dominance" (e.g., Collier and Hoeffler, 2004), and *additional* time-varying institutional correlates of conflict onset, capturing the lagged annual values of an index of political instability and an indicator for the emergence of a newly independent state from colonial powers (e.g., Fearon and Laitin, 2003). In light of constraints imposed by the availability of data on these additional control variables, the analysis is restricted to a smaller sample of countries and to the 1960–1999 (as opposed to the 1960–2017) time period. Therefore, the specification presented in each odd-numbered column of the table is intended to provide a relevant baseline for the robustness check in the subsequent even-numbered column (i.e., by holding fixed the regression sample). Turning to the results in Table SA.XVI, the lagged index of political instability does appear to enter some of our specifications with a positive and statistically significant coefficient, although the other additional controls considered by the analysis do not seem to be significantly correlated with conflict onset. However, despite the substantial reduction in both the sample time-frame and the number of countries in the cross-section, our coefficient of interest reassuringly remains positive and precisely estimated, regardless of the inclusion of these additional controls to the specifications.

Robustness to Accounting for Commodity Export Price Shocks The analysis in Table SA.XVII checks the robustness of our baseline results for the annual onset of civil conflict in repeated cross-country data, as shown in Columns 5–8 of Table II, to *additionally* accounting for the potentially confounding "income effect" of commodity export price shocks (e.g., Bazzi and Blattman, 2014), as captured by the contemporaneous, lagged, and twice lagged values of either an annual price shock that has been aggregated across commodity export types (Columns 1–2 and 5–6) or annual price shocks disaggregated by type of commodity export, including export price shocks associated with annual crops, perennial crops, and extractive crops (Columns 3–4 and 7–8). These export price shock variables are all obtained from the data set of Bazzi and Blattman (2014), so the reader is referred to that work for additional details on these variables. In light of constraints imposed by the availability of data on these additional covariates, the analysis is restricted to a smaller sample of countries and to the 1960–2007 (as opposed to the 1960–2017) time period. As is evident from the results in Table SA.XVII, there is indeed a significant mitigating "income effect" on the annual likelihood of a conflict onset associated with the contemporaneous and twice lagged values of commodity export price shocks (for both aggregated and disaggregated variants of these shocks). Nonetheless, despite the reduction in both the number of countries in the cross-section and the sample time-frame, our coefficient of interest reassuringly remains positive and statistically significant when subjected to these additional covariates in the specifications.

TABLE SA.X: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Ecological/Climatic Covariates

Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	14.367***	10.178**	17.325***	15.651***	6.172*	6.001*	7.063**	9.482**
	(4.264)	(4.488)	(4.387)	(5.167)	(3.306)	(3.538)	(3.425)	(4.282)
Ecological fractionalization	−0.368	−0.080	−0.503	−0.394	0.018	−0.401	−0.027	−0.432
	(0.456)	(0.524)	(0.432)	(0.494)	(0.274)	(0.371)	(0.275)	(0.376)
Ecological polarization	0.865**	0.327	1.086***	0.927**	0.238	0.330	0.406	0.529
	(0.417)	(0.504)	(0.398)	(0.471)	(0.301)	(0.419)	(0.303)	(0.420)
Lagged temperature	0.078***	0.002	0.067***	0.023	0.033*	−0.004	0.032*	0.009
	(0.027)	(0.034)	(0.021)	(0.025)	(0.019)	(0.024)	(0.016)	(0.020)
Lagged precipitation	0.177	−0.042	0.248	0.148	0.096	−0.002	0.110	0.086
	(0.178)	(0.166)	(0.167)	(0.176)	(0.124)	(0.138)	(0.122)	(0.140)
Lagged temperature volatility	−0.576*	−0.416	−0.356	−0.274	0.307	0.249	0.218	0.239
	(0.342)	(0.382)	(0.307)	(0.332)	(0.287)	(0.281)	(0.272)	(0.263)
Lagged precipitation volatility	−1.326	−1.363	−0.504	−0.439	−0.282	−0.152	−0.566	−0.221
	(0.814)	(1.096)	(0.603)	(0.742)	(0.592)	(0.708)	(0.595)	(0.647)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	1,270	1,045	1,583	1,311	5,452	4,377	6,996	5,757
Countries	123	121	150	147	123	121	150	147
Pseudo R^2	0.431	0.443			0.135	0.163		
Marginal effect of diversity	2.675***	1.873**	3.364***	2.981***	0.322*	0.312*	0.333*	0.454*
	(0.796)	(0.833)	(0.908)	(1.046)	(0.177)	(0.186)	(0.173)	(0.233)
First-stage F statistic			83.318	70.585			94.679	77.102

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-country data, as shown in Table II. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of (i) time-invariant fractionalization and polarization measures of the ecological diversity of land (e.g., Fenske, 2014); and (ii) climatic experience in the recent past (e.g., Burke et al., 2015), as captured by either (a) the temporal mean and volatility of annual temperature and annual precipitation over the previous 5-year interval for the quinquennial incidence regressions; or (b) the lagged values of annual temperature and annual precipitation as well as their temporal volatility over the previous 5 years for the annual onset regressions. The specifications examined in this table are otherwise identical to corresponding ones reported in Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the incidence or onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XI: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Deep-Rooted Determinants of Economic Development

Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	15.404***	9.821**	19.297***	15.653**	5.222*	4.777*	8.565**	11.664***
	(4.670)	(4.781)	(5.404)	(6.386)	(2.939)	(2.784)	(3.657)	(4.255)
Log years since Neolithic Revolution	0.085	0.187	−0.290	−0.243	0.333**	0.324*	0.029	−0.160
	(0.270)	(0.296)	(0.285)	(0.334)	(0.147)	(0.174)	(0.194)	(0.232)
Log index of state antiquity	0.244***	0.076	0.286***	0.143	0.093**	0.035	0.125**	0.096
	(0.088)	(0.103)	(0.101)	(0.116)	(0.041)	(0.057)	(0.051)	(0.070)
Log duration of human settlement	0.000	0.070	−0.024	−0.009	0.039	0.044	0.004	0.019
	(0.131)	(0.131)	(0.097)	(0.118)	(0.066)	(0.071)	(0.059)	(0.069)
Log distance from regional frontier in 1500	−0.031	0.001	−0.057	−0.025	0.049	0.050	−0.004	−0.018
	(0.052)	(0.051)	(0.040)	(0.047)	(0.032)	(0.038)	(0.026)	(0.031)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	1,141	953	1,447	1,219	4,810	4,481	6,280	5,886
Countries	110	109	136	135	110	109	136	135
Pseudo R^2	0.425	0.432			0.143	0.151		
Marginal effect of diversity	2.992***	1.901**	3.885***	3.105**	0.293*	0.263*	0.437**	0.604**
	(0.896)	(0.936)	(1.140)	(1.333)	(0.165)	(0.154)	(0.203)	(0.257)
First-stage F statistic			41.126	39.893			48.227	44.985

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-country data, as shown in Table II. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of other deep-rooted determinants of comparative economic development, including (i) the time elapsed since the onset of the Neolithic Revolution (e.g., Ashraf and Galor, 2013a); (ii) an index of experience with institutionalized statehood since antiquity (e.g., Bockstette et al., 2002); (iii) the time elapsed since initial human settlement in prehistory (e.g., Ahlerup and Olsson, 2012); and (iv) the great-circle distance to the closest regional technological frontier in the year 1500 (e.g., Ashraf and Galor, 2013a). The specifications examined in this table are otherwise identical to corresponding ones reported in Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the incidence or onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XII: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Ethnic and Spatial Inequality

Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) Probit	(6) Probit	(7) IV Probit	(8) IV Probit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	14.732***	14.259***	16.367***	16.080***	6.687**	6.812**	7.892***	9.098***
	(3.867)	(3.801)	(3.782)	(4.046)	(2.862)	(2.952)	(2.971)	(3.367)
Ethnic inequality in luminosity	0.593	0.675	0.331	0.277	0.330	0.330	0.263	0.142
	(0.372)	(0.451)	(0.376)	(0.445)	(0.261)	(0.262)	(0.257)	(0.255)
Spatial inequality in luminosity	−0.035	0.150	0.294	0.519	−0.053	−0.017	0.070	0.086
	(0.409)	(0.425)	(0.392)	(0.410)	(0.256)	(0.259)	(0.247)	(0.279)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Controls for oil, population, and income		×		×		×		×
Observations	1,234	1,038	1,547	1,304	5,206	4,342	6,840	5,722
Countries	120	119	147	145	120	119	147	145
Pseudo R^2	0.408	0.442			0.133	0.172		
Marginal effect of diversity	2.838***	2.626***	3.272***	3.094***	0.348**	0.347**	0.370**	0.431**
	(0.717)	(0.702)	(0.787)	(0.843)	(0.154)	(0.153)	(0.153)	(0.182)
First-stage F statistic			125.548	93.701			133.266	99.940

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-country data, as shown in Table II. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding influence of measures of intrastate economic inequality (e.g., Alesina et al., 2016), as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across either (i) the georeferenced homelands of ethnic groups (ethnic inequality); or (ii) 2.5×2.5-degree geospatial grid cells (spatial inequality). The specifications examined in this table are otherwise identical to corresponding ones reported in Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the incidence or onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XIII: Population Diversity and the Incidence of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Alternative Correlates of Conflict Incidence

Cross-country sample:	Old World					Global				
	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) Probit	(6) IV Probit	(7) IV Probit	(8) IV Probit	(9) IV Probit	(10) IV Probit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017									
Population diversity (ancestry adjusted)	12.439***	12.412***	13.672***	9.587**	13.200***	13.115***	13.929***	14.428***	10.985**	14.758***
	(3.718)	(3.745)	(4.027)	(4.202)	(4.052)	(4.107)	(4.149)	(4.427)	(4.442)	(4.774)
Ethnic fractionalization (Fearon, 2003)	−0.266					−0.147				
	(0.332)					(0.329)				
Linguistic fractionalization (Alesina et al., 2003)		0.348					0.276			
		(0.354)					(0.317)			
Religious fractionalization (Alesina et al., 2003)			−0.463*					−0.705**		
			(0.280)					(0.276)		
Ethnolinguistic fractionalization (Esteban et al., 2012)				0.106					0.179	
				(0.365)					(0.346)	
Ethnolinguistic polarization (Esteban et al., 2012)				0.717					3.225**	
				(1.488)					(1.374)	
Gini index of ethnolinguistic diversity (Esteban et al., 2012)				−0.519					−1.358	
				(0.716)					(1.053)	
Log percentage mountainous terrain					0.099					0.112*
					(0.063)					(0.062)
Noncontiguous state dummy					0.371*					0.560***
					(0.214)					(0.182)
Disease richness					0.000					−0.007
					(0.010)					(0.010)
Controls for all baseline covariates	×	×	×	×	×	×	×	×	×	×
Observations	1,020	1,035	1,046	950	1,015	1,286	1,278	1,312	1,177	1,281
Countries	119	120	121	106	118	145	143	147	128	144
Pseudo R^2	0.429	0.436	0.438	0.451	0.436					
Marginal effect of diversity	2.387***	2.309***	2.547***	1.779**	2.499***	2.577***	2.664***	2.759***	2.124**	2.853***
	(0.722)	(0.700)	(0.762)	(0.789)	(0.784)	(0.852)	(0.833)	(0.894)	(0.891)	(0.978)
First-stage F statistic						100.578	104.976	98.705	68.499	70.482

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on the quinquennial incidence of civil conflict in repeated cross-country data, as shown in Columns 2 and 4 of Table II. Specifically, it establishes robustness to accounting for the potentially confounding influence of *alternative* distributional indices of intergroup diversity (e.g., Fearon, 2003; Alesina et al., 2003; Esteban et al., 2012) and *additional* geographical correlates of conflict (e.g., Fearon and Laitin, 2003; Cervellati et al., 2017). The specifications examined in this table are identical to the fully specified baseline models of conflict incidence, as reported in Columns 2 and 4 of Table II, with the exception that in Columns 1–3 and 6–8 of the current analysis, each of the reported control variables is employed in lieu of the baseline control for ethnic fractionalization (Alesina et al., 2003), whereas in Columns 4 and 9, the set of reported control variables replaces the baseline controls for both ethnic fractionalization and ethnolinguistic polarization (Desmet et al., 2012), in the interest of mitigating multicollinearity. Further, in Columns 5 and 10 of the current analysis, the set of reported geographical controls augment the fully specified baseline models from Columns 2 and 4 of Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the incidence of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XIV: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Employing the Classical Logit and Rare-Events Logit Estimators

	(1) Classical Logit	(2) Rare-Events Logit	(3) Classical Logit	(4) Rare-Events Logit	(5) Classical Logit	(6) Rare-Events Logit	(7) Classical Logit	(8) Rare-Events Logit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	24.420***	23.755***	22.262***	20.941***	13.857**	13.409**	13.175**	12.442*
	(6.653)	(6.529)	(6.703)	(6.479)	(6.266)	(6.177)	(6.584)	(6.517)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity			×	×			×	×
Controls for institutions			×	×			×	×
Controls for oil, population, and income			×	×			×	×
Observations	1,270	1,270	1,045	1,045	5,452	6,280	4,377	5,221
Countries	123	123	121	121	123	123	121	121
Pseudo R^2	0.414		0.441		0.133		0.164	
Marginal effect of diversity	3.733***	3.964***	2.992***	3.230***	0.191**	0.194**	0.156*	0.171*
	(1.009)	(1.128)	(0.937)	(1.088)	(0.086)	(0.097)	(0.081)	(0.095)

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-sectional data for the Old World sample of countries, as shown in Columns 1–2 and 5–6 of Table II. Specifically, it establishes robustness to employing the ordinary logit and rare-events logit (King and Zeng, 2001) estimators, rather than the probit estimator, for estimating the relevant empirical models of conflict incidence and onset. The specifications examined in this table are otherwise identical to corresponding ones reported in Columns 1–2 and 5–6 of Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis. Given the absence of readily available ordinary logit and rare-events logit estimators that permit instrumentation, the current analysis is unable to implement the global-sample identification strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. The estimated marginal effect of a 1 percentage point increase in population diversity is the marginal effect at the mean value of diversity in the cross-section, and it reflects the increase in either the quinquennial likelihood of a conflict incidence (Columns 1–4) or the annual likelihood of a conflict onset (Columns 5–8), both expressed in percentage points. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XV: Population Diversity and the Incidence or Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Spatiotemporal Dependence using Two-Way Clustering of Standard Errors

	(1) Probit	(2) Logit	(3) Probit	(4) Logit	(5) Probit	(6) Logit	(7) Probit	(8) Logit
	Quinquennial PRIO25 civil conflict incidence, 1960–2017				Annual PRIO25 civil conflict onset, 1960–2017			
Population diversity (ancestry adjusted)	13.366***	24.420***	12.203***	22.262***	6.172**	13.857**	6.356*	13.175*
	(2.616)	(4.261)	(3.381)	(6.025)	(2.906)	(6.528)	(3.478)	(7.368)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity			×	×			×	×
Controls for institutions			×	×			×	×
Controls for oil, population, and income			×	×			×	×
Observations	1,270	1,270	1,045	1,045	5,452	5,452	4,377	4,377
Countries	123	123	121	121	123	123	121	121
Pseudo R^2	0.416	0.414	0.440	0.441	0.131	0.133	0.161	0.164

Notes: This table conducts a robustness check on the results from the baseline probit and logit analyses of the reduced-form impact of contemporary population diversity on either the quinquennial incidence or the annual onset of civil conflict in repeated cross-sectional data for the Old World sample of countries, as shown in Columns 1–2 and 5–6 of Table II and in odd-numbered columns of Table SA.XIV. Specifically, it establishes robustness of the standard-error estimates to accounting for spatiotemporal dependence across country-time observations by implementing multi-dimensional clustering of standard errors, following the methodology of Cameron et al. (2011). To implement this robustness check, the standard errors across country-time observations are clustered in two dimensions: (i) the country level, which allows for temporal dependence within a country over time (i.e., across either 5-year intervals or years); and (ii) the time level, which allows for spatial dependence across countries within a given time period (i.e., either a 5-year interval or a year). The specifications examined in this table are otherwise identical to corresponding ones reported in Columns 1–2 and 5–6 of Table II and in odd-numbered columns of Table SA.XIV. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis. Given the absence of readily available probit and logit estimators that not only allow for multi-dimensional clustering of standard errors but also permit instrumentation, the current analysis is unable to implement the global-sample identification strategy of exploiting prehistoric migratory distance from East Africa to the indigenous (precolonial) population of a country as an excluded instrument for the country's contemporary population diversity. Heteroskedasticity-robust standard errors, clustered multi-dimensionally at both the country and time levels, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XVI: Population Diversity and the Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Alternative Correlates of Conflict Onset

Cross-country sample:	Old World				Global			
	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) IV Probit	(6) IV Probit	(7) IV Probit	(8) IV Probit
	Annual PRIO25 civil conflict onset, 1960–1999							
Population diversity (ancestry adjusted)	7.791**	6.872**	8.267**	8.330*	8.808**	8.111**	11.955**	11.507**
	(3.657)	(3.469)	(4.181)	(4.342)	(3.516)	(3.417)	(4.838)	(4.975)
Ethnic dominance		0.147		−0.002		0.147		0.040
		(0.115)		(0.135)		(0.103)		(0.129)
Political instability, lagged		0.264**		0.165		0.245**		0.056
		(0.106)		(0.136)		(0.098)		(0.128)
New state dummy, lagged		0.125				−0.149		
		(0.527)				(0.494)		
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity			×	×			×	×
Controls for institutions			×	×			×	×
Controls for oil, population, and income			×	×			×	×
Observations	2,761	2,761	2,139	2,139	3,728	3,728	3,031	3,031
Countries	96	96	94	94	121	121	119	119
Pseudo R^2	0.137	0.145	0.155	0.157				
Marginal effect of diversity	0.472**	0.413*	0.516*	0.519*	0.495**	0.448**	0.706**	0.672*
	(0.231)	(0.216)	(0.267)	(0.277)	(0.224)	(0.210)	(0.349)	(0.350)
First-stage F statistic					132.831	132.602	78.279	73.849

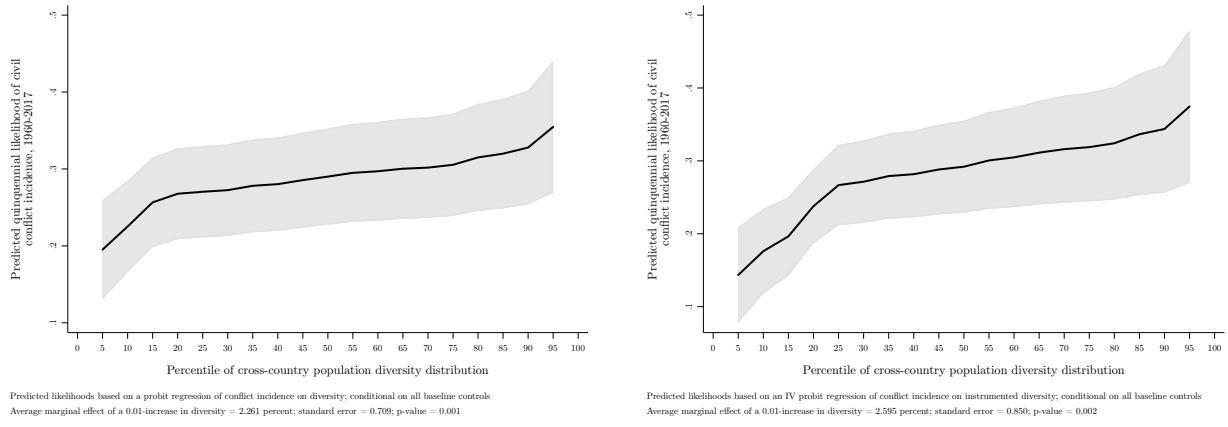
Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on the annual onset of civil conflict in repeated cross-country data, as shown in Columns 5–8 of Table II. Specifically, it establishes robustness to accounting for the potentially confounding influence of an *additional* distributional index of intergroup diversity (e.g., Collier and Hoeffler, 2004) and *additional* time-varying institutional correlates of conflict (e.g., Fearon and Laitin, 2003). The lagged indicator for the emergence of a newly independent state from colonial powers is dropped from the specifications in Columns 4 and 8 due to multicollinearity. In light of constraints imposed by the availability of data on the additional control variables in this table, the analysis is restricted to the 1960–1999 as opposed to the 1960–2017 time period. Therefore, the specification presented in each odd-numbered column of the table is intended to provide a relevant baseline for the robustness check in the subsequent even-numbered column (i.e., by holding fixed the regression sample). The specifications examined in this table are otherwise identical to the baseline models of conflict onset, as reported in Columns 5–8 of Table II. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SA.XVII: Population Diversity and the Onset of Civil Conflict in Repeated Cross-Country Data – Robustness to Accounting for Commodity Export Price Shocks

Cross-country sample:	Old World				Global			
	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) IV Probit	(6) IV Probit	(7) IV Probit	(8) IV Probit
	Annual PRIO25 civil conflict onset, 1960–2007							
Population diversity (ancestry adjusted)	8.596**	8.946**	8.632**	8.734**	9.007***	10.656**	9.086***	10.592**
	(3.665)	(3.894)	(3.622)	(3.899)	(3.401)	(4.537)	(3.388)	(4.570)
Aggregate price shock	−0.128**	−0.159***			−0.137***	−0.190***		
	(0.052)	(0.059)			(0.053)	(0.056)		
Aggregate price shock, lagged	0.026	0.021			0.014	0.017		
	(0.060)	(0.069)			(0.058)	(0.062)		
Aggregate price shock, twice lagged	−0.172***	−0.179***			−0.113*	−0.121*		
	(0.060)	(0.066)			(0.058)	(0.064)		
Annual crop price shock			−0.161**	−0.191**			−0.156**	−0.223***
			(0.071)	(0.083)			(0.071)	(0.075)
Annual crop price shock, lagged			−0.039	−0.048			−0.049	−0.045
			(0.083)	(0.093)			(0.082)	(0.088)
Annual crop price shock, twice lagged			−0.176**	−0.178*			−0.101	−0.112
			(0.084)	(0.094)			(0.084)	(0.095)
Perennial crop price shock			−0.127*	−0.144**			−0.127**	−0.154***
			(0.066)	(0.070)			(0.058)	(0.059)
Perennial crop price shock, lagged			0.116***	0.120**			0.094**	0.089*
			(0.045)	(0.054)			(0.046)	(0.051)
Perennial crop price shock, twice lagged			−0.130***	−0.145***			−0.076	−0.083*
			(0.050)	(0.053)			(0.046)	(0.049)
Extractive crop price shock			−0.187**	−0.247***			−0.185**	−0.275***
			(0.081)	(0.092)			(0.081)	(0.086)
Extractive crop price shock, lagged			0.051	0.055			0.031	0.041
			(0.088)	(0.098)			(0.088)	(0.094)
Extractive crop price shock, twice lagged			−0.330***	−0.332***			−0.256***	−0.264**
			(0.103)	(0.111)			(0.096)	(0.104)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Observations	2,876	2,626	2,876	2,626	3,906	3,599	3,906	3,599
Countries	82	81	82	81	105	103	105	103
Pseudo R^2	0.122	0.150	0.133	0.162				
Marginal effect of diversity	0.531**	0.535**	0.528**	0.516**	0.501**	0.577**	0.500**	0.568**
	(0.237)	(0.242)	(0.232)	(0.240)	(0.213)	(0.281)	(0.211)	(0.280)
First-stage F statistic					102.975	51.265	102.702	51.169

Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on the annual onset of civil conflict in repeated cross-country data, as shown in Columns 5–8 of Table II. Specifically, it establishes robustness to *additionally* accounting for the potentially confounding "income effect" of commodity export price shocks (e.g., Bazzi and Blattman, 2014), as captured by the contemporaneous, lagged, and twice lagged values of either an annual price shock that has been aggregated across commodity export types (Columns 1–2 and 5–6) or annual price shocks disaggregated by type of commodity export, including export price shocks associated with annual crops, perennial crops, and extractive crops (Columns 3–4 and 7–8). These export price shock variables are all obtained from the data set of Bazzi and Blattman (2014), so the reader is referred to that work for additional details on these variables. In light of constraints imposed by the availability of data on these export price shock variables, the analysis is restricted to the 1960–2007 as opposed to the 1960–2017 time period. The specifications examined in this table are otherwise identical to those reported in Columns 5–8 of Table II, with the exception that the fully specified models in the current analysis omit the controls for oil presence, total population, and GDP per capita, in the interest of minimizing endogeneity with the export price shock variables and maximizing degrees of freedom. The reader is therefore referred to Table II and the corresponding table notes for additional details on the baseline set of covariates considered by the current analysis, the identification strategy employed by the IV probit regressions, and the estimation and interpretation of the marginal effect of population diversity on the onset of conflict. Heteroskedasticity-robust standard errors, clustered at the country level, are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

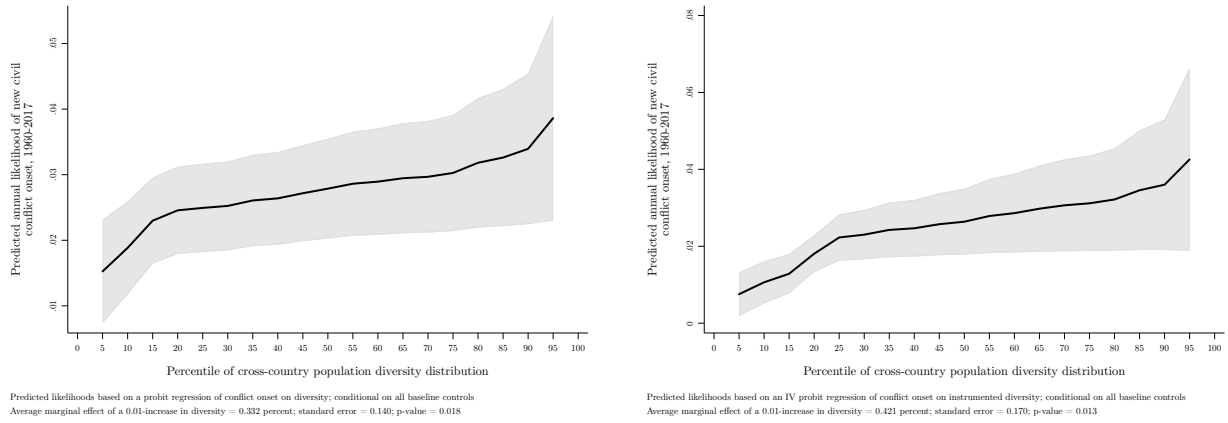
A.3 Supplementary Figures



(a) Old-World sample (b) Global sample

FIGURE SA.1: Population Diversity and the Incidence of Civil Conflict

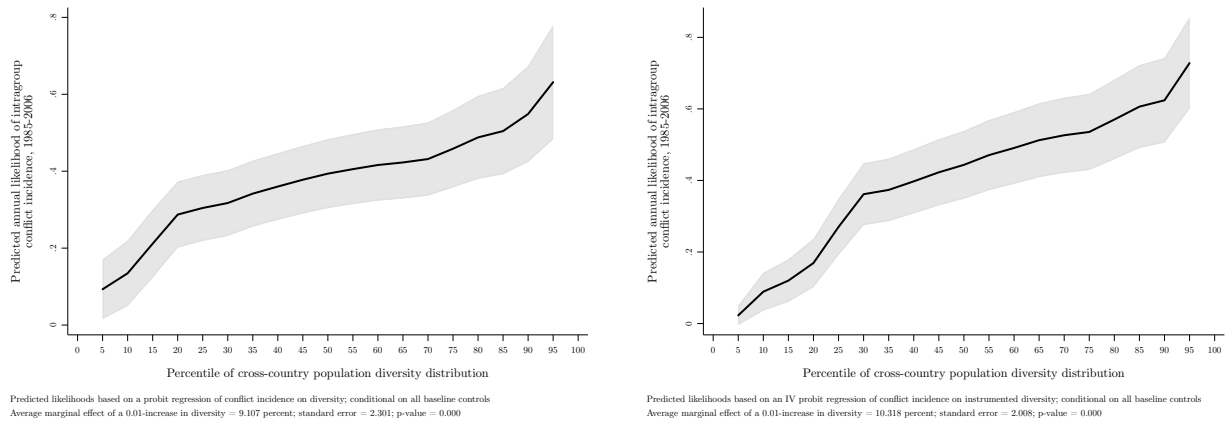
Notes: This figure depicts the influence of contemporary population diversity on the *predicted* likelihood of observing the incidence of a PRIO25 civil conflict in any given 5-year interval during the 1960–2017 time period, conditional on the full set of control variables, as considered by the specifications in Columns 2 and 4 of Table II. In each panel, the predicted likelihood of civil conflict incidence is illustrated as a function of the percentile of the cross-country diversity distribution in the relevant estimation sample, and the shaded area reflects the 95-percent confidence-interval region of the depicted relationship.



(a) Old-World sample (b) Global sample

FIGURE SA.2: Population Diversity and the Onset of Civil Conflict

Notes: This figure depicts the influence of contemporary population diversity on the *predicted* likelihood of observing the onset of a new PRIO25 civil conflict in any given year during the 1960–2017 time period, conditional on the full set of control variables, as considered by the specifications in Columns 6 and 8 of Table II. In each panel, the predicted likelihood of civil conflict onset is illustrated as a function of the percentile of the cross-country diversity distribution in the relevant estimation sample, and the shaded area reflects the 95-percent confidence-interval region of the depicted relationship.



(a) Old-World sample

(b) Global sample

FIGURE SA.3: Population Diversity and the Incidence of Intragroup Conflict

Notes: This figure depicts the influence of contemporary population diversity on the *predicted* likelihood of observing the incidence of one or more intragroup conflicts in any given year during the 1985–2006 time period, conditional on the full set of control variables, as considered by the specifications in Columns 2 and 5 in Panel B of Table III. In each panel, the predicted likelihood of intragroup conflict incidence is illustrated as a function of the percentile of the cross-country diversity distribution in the estimation relevant sample, and the shaded area reflects the 95-percent confidence-interval region of the depicted relationship.

A.4 Variable Definitions for the Country-level Analyses

Migratory Distance and Population Diversity

1. **Migratory distance from East Africa:** The great circle distance from Addis Ababa, Ethiopia to a country's capital city along a land-restricted path forced through one or more of five intercontinental waypoints, including Cairo, Egypt; Istanbul, Turkey; Phnom Penh, Cambodia; Anadyr, Russia; and Prince Rupert, Canada. Distances are calculated using the Haversine formula and are measured in units of ten thousand kilometers. The methodology underlying the construction of this measure is adopted from Ramachandran et al. (2005). The geographical coordinates of the waypoints are obtained from Ramachandran et al. (2005) and those of the capital cities are obtained from the Central Intelligence Agency's (CIA) World Factbook. See Ashraf and Galor (2013a) for additional details.

2. **Population diversity (precolonial):** The expected heterozygosity (neutral genetic diversity) of a country's precolonial population as predicted by migratory distance from East Africa (i.e., Addis Ababa, Ethiopia) to the country's capital city. This measure is calculated by applying the regression coefficients obtained from regressing expected heterozygosity on migratory distance at the ethnic group level, using a worldwide sample of 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel. The expected heterozygosities and geographical coordinates of the ethnic groups are from Ramachandran et al. (2005). See Ashraf and Galor (2013a) for additional details.

3. **Population diversity (ancestry adjusted):** The expected heterozygosity (neutral genetic diversity) of a country's contemporary national population, as developed by Ashraf and Galor (2013a). This measure is based on migratory distances from East Africa to the year 1500 locations of the ancestral populations of the country's component ethnic groups in 2000 and on the pairwise migratory distances among these ancestral populations. The source countries of the ancestral populations are identified from the World Migration Matrix, 1500–2000 (Putterman and Weil, 2010), and the capital cities of these countries are used to compute the aforementioned migratory distances. The measure of population diversity is then computed by applying (i) the coefficients obtained from regressing expected heterozygosity on migratory distance from East Africa at the ethnic group level, using a worldwide sample of 53 ethnic groups from the HGDP-CEPH Human Genome Diversity Cell Line Panel; (ii) the coefficients obtained from regressing pairwise genetic distance on pairwise migratory distance in a sample of 1,378 HGDP-CEPH ethnic group pairs, and (iii) the ancestry weights representing the fractions of the year 2000 national population (i.e., of the country for which the measure is being computed) that can trace their ancestral origins to different source countries in the year 1500. The data at the ethnic-group (or group-pair) level on expected heterozygosities, geographical coordinates, and pairwise genetic distances are obtained from Ramachandran et al. (2005), and the country-level data on ancestry weights are obtained from the World Migration Matrix, 1500–2000. See Ashraf and Galor (2013a) for a detailed discussion of the methodology underlying the construction of this measure.

Conflict outcomes

1. **PRIO civil conflict and civil war outcomes:** Our primary measures of civil conflict are based on Version 18.1 of the UCDP/PRIO Armed Conflict Dataset (ACD), covering the 1946–2017 time period (Gleditsch et al., 2002; Pettersson and Eck, 2018). In this dataset, an armed

conflict is defined as "a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in a calendar year." In our study, the term *PRIO25 civil conflict* indicates an internal armed conflict between the government of a state and one or more internal opposition group(s), without any intervention from other states as independent actors or intervention from other states to support either side of the conflict. Thus, the measures of civil conflict in our study exclude internationalized internal armed conflicts. In addition, extrasytemic and interstate conflicts are also excluded from the analysis, following the standard definition of civil conflict. For further information on the data underlying our various civil conflict measures (discussed below), the interested reader is referred to the codebook for Version 18.1 of the UCDP/PRIO ACD.

The main conflict variable examined in our cross-sectional analyses of civil conflict is the ***log number of new PRIO25 civil conflict onsets per year*** during the 1960–2017 time period. This measure is obtained by first computing the total count of *new* civil conflicts that took place on the territory of a country in our sample during this period. Then, this count is divided by the number of years over the same time period in which the territory was home to one or more entities included in the Gleditsch and Ward list of independent states, as employed by the UCDP/PRIO ACD. Finally, the resulting average annual conflict frequency is scaled up by 1 and log-transformed. Each *new* conflict is identified by a unique conflict identifier provided by the UCDP/PRIO ACD. In this definition, two or more conflict episodes involving the same actors fighting over the same incompatibility are not treated as separate (new) conflicts. Instead, they are assigned the same conflict identifier.

The main outcome examined by our regressions using annually repeated cross-country data is ***annual PRIO25 civil conflict onset***. It is equal to 1 for each year when at least one new PRIO25 conflict broke out and zero otherwise. The date of a *new* conflict outbreak (or onset) is the starting year of the first conflict episode for a given conflict, and it reflects the first year in which the conflict reached or surpassed the annual fatality threshold of 25 battle-related deaths. Subsequent years of a given conflict episode or outbreaks of subsequent conflict episodes of the same conflict are not considered *new* conflict onsets.

Quinquennial PRIO25 civil conflict incidence is the main outcome examined by our regressions using quinquennially repeated cross-country data over the 1960–2017 time period. It is equal to 1 for a given 5-year interval for a country if there was an active (ongoing) PRIO25 civil conflict in at least one year during that time interval and zero otherwise. A conflict is deemed active in a given calendar year if it resulted in at least 25 battle-related deaths during that year. ***Annual PRIO25 civil conflict incidence*** is defined in a similar manner except that the incidence is coded for each country-year observation instead of a 5-year time interval for a country.

Quinquennial PRIO1000 civil war incidence is an alternative outcome examined by our robustness checks in regressions using quinquennially repeated cross-country data. This variable is constructed in a manner similar to ***quinquennial PRIO25 civil conflict incidence***. The only difference is that for civil wars, a conflict is deemed as active (ongoing) in a given year only if a much higher fatality threshold of 1,000 (instead of 25) battle-related deaths is exceeded in that year.

2. **Intragroup (intracommunal) factional conflict:** The outcome variables employed by the analysis of intragroup conflict are based on the All Minorities At Risk (AMAR) Sample

Data of the AMAR Phase I Project (Birnir et al., 2018). The AMAR sample contains longitudinal data on 365 AMAR ethnic groups. Of these groups, 291 were included in the original Minorities At Risk (MAR) Project (Phases I–V), and the remaining 74 were selected randomly from the sample frame of socially relevant groups outlined by Birnir et al. (2015), according to the new AMAR criteria summarized in the AMAR codebook.

The measures of intragroup factional conflict we employ are constructed using the *INTRACON* variable in the AMAR Sample Data. This is a dummy variable, coded for each group in the AMAR sample, indicating the presence of an intracommunal conflict within that group in a given year. Specifically, the variable is coded for each year during the 1980–2006 time period. However, since the coverage of AMAR groups for the 1980–1984 time period is rather limited, our measures of intragroup conflict are based on information for the 1985–2006 time frame. Thus, the outcome variable in our cross-country analysis of intragroup conflict is the ***share of AMAR group-years with at least one intracommunal conflict*** within a country during this time period. Further, the outcome variable in our analysis of intragroup conflict using annually repeated cross-country data is ***annual intracommunal conflict incidence***, coded 1 for each country-year in which there was at least one AMAR sample group with an active intracommunal conflict and zero otherwise. For further information on the data underlying our measures of intragroup conflict, the reader is referred to Version 1 of the codebook for the AMAR Phase I Project.

3. **Historical conflict outcomes:** To construct historical conflict outcomes between the 15th and 19th centuries, we make use of information on the locations of violent conflicts during the 1400–1799 time period, as compiled by Brecke (1999) and georeferenced by Dincecco et al. (2015). The georeferenced conflict locations are used to map historical conflicts to territories, as defined by their contemporary national borders. It may be noted that in the catalog of conflicts from Dincecco et al. (2015), there were a small number of instances where the country assignment did not match the country implied by the georeferenced location of the conflict in ArcGIS. In such cases, supplementary information from the catalog (e.g., the actors in the conflict or the place where the conflict occurred) was consulted to first determine if the mismatch was due to an error in the original country assignment or an error in the supplied coordinates. Then, either the country assignment or the coordinates were altered to match our understanding of the true location of the conflict. In addition, for naval conflicts or for conflicts between actors that took place on lands to which neither actor was native, these specific conflicts were assigned to either one of the actors' countries (rather than the country implied by the location of the conflict) but only if the actors possessed comparable levels of diversity (e.g., if the actors were both European colonial powers engaged in a conflict on a colonized territory).

   As for the underlying conflict data, the definition of a violent conflict in Brecke's dataset is based on Cioffi-Revilla (1996): "An occurrence of purposive and lethal violence among 2+ social groups pursuing conflicting political goals that results in fatalities, with at least one belligerent group organized under the command of authoritative leadership. The state does not have to be an actor. Data can include massacres of unarmed civilians or territorial conflicts between warlords." The list is comprised of conflicts that resulted in at least 32 fatalities. This fatality level corresponds to a magnitude of 1.5 or higher on Richardson's (1960) base-10 log conflict scale. Although the dataset does not systematically distinguish between intrastate and interstate conflicts, the latter appear to form the basis of the recorded conflicts, and while the recorded conflicts do not necessarily represent the whole universe

of conflict events during the sample period, the list contains almost all major conflicts that have been documented by historians. The conflict catalog is also considered to be fairly comprehensive in terms of its broad regional coverage, including five regions of the world: Western Europe, Eastern Europe, North Africa, West & Central Africa, East & Southern Africa, as well as Central Asia & Siberia.

Based on these conflict data, our study employs two distinct categories of country-level outcome measures: (1) the number of distinct conflicts, occurring in each century of the 1400–1799 time period or across this entire time frame; and (2) the likelihood of observing one or more conflicts, either during the entire 1400–1799 time period or in each century therein.

4. **MEPV civil conflict severity:** This variable is constructed using information provided by the Major Episodes of Political Violence (MEPV) War List (1946–2017), maintained by the Center for Systemic Peace. This list is a regularly updated version of Appendix C from Marshall (1999) and further detailed in Marshall (2002).

   A major episode of political violence is defined as the systematic and sustained use of lethal violence by one or more organized groups, resulting in at least 500 directly-related deaths over the course of the episode. Episodes are coded for both time span and a general magnitude of societal-systemic impact (an eleven-point scale, 0-10). These magnitude scores are considered to be consistent and comparable across categories and cases. Further, each episode is assigned to one of seven categories of armed conflict: international violence (IV), international war (IW), international independence war (IN), civil violence (CV), civil war (CW), ethnic violence (EV), and ethnic war (EW). Episodes belonging to the last four of these categories constitute the universe of intrastate episodes that are of interest to our analysis. The magnitude scores for these episodes are aggregated into the *CIVTOT* variable in the MEPV dataset. *CIVTOT* is an annual ordinal index of civil conflict intensity at the country level that underlies the particular measure of ***quinquennial MEPV civil conflict severity*** we employ – namely, the maximum value of *CIVTOT* across all years in any given 5-year interval during the 1960–2017 time period. For further information on the data underlying our measure of civil conflict severity, the reader is referred to the codebook for the MEPV dataset.

5. **CNTS social conflict index:** This variable is based on the Domestic Conflict Event Data from the Cross-National Time Series (CNTS) Data Archive 2018 Edition (Banks and Wilson, 2018), which covers the 1815–2017 time period.

   Specifically, the basis of our CNTS social conflict index is the variable *Domestic9* from the CNTS Data Archive. *Domestic9* is an annual continuous index of the degree of social unrest, computed by first taking the weighted sum of the counts of different unrest/conflict events (given by the variables *domestic1-8*) in a country-year. As of October 2007, the weights employed were as follows: Assassinations (25), Strikes (20), Guerrilla Warfare (100), Government Crises (20), Purges (20), Riots (25), Revolutions (150), and Anti-Government Demonstrations (10). In a second step, the weighted sum is multiplied by 100/8 to obtain *Domestic9*. The specific measure used in our study is a ***quinquennial CNTS social conflict index***, calculated for each country as the maximum value of *Domestic9* across all years in any given 5-year interval during the 1960–2017 time period. For further information on the source data for our social conflict index, the reader is referred to the website of the CNTS Data Archive.

6. **UCDP nonstate conflict incidence:** This measure is based on information from Version 18.1 of the UCDP Non-State Conflict Dataset, covering the 1989–2017 time period (Sundberg et al., 2012).

   A non-state conflict is defined by the Uppsala Conflict Data Program (UCDP) as "the use of armed force between two organized armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year." An *organized group* can be either (i) a *formally organized group*, i.e., any non-governmental group of people having announced a name for their group and using armed force against another similarly organized group; or (ii) an *informally organized group*. The latter type of group does not have an announced name, but it uses armed force against another similarly organized group such that there is a clear pattern of violent incidents that are connected and in which both groups use armed force against the other. ***Quinquennial UCDP nonstate conflict incidence*** is coded 1 for any 5-year interval for a country if in any year during this interval there was at least one active (ongoing) non-state conflict in the country. A conflict is deemed active in a given calendar year if it resulted in at least 25 battle-related deaths during that year. For further information on the source data for our measure of non-state conflict incidence, the reader is referred to the codebook for Version 18.1 of the UCDP Non-State Conflict Dataset.

Other outcomes

1. **Number of ethnic groups:** The total number of distinct ethnic groups in a country's population, as compiled by Fearon (2003). The specific variable employed by our analysis is the natural logarithm of one plus the number of ethnic groups. See Fearon (2003) for additional details on primary data sources and methodological assumptions.

2. **Prevalence of interpersonal trust:** This variable is constructed using information from the World Values Survey (2006, 2009) (henceforth, WVS) on the prevalence of generalized interpersonal trust in a country's population. In particular, this well-known measure of social capital at the country level reflects the proportion of all respondents (from across five different waves of the WVS, conducted over the 1981–2009 time period) that opted for the answer "Most people can be trusted" (as opposed to "Can't be too careful") when responding to the survey question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" For additional details, the reader is referred to documentation available on the WVS website.

3. **Variation in political attitudes:** The intra-country dispersion in self-reported individual political positions on a "left"–"right" categorical scale, based on data from the WVS. Specifically, this measure of heterogeneity in political attitudes at the country level is calculated as the intra-country standard deviation across all respondents (sampled over five different waves of the WVS during the 1981–2009 time period) of their self-reported positions on a categorical scale from 1 (politically "left") to 10 (politically "right") when answering the survey question "In political matters, people talk of 'the left' and 'the right.' How would you place your views on this scale, generally speaking?" Given that this variable's unit of measurement does not possess a natural interpretation, we standardize the cross-country distribution of this variable prior to conducting our regressions. For additional details, the reader is referred to documentation available on the WVS website.

Main Control Variables

1. **Ethnic fractionalization:** This is the well-known ethnic fractionalization index of a country, reflecting the probability that two individuals, randomly selected from the country's population, will belong to different ethnic groups. Formally, a country's ethnic fractionalization index is calculated as follows:

$$FRAC = 1 - \sum_{i=1}^{n} p_i^2,$$

where p_i is the proportional representation of ethnic group i in the national population; and n is the total number of ethnic groups in the country. The specific variable we employ is based on the list of ethnic groups (and their national population shares) by country as compiled by Alesina et al. (2003). See Alesina et al. (2003) for additional details on primary data sources and methodological assumptions.

2. **Ethnolinguistic polarization:** An ethnolinguistic polarization index at the country level, calculated by applying the following definition of polarization due to Reynal-Querol (2002) and Montalvo and Reynal-Querol (2005):

$$POL = 4 \sum_{i=1}^{n} p_i^2 \left[1 - p_i\right],$$

where p_i is the proportional representation of linguistic group i in the national population; and n is the total number of linguistic groups in the country. The employed ethnolinguistic polarization index is sourced from the replication dataset of Desmet et al. (2012). The authors provide measures of several such polarization indices, constructed at different levels of aggregation of linguistic groups in a country's population (based on hierarchical linguistic trees). The specific polarization measure we use corresponds to the most disaggregated level of the linguistic tree, and it reflects the extent of polarization across subnational groups classified according to modern-day languages. See Desmet et al. (2012) for additional details on primary data sources and methodological assumptions.

3. **Absolute latitude:** The absolute value of the latitude of a country's geodesic centroid, as reported by the *At These Coordinates* resource repository, based on metadata from (i) the National Geospatial-Intelligence Agency's (NGA) GEOnet Names Server (GNS); and (ii) the United States Geological Survey's (USGS) Geographic Names Information System (GNIS).

4. **Ruggedness:** A measure of the degree of terrain ruggedness of a country's territory. Based on Riley et al. (1999), the ruggedness of a grid cell, i, is defined as

$$RIX(i) = \sqrt{\sum_{k=1}^{8} (h_i - h_{j_k})^2},$$

where h_l is the elevation (in meters above sea level) of cell $l = i, j_1, j_2, ..., j_8$, and the cells indexed by j are the eight neighboring cells of i. The country-level measure of ruggedness used by our study is the mean value of $RIX(i)$ across all 1 km $\times$ 1 km grid cells of a country. The cell-level ruggedness index is computed by Özak (2010), based on topographical data from

the Global Land One-Kilometer Base Elevation (GLOBE) digital elevation model (Hastings et al., 1999).

5. **Mean and range of elevation:** The country-level mean and range of elevation (in thousands of kilometers above sea level), calculated using geospatial elevation data at a 1-degree resolution from the Geographically based Economic data (G-ECON) project (Nordhaus, 2006), based on similar data at a 10-minute resolution from New et al. (2002). The mean of elevation at the country level reflects the average value across the grid cells that are located within a country's national borders, whereas the range of elevation reflects the difference between the maximum and minimum values across the same set of grid cells. See the G-ECON project website for additional details.

6. **Mean and range of land suitability:** The country-level mean and range of a geospatial index of the suitability of land for agriculture, based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as on ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH. This index was initially developed at a half-degree resolution by Ramankutty et al. (2002), and it has been aggregated to the country level by Michalopoulos (2012), with the mean at the country level reflecting the average value of the index across the grid cells that are located within a country's national borders, and the range reflecting the difference between the maximum and minimum values of the index across the same set of grid cells. See Michalopoulos (2012) for additional details.

7. **Island nation:** An indicator for whether a country shares a land border with any other country, as reported by the CIA's World Factbook. Of the 147 countries in our baseline sample, the following 7 are coded as island nations: Australia, Cuba, Japan, Sri Lanka, Madagascar, New Zealand, and Philippines.

8. **Distance to nearest waterway:** The distance (in thousands of kilometers) from a grid cell to the nearest ice-free coastline or sea-navigable river, averaged across the grid cells of a country. This variable was originally constructed by Gallup et al. (1999) and is available from the Research Datasets online repository maintained by Harvard University's Center for International Development.

9. **Colonial history:** A set of three indicators reflecting a country's experience of colonial rule by (i) the U.K., (ii) France, or (iii) any other major colonizing power, respectively. Therefore, the omitted category is the absence of colonial rule. These variables are constructed based on information from various sources, including the CIA's World Factbook, the Encyclopaedia Brittanica, Country Studies of the Library of Congress, and rulers.org amongst others. Additional details are available from the authors upon request.

   In cross-sectional regressions at the country level, the relevant measures comprise time-invariant indicators for the historical presence of colonial rule – i.e., whether the country has ever been ruled by the colonizing power in question. In regressions using repeated cross-country data, the relevant measures comprise time-varying indicators of the lagged prevalence of colonial rule – i.e., whether the country was ruled by the colonizing power in question at any point in the preceding 5-year time interval or in the preceding year, depending on the temporal dimension of the repeated cross-section.

10. **Legal origins:** A set of two time-invariant indicators for British and French legal origins, as reported by La Porta et al. (1999). Specifically, these indicators identify whether the legal origin of country's Company Law or Commercial Code is (i) the English Common Law or (ii) the French Commercial Code, respectively. The omitted category is German, Scandinavian, or Socialist legal origins, as recognized by La Porta et al. (1999).

11. **Executive constraints:** An index, reported at an annual frequency as a 7-point categorical variable (from 1 to 7) by the Polity IV Project (Version 2017), quantifying the extent of institutionalized constraints on the decision-making power of chief executives (Marshall et al., 2017). The specific version of the Polity IV Project dataset employed by our study covers the 1800–2017 time period. For further information on the index of executive constraints, the reader is referred to the codebook for Version 2017 of the Polity IV Project dataset.

    In cross-sectional regressions at the country level, the relevant measure is the temporal average of the index across all years in the 1960–2017 time period. In regressions using quinquennially repeated cross-country data, the relevant measure is the temporal average of the index across all years in the preceding 5-year time interval. Finally, in regressions based on annually repeated cross-country data, the relevant measure is the value of the index from the preceding year.

12. **Type of political regime:** Our measures of the type of political regime are based on two indicators reflecting whether a country is classified as a democracy (or not) and as an autocracy (or not) in a given year. The omitted category is anocracy, a hybrid regime that constitutes the middle range of the autocracy-democracy political spectrum. This regime classification is based on the *POLITY2* index (the Revised Combined Polity Score), as reported at an annual frequency by the Polity IV Project (Version 2017) for the 1800–2017 time period (Marshall et al., 2017). *POLITY2* is a discrete index that ranges from -10 (strongly autocratic) to +10 (strongly democratic). Following the norm in the literature, a country-year is coded as a *democracy* if the *POLITY2* score is above 5 or as an *autocracy* if the score is below -5. The prevalence of *anocracy*, occurring when the *POLITY2* score is between -5 and 5 for a country-year, therefore serves as the omitted political regime category. For further information on the *POLITY2* index, the reader is referred to the codebook for Version 2017 of the Polity IV Project dataset.

    In cross-sectional regressions at the country level, the relevant measures of regime type are the fractions of years during the 1960–2017 time period that a country spent as a democracy and as an autocracy, respectively. In regressions using quinquennially repeated cross-country data, the relevant measures are the fractions of years during the preceding 5-year time interval that a country spent as a democracy and as an autocracy, respectively. Finally, in regressions based on annually repeated cross-country data, the relevant measures are the indicators for democracy and autocracy for the preceding year.

13. **Oil or gas reserve discovery:** A time-invariant indicator of at least one petroleum (oil or gas) reserve on the land territory of a country. This variable is based on information provided in the Petroleum Dataset (Version 1.2), covering the 1946–2003 time period (Lujala et al., 2007). Therefore, the available data does not provide information about any petroleum deposit discovered after 2003. The dataset is compiled for the main purpose of investigating the relationship between armed civil conflict and natural resources. Each on-shore petroleum (oil or gas) reserve – identified as polygons in the shapefile accompanying the dataset – is

assigned to a modern-day country based on the coordinates of the centroids of the deposit polygons. For additional information, the reader is referred to the codebook for Version 1.2 of the Petroleum Dataset, available from the Geographical and Resource Datasets online repository maintained by PRIO.

14. **Log population size:** The log-transformed size of a country's population, as reported by the World Bank's World Development Indicators (WDI) online data catalog.

    In cross-sectional regressions at the country level, the relevant measure is the log-transformed temporal average of annual population observations across all years in the 1960–2017 time period. In regressions using quinquennially repeated cross-country data, the relevant measure is the log-transformed temporal average of observations across all years in the preceding 5-year time interval. Finally, in regressions based on annually repeated cross-country data, the relevant measure is the log-transformed observation from the preceding year.

15. **Log GDP per capita:** The log-transformed per-capita GDP (in current US$) of a country, as reported by the World Bank's World Development Indicators (WDI) online data catalog.

    In cross-sectional regressions at the country level, the relevant measure is the log-transformed temporal average of annual per-capita GDP observations across all years in the 1960–2017 time period. In regressions using quinquennially repeated cross-country data, the relevant measure is the log-transformed temporal average of observations across all years in the preceding 5-year time interval. Finally, in regressions based on annually repeated cross-country data, the relevant measure is the log-transformed observation from the preceding year.

Other Control Variables (for Robustness Checks)

1. **Ecological fractionalization and polarization:** These measures of ecological diversity are motivated by Fenske (2014). The measure of *ecological fractionalization* is a Herfindahl index, constructed as

$$\text{Ecological fractionalization}_i = 1 - \sum_{t=1}^{t=18} (s_i^t)^2;$$

   and *ecological polarization* index is given by

$$\text{Ecological polarization}_i = 1 - \sum_{t=1}^{t=18} \left(\frac{0.5 - s_i^t}{0.5} \right)^2 s_i^t,$$

   where s_i^t is the share of the area of country i that is occupied by ecological type t. The polarization index measures the degree to which a country's area approximates a territory in which two ecological types each occupy half the total area. The relevant information on the spatial distribution of ecological types across the land surface of the earth is derived from global maps of agro-ecological zones from the Food and Agriculture Organization (FAO) of the United Nations.

2. **Mean and volatility of temperature and precipitation:** These four variables are constructed using information on mean temperature (in degree Celcius) per annum and total precipitation (in mm) per annum as reported by the Climate Research Unit (CRU) (Harris et al., 2014). Specifically, we employ the country-level spatial aggregates of annual mean

temperature and annual total precipitation, provided the CRU CY Version 4.01 dataset, which spans the 1901–2016 time period.

In cross-sectional regressions at the country level, the relevant measures of mean temperature and total precipitation reflect the temporal averages of the annual observations of these variables across all years in the 1960–2017 time period, whereas the corresponding volatility measures capture their respective temporal standard deviations during the same time span. In regressions using quinquennially repeated cross-country data, the relevant mean and volatility measures are similarly defined, except that the temporal averages and standard deviations are calculated across the years of the preceding 5-year time interval (rather than the full sample period). Finally, in regressions based on annually repeated cross-country data, the relevant measures are the one-year lags of annual mean temperature and annual total precipitation as well as the interannual standard deviations of temperature and precipitation over a 5-year rolling window that ends in the preceding year.

3. **Log years since Neolithic Revolution:** The log-transformed number of thousand years elapsed (as of the year 2000) since the majority of the population residing in a territory defined by a country's modern national borders began practicing sedentary agriculture as the primary mode of subsistence. This measure, initially reported by Putterman (2008), is compiled using a host of both region- and country-specific archaeological studies as well as more general encyclopedic works on the transition from hunting and gathering to agriculture during the Neolithic Revolution. The reader is referred to Putterman's website for a detailed description of the primary and secondary data sources employed in the construction of this variable.

4. **Log index of state antiquity:** The log-transformation of an index reflecting a country's cumulative experience with institutionalized statehood since antiquity. Specifically, we employ the State Antiquity Index (version 3.1), first introduced by Bockstette et al. (2002). The underlying index quantifies the exposure of a territory – as defined by a country's modern national borders – to formal statehood (i.e., being an independent nation-state or part of a larger kingdom or an empire) since the year 1 CE and until 1950. In particular, for each 50-year time interval, information on a territory's status with respect to the following 3 questions (each with specific weights applied) is employed: (i) is there a government above the tribal level?; (ii) is this government foreign or locally based?; and (iii) how much of the territory of the modern country was ruled by this government? These information are then aggregated over time to produce an index that ranges between 0 and 1. The reader is referred to Putterman's website for a detailed description of the methodology and data sources employed in the construction of this index.

5. **Log duration of human settlement:** The natural logarithm of the maximum duration (in tens of thousands of years) of uninterrupted settlement by anatomically modern humans across locations in a territory defined by a country's modern national borders. The underlying measure is obtained from the dataset of Ahlerup and Olsson (2012). The reader is therefore referred to that work for additional details on data sources and methodological assumptions.

6. **Log distance from regional frontier in 1500:** The great circle distance from a country's capital city to the closest regional technological frontier around the year 1500. The variable is obtained from the dataset of Ashraf and Galor (2013a). The set of regional frontiers

comprises the two most populous cities, reported for the year 1500 and belonging to different civilizations or sociopolitical entities, from each of Africa, Europe, Asia, and the Americas. Distances are calculated using the Haversine formula and are measured in kilometers. The historical urban population data used to identify the frontiers are sourced from Chandler (1987) and Modelski (2003), and the geographical coordinates of ancient urban centers are sourced from online resources such as Wikipedia.

7. **Ethnic inequality in luminosity:** A measure of intra-country economic inequality as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across the georeferenced homelands of ethnic groups. This measure is sourced from the replication dataset of Alesina et al. (2016). The reader is therefore referred to that work for additional details on data sources and methodological assumptions.

8. **Spatial inequality in luminosity:** A measure of intra-country economic inequality as captured by the subnational spatial distribution of per-capita adjusted nighttime luminosity in the year 2000 across 2.5×2.5-degree geospatial grid cells. This measure is sourced from the replication dataset of Alesina et al. (2016). The reader is therefore referred to that work for additional details on data sources and methodological assumptions.

9. **Linguistic fractionalization and polarization (georeferenced):** These are the country-level counterparts of the measures of linguistic fractionalization and polarization that are used in our analysis of conflicts at the ethnic homelands level. Specifically, these measures are constructed using georeferenced information on the spatial distribution of language homelands from the World Language Mapping System (WLMS) along with gridded population data from the Gridded Population of the World (GPW) dataset.

10. **Ethnic fractionalization (Fearon, 2003):** The ethnic fractionalization index compiled by Fearon (2003). The index reflects the probability that two individuals, randomly selected from a country's population, will belong to different ethnic groups.

11. **Linguistic fractionalization (Alesina et al., 2003):** The linguistic fractionalization index compiled by Alesina et al. (2003). The index reflects the probability that two individuals, randomly selected from a country's population, will belong to different linguistic groups.

12. **Religious fractionalization (Alesina et al., 2003):** The religious fractionalization index compiled by Alesina et al. (2003). The index reflects the probability that two individuals, randomly selected from a country's population, will belong to different religions.

13. **Ethnolinguistic fractionalization (Esteban et al., 2012):** An index of ethnolinguistic fractionalization, as represented by the *frac_fear* variable in the replication dataset of Esteban et al. (2012). The underlying ethnolinguistic population shares are sourced from Fearon (2003).

14. **Ethnolinguistic polarization (Esteban et al., 2012):** The Esteban-Ray index of ethnolinguistic polarization with $\delta = 0.05$, as represented by the *er_fear_delta005* variable in the replication dataset of Esteban et al. (2012). The underlying ethnolinguistic population shares are sourced from Fearon (2003).

15. **Gini index of ethnolinguistic diversity (Esteban et al., 2012):** The gini index of ethnolinguistic diversity per capita with $\delta = 0.05$, as represented by the variable named *gini_fear_delta005_PERCAPTA* in the replication dataset of Esteban et al. (2012). It is obtained after dividing the gini index of ethnolinguistic diversity by population size. The underlying ethnolinguistic population shares are sourced from Fearon (2003).

16. **Log percentage mountainous terrain:** The log-transformation of the proportion (in percentage) of a country's territory that is "mountainous" according to the codings of the geographer A.J. Gerard. This variable is sourced from the replication dataset of Fearon and Laitin (2003), where it is used to test the hypothesis that "rough terrain, poorly served by roads, at a distance from the centers of state power should favor insurgency and civil war."

17. **Noncontiguous state dummy:** A time-invariant indicator of whether a country possesses a territory with a population of at least 10,000 that is separated from the region containing its capital city either by land or 100 kilometers of water. This variable is sourced from the replication dataset of Fearon and Laitin (2003), where it is used to test the hypothesis that "the presence of a territory that is separated from the center of national governance by water or distance can help rebels more easily sustain insurgent activity and, thereby, make civil war more likely."

18. **Disease richness:** The total number of different types of infectious diseases in a country as reported by Fincher and Thornhill (2008), based on the Global Infectious Disease and Epidemiology Network (GIDEON; www.gideononline.com).

19. **Ethnic dominance:** A time-invariant indicator of whether the largest ethnic group in a country constitutes 45-90% of the national population. This variable is sourced from the replication dataset of Hegre and Sambanis (2006), but the primary source of the measure is Collier and Hoeffler (2004).

20. **Political instability:** A time-varying indicator at the country-year level of whether there was a change in the Polity IV regime index by 3 or more points in any of the three years prior to the country-year in question. Periods of regime transition (-88) and "interruptions" (indicating a complete collapse of central authority) are also coded as cases of political instability. Episodes of foreign occupation, however, are treated as missing observations. In robustness checks of our civil conflict onset regressions, the one-year lagged value of this variable is employed. This variable is sourced from the replication dataset of Hegre and Sambanis (2006), but the primary source is Fearon and Laitin (2003).

21. **New state dummy:** A time-varying indicator at the country-year level for whether the current year is the first year of the country's existence (e.g., as a newly independent state from colonial rule). In robustness checks of our civil conflict onset regressions, the one-year lagged value of this variable is employed. This variable is sourced from the replication dataset of Hegre and Sambanis (2006).

22. **Commodity export price shocks:** A set of four variables capturing different types of commodity export price shocks on an annual basis, sourced from the replication dataset of Bazzi and Blattman (2014). The first variable reflects *aggregate price shocks* and is computed

as the annual change in a country's log commodity export price index (a geometric average of all commodity export prices weighted by lagged export shares). The remaining variables reflect three types of disaggregated price shocks. The first of these reflects *annual crop price shocks*, i.e., price shocks to annual agricultural goods, such as oilseeds, food crops, and livestock, that are more likely to accrue to households. The second reflects *perennial crop price shocks*, i.e., price shocks to perennial tree crops like cocoa, coffee, rubber, or lumber. Finally, the third type of disaggregated price shocks captures *extractive crop price shocks*, i.e., price shocks to extractive products, namely, minerals, oil, and gas, that are more likely accrue to states. By construction, the sum of the three disaggregated types of shocks yields the *aggregate price shock* variable. In robustness checks of our civil conflict onset regressions, we employ the contemporaneous as well as the one- and two-year lagged values of these various commodity export price shock variables. For additional details, the reader is referred to Bazzi and Blattman (2014).

Supplement B Supplement to the Ethnicity-Level Analyses

B.1 Construction of the Georeferenced Ethnicity-Level Dataset

This research constructs a novel geo-referenced data set of population diversity for a large number of ethnic groups across the globe. Two measures are constructed: (i) a measure of genetic diversity for 207 ethnic homelands for all individuals covered in the Pemberton et al. (2013) dataset that can be mapped to an ethnic homeland, and (ii) a measure of predicted population diversity for 901 ethnic homelands covered in the Geo-Referencing of Ethnic Groups (GREG) map of Weidmann et al. (2010).

The geo-referenced dataset for observed genetic diversity maps all 10,386 linkable individuals in the Pemberton et al. (2013) dataset into their ethnic homelands. This mapping results in a sample of 207 ethnic homelands for which, in addition to the measure of genetic diversity, spatial characteristics (e.g., geographic, climatic, and societal attributes) are available. Furthermore, using data on the spatial distribution of language areas in conjunction with data on the spatial distribution of population sizes, the study generates measures of linguistic fractionalization and polarization for each ethnic homeland. Finally, using gridded PRIO data (PRIO-GRID version 1.01) as reported by Tollefsen et al. (2012) based on the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002) as well as data on UCDP Georeferenced conflict events (Sundberg et al., 2012; Croicu and Sundberg, 2015) the study generates a range of measures of conflict within each ethnic homeland.

The mapping of the 10,386 linkable individuals in the Pemberton et al. (2013) dataset into their ethnic homelands was based on the individual's ethnic identity, location, and geographical coordinates, where the polygons for the ethnic homelands were based on (i) polygons found in Murdock (1959) and digitized by Nunn (2008); Nunn and Wantchekon (2011), (ii) the Handbook of North American Indians (Heizer, 1978), (iii) Global Mapping International's World Language Mapping System (WLMS) (see http://worldgeodatasets.com/language), (iv) the Geo-Referencing of Ethnic Groups (GREG) map of Weidmann et al. (2010), and (v) the Database of Global Administrative Areas (GADM) map version 3.6 (gadm.org).

The geo-referenced dataset for predicted predicted population diversity for 901 ethnic homelands covered in the Geo-Referencing of Ethnic Groups (GREG) map of Weidmann et al. (2010) is constructed based on the migratory distance from Addis Ababa in East Africa to the centroid of the homeland.[1]

B.2 Variable Definitions for the Ethnic-level Analyses

Conflict measures

1. **Conflict prevalence:** The average yearly share of the area of each ethnic homeland, over the period 1989–2008, that was within the boundaries of internal armed conflict event (between the government of a state and internal opposition groups). This measure is calculated using the gridded PRIO data (PRIO-GRID version 1.01) as reported by Tollefsen et al. (2012) based on the UCDP/PRIO Armed Conflict Dataset (Gleditsch et al., 2002).

2. **Number of conflict events:** The number of conflict events within each ethnic homeland in the UCDP Georeferenced Event Dataset covering the period 1989–2017 (Sundberg et al., 2012; Croicu and Sundberg, 2015).

[1] One homeland spanning territories in South America and Mauritius labeled "Indians of India and Pakistan" is excluded from the sample. The qualitative results would not be affected by the inclusion of this territory.

3. **Number of deaths:** The best (i.e., most likely) estimate of total fatalities resulting from a conflict event within each ethnic homeland in the UCDP Georeferenced Event Dataset covering the period 1989–2017 (Sundberg et al., 2012; Croicu and Sundberg, 2015).

4. **Number of deaths per event:** The number of deaths per event within each ethnic homeland in the UCDP Georeferenced Event Dataset covering the period 1989–2017 (Sundberg et al., 2012; Croicu and Sundberg, 2015).

Trust-related measures

1. **Intra-group trust (Africa):** The measure of an individual's trust in individuals from the same ethnic group in the 2005 Afrobarometer survey (3rd wave), as linked by Nunn and Wantchekon (2011) to the ethnicity names used in the Ethnographic Atlas. The measure takes the value 0 if the response to the question "How much do you trust each of the following types of people: People from your own ethnic group?" is "not at all", 1 if the response is "just a little", 2 if the value is "I trust them somewhat" and 3 if the value is "I trust them a lot".

2. **Slave exports (Africa):** A measure of the number of slaves taken from each ethnicity in transatlantic and Indian Ocean slave trades. The measure comes from Nunn and Wantchekon (2011) and is based on data from Nunn (2008).

3. **Other control variables (Africa):** The measures come from Nunn and Wantchekon (2011) and are based on data from 2005 Afrobarometer survey (3rd wave).

4. **Trust (US):** A measure of an individual's trust in people in general based on data from the General Social Survey 1972–2014 Release 6b Smith et al. (2018). The measure takes the value 1 if the response to the question "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" is "cannot trust", 2 if the response is "depends", and 3 if the value is "can trust".

Migratory distance and interpersonal population diversity

1. **Observed population diversity:** The expected heterozygosity (genetic diversity) of individuals in each of the 207 ethnic homelands, as calculated using Nei's formula (Nei, 1973), based on the individual-level data from Pemberton et al. (2013).

2. **Predicted population diversity:** The predicted level of population diversity of an ethnic homeland based on the migratory distance from East Africa to the centroid of the homeland, using the linear regression fit between observed population diversity and migratory distance from Addis Ababa obtained in sample of 207 ethnic homelands for which observed genetic diversity is available. The migratory distance from Addis is defined as the shortest traversable paths from Addis Ababa to the centroid of each ethnic group was computed. Given the limited ability of humans to travel across large bodies of water, the traversable area included bodies of water at a distance of 100km from land mass (excluding migration from Africa into Europe via Italy or Spain).[2]

[2]For the computation of predicted population diversity, distances to islands, where travel on water exceeds 100kms, are ignored since the Serial Founder Effect requires the serial foundation of populations along the migratory path and this was not feasible on water.

Control variables

1. **Linguistic fractionalization and polarization:** The degree of fractionalization in the ethnic homeland, using the formula $1 - \sum_i s_i^2$, and the degree of polarization in the ethnic homeland, using the formula $4 \sum_i s_i^2(1 - s_i)$, where s_i is an estimate of the population share of language group i in the homeland. Using the WLMS map of the spatial distribution of language areas in conjunction with the Gridded Population of the World dataset, the study estimates the number of individuals living in each intersection between ethnic homelands and language areas, assuming that population counts in overlapping language areas are equally split between these languages.

2. **Absolute latitude:** The absolute value of the latitude of an ethnic homeland's geodesic centroid, or, when the centroid is outside of the homeland, a representative interior point.

3. **Ruggedness:** The average level of the Terrain Ruggedness Index measure of Nunn and Puga (2012) across the grid cells that are located within a homeland.

4. **Mean and range of elevation:** The mean and range of elevation above sea level of an ethnic homeland, calculated using geospatial data from the *Atlas of the Biosphere* project (nelson.wisc.edu/sage/data-and-models/atlas/), across the grid cells that are located within a homeland.[3]

5. **Mean and range of land suitability:** The mean and range of the post-1500 optimal Caloric Suitability Index, measured by Galor and Özak (2016), across the grid cells that are located within a homeland.

6. **Island location:** A dummy variable indicating if the land type of an ethnic homeland's geodesic centroid (or a representative interior point) is a "small island" or a "very small island" as reported in the *World Countries* geographical dataset provided by ESRI (arcgis.com/home/item.html?id=ac80670eb213440ea5899bbf92a04998).

7. **Distance to nearest waterway:** The mean of the geodesic distance to the nearest coast or river, across the grid cells that are located within a homeland. Coastline locations are reported in the *Global Self-consistent, Hierarchical, High-resolution Geography Database* (http://soest.hawaii.edu/pwessel/gshhg). River locations are reported in the 1:10m *Natural Earth River + Lake Centerlines* dataset version 4 (http://naturalearthdata.com/downloads/10m-physical-vectors/10m-rivers-lake-centerlines).

8. **Temperature:** The mean of the daily average temperature (in degree Celcius), across the grid cells that are located within a homeland, based on data from the CRU TS dataset version 3.21 for the period 1901–2012, as reported by Climate Research Unit (CRU) (Harris et al., 2014).

9. **Precipitation:** The mean of the annual total precipitation (in mm), across the grid cells that are located within a homeland, based on data from the CRU TS dataset version 3.21 for the period 1901–2012, as reported by Climate Research Unit (CRU) (Harris et al., 2014).

[3]The mean elevation can be negative in some cases due to the existence of places on land with elevation below sea level or the inclusion of territories at sea in the homeland polygon, for which the elevation is negative.

10. **Time since settlement:** The earliest year with a positive population count estimate in the ethnic homeland. Specifically, the study employs the population count data from the *History Database of the Global Environment* dataset version 3.1 (themasites.pbl.nl/tridion/en/themasites/hyde/download/index-2.html), described in Klein Goldewijk et al. (2010, 2011).

11. **Malaria:** The mean level of plasmodium falciparum malaria endemicity in 2010, across the grid cells that are located within a homeland. Specifically, the current study employs the data on the age-standardised plasmodium falciparum Parasite Rate from Gething et al. (2011). It represents the estimated proportion of 2–10 year olds in the general population that are infected with plasmodium falciparum, averaged over the months of 2010. The estimates are based on data from parasite rate surveys and a geostatistical model that produces a range of predicted endemicities for each location. The model includes environmental covariates which improves the accuracy of the prediction. The environmental covariates include rainfall, temperature, land cover and urban/rural status. The endemicity data reports the mean value for the probability distribution at each location (approx. $1km^2$).

12. **Oil or gas reserve discovery:** A time-constant dummy for the presence of at least one petroleum (oil or gas) reserve on the territory of an ethnic homeland. The variable is based on information provided in the Petroleum Dataset (version 1.2, dated 2009) covering the period 1946–2003 (Lujala et al., 2007). The dataset is compiled for the main purpose of investigating the relationship between armed civil conflict and natural resources. Each on-shore petroleum reserve (oil or gas) – indicated as polygons in the shapefile accompanying the dataset – is assigned to an ethnic homeland using the coordinates of the centerpoints of the deposit polygons.

13. **Luminosity:** The mean level of cloud-free nighttime light intensity for the years 1992–2013, accross the grid cells that are located within a homeland. Specifically, the current study employs all available data in version 4 of the Defense Meteorological Satellite Program – Operational Linescan System (DMSP-OLS) Nighttime Lights Time Series (ngdc.noaa.gov/eog/dmsp/downloadV4composites.html). Since the log of zero is undefined, log luminosity is defined as the log of the sum of 0.001 and the luminosity measure.

B.3 Robustness Checks

TABLE SB.I: Population Diversity and Conflict across Ethnic Homelands – Robustness to Accounting for Alternative Distances

	Log conflict prevalence					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Observed population diversity	28.338***	31.342***	30.591***			
	(9.622)	(9.692)	(9.735)			
Predicted population diversity				73.828***	70.194***	75.334***
				(7.390)	(7.313)	(7.305)
Distance to Technological Frontier in Year 1 (in 1000 kms)	-0.045			-0.172***		
	(0.163)			(0.066)		
Distance to Technological Frontier in Year 1000 (in 1000 kms)		-0.324*			-0.268***	
		(0.168)			(0.062)	
Distance to Technological Frontier in Year 1500 (in 1000 kms)			-0.210			-0.124**
			(0.148)			(0.061)
Ethnolinguistic fractionalization	1.633	1.446	1.474	0.279	0.340	0.330
	(1.219)	(1.171)	(1.196)	(0.383)	(0.381)	(0.381)
Ethnolinguistic polarization	-0.353	-0.213	-0.237	0.332	0.315	0.296
	(1.029)	(0.990)	(1.010)	(0.348)	(0.344)	(0.347)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Observed	Observed	Observed	Predicted	Predicted	Predicted
Observations	207	207	207	901	901	901
Effect of 10th90th %ile move in diversity	0.443***	0.490***	0.478***	1.639***	1.558***	1.672***
	(0.150)	(0.152)	(0.152)	(0.164)	(0.162)	(0.162)
First-stage F statistic						
Adjusted R^2	0.304	0.316	0.310	0.367	0.375	0.365
β^*	26.359	28.224	29.899	80.379	77.719	77.280

Notes: This table exploits variations across ethnic homelands to establish a significant positive impact of observed and predicted population diversity on the log conflict prevalence during the 1989–2008 period, conditional on migratory distances from historical technological frontiers as well as the baseline geographical characteristics. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its distribution is expressed in terms of the change in the prevalence of conflicts within the territory of a homeland over the years 1989–2008. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SB.II: Observed Population Diversity and Conflict across Ethnic Homelands – Robustness to Accounting for Measures of Ecological Diversity

	Log conflict prevalence						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS
Observed population diversity	27.700***	32.958***	24.748***	25.591***	24.996***	26.869**	26.325**
	(10.372)	(10.482)	(9.315)	(9.313)	(9.287)	(10.427)	(10.425)
Ecological diversity	-0.838	-0.637	1.029	0.748	0.909	0.733	0.843
	(1.430)	(1.595)	(1.429)	(1.418)	(1.414)	(1.384)	(1.379)
Ecological polarization	0.942	1.103	0.675	0.702	0.687	1.006	1.009
	(1.141)	(1.228)	(1.065)	(1.045)	(1.054)	(1.024)	(1.025)
Ethnolinguistic fractionalization				1.140*		0.893	
				(0.636)		(0.652)	
Ethnolinguistic polarization					0.734		0.641
					(0.527)		(0.530)
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	No	No	Yes	Yes	Yes	Yes	Yes
Development outcomes	No	No	No	No	No	Yes	Yes
Disease environment controls	No	No	No	No	No	Yes	Yes
Sample	Observed	Observed	Observed	Observed	Observed	Observed	Observed
Observations	205	205	205	205	205	205	205
Effect of 10th-90th %ile move in diversity	0.433***	0.515***	0.387***	0.400***	0.391***	0.420***	0.411**
	(0.162)	(0.164)	(0.146)	(0.146)	(0.145)	(0.163)	(0.163)
Adjusted R^2	0.106	0.168	0.308	0.317	0.312	0.330	0.328
β^*		37.005	23.299	24.574	23.683	26.483	25.685

Notes: This table exploits cross-ethnicity variations to establish a significant positive impact of contemporary population diversity on the log spatio-temporal prevalence of UCDP/PRIO conflicts during the 1989–2008 period, conditional on ecological diversity and ecological polarization as well as the baseline control variables. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The 2SLS regressions exploit prehistoric migratory distance from East Africa to each ethnic homeland as an excluded instrument for the observed population diversity of this ethnic group. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the change in the average yearly share of the area of each ethnic homeland that was within the boundaries of internal armed conflict over the period 1989–2008. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SB.III: Predicted Population Diversity and the Spatiotemporal Prevalence of Conflict across Ethnic Homelands – Robustness to Accounting for Measures of Ecological Diversity

	Log conflict prevalence						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Predicted population diversity	77.597***	79.803***	76.148***	75.668***	77.910***	77.646***	
	(6.245)	(7.314)	(7.425)	(7.458)	(9.700)	(9.807)	
Observed population diversity							130.105***
							(33.284)
Ecological diversity	0.711	0.808	1.064*	1.070*	1.565**	1.496**	-0.078
	(0.631)	(0.638)	(0.629)	(0.634)	(0.714)	(0.719)	(1.722)
Ecological polarization	0.396	0.466	0.317	0.299	-0.455	-0.435	0.263
	(0.587)	(0.541)	(0.533)	(0.536)	(0.596)	(0.599)	(1.233)
Ethnolinguistic fractionalization			0.341		0.174		
			(0.300)		(0.354)		
Ethnolinguistic polarization				0.450*		0.565*	
				(0.267)		(0.315)	
Regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Climatic controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Development outcomes	No	No	Yes	Yes	Yes	Yes	No
Disease environment controls	No	No	Yes	Yes	Yes	Yes	No
Sample	Predicted	Predicted	Predicted	Predicted	Old World	Old World	Observed
Observations	891	891	891	891	697	697	205
Effect of 10th-90th %ile move in diversity	1.748***	1.797***	1.715***	1.704***	0.976***	0.972***	2.034***
	(0.141)	(0.165)	(0.167)	(0.168)	(0.121)	(0.123)	(0.520)
Adjusted R^2	0.207	0.365	0.381	0.382	0.406	0.409	
β^*		81.333	75.203	74.414	69.099	68.719	
Migratory distance from East Africa (in 10,000 km)							-0.043***
							(0.009)
First-stage F-statistic							23.605

Notes: This table exploits cross-ethnicity variations to establish a significant positive impact of predicted population diversity on the log spatio-temporal prevalence of UCDP/PRIO conflicts during the 1989–2008 period, conditional on ecological diversity and ecological polarization as well as the baseline control variables. The set of continent and regional dummies includes indicators for Europe, Asia, North America, South America, Oceania, North Africa, and Sub-Saharan Africa. Additional climatic covariates refer to the average diurnal temperature range, average cloud cover, and average temperature range in the homeland. The 2SLS regressions exploit prehistoric migratory distance from East Africa to each ethnic homeland as an excluded instrument for the observed population diversity of this ethnic group. The estimated effect associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the change in the average yearly share of the area of each ethnic homeland that was within the boundaries of internal armed conflict over the period 1989–2008. Heteroskedasticity-robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

B.4 Descriptive Statistics for the Trust Analyses

TABLE SB.IV: Summary Statistics

			Percentile		
	Mean	SD	10th	90th	N
PANEL A	African sample				
Intra-group trust	1.52	1.00	0.00	3.00	3,212
Population diversity (observed)	0.76	0.00	0.76	0.77	3,212
Age	35.82	14.54	20.00	58.00	3,212
Male	0.49	0.50	0.00	1.00	3,212
Ethnic fractionalization	0.27	0.28	0.00	0.72	3,212
Ethnolinguistic polarization	0.53	0.13	0.30	0.62	3,212
Proportion of ethnic group in district	0.73	0.33	0.12	1.00	3,212
School present	0.84	0.37	0.00	1.00	3,208
Electricity present	0.65	0.48	0.00	1.00	3,210
Piped water present	0.44	0.50	0.00	1.00	3,157
Sewage present	0.23	0.42	0.00	1.00	3,054
Health clinic present	0.58	0.49	0.00	1.00	3,060
Living in an urban area	0.44	0.50	0.00	1.00	3,212
Living condition categories	2.65	1.25	1.00	4.00	3,206
Education categories	3.51	2.10	0.00	6.00	3,207
Occupation categories	18.92	92.10	1.00	23.00	3,201
Religion categories	10.52	51.36	2.00	12.00	3,204
Slave exports (Atlantic and Indian)	277.44	262.45	0.17	665.97	3,212
PANEL B	US sample				
Trust	1.88	0.97	1.00	3.00	2,294
Population diversity (predicted)	0.72	0.02	0.67	0.74	2,294
GSS year	1993.94	10.59	1980.00	2010.00	2,294
Age	54.37	19.46	27.00	80.00	2,284
Sex	1.55	0.50	1.00	2.00	2,294
Family income categories	2.73	0.89	2.00	4.00	1,803
Religion categories	2.02	1.29	1.00	3.00	2,283
Highest educational degree categories	1.30	1.20	0.00	3.00	2,290
Ethnic fractionalization (ancestral)	0.23	0.18	0.11	0.54	2,294
Ethnolinguistic polarization (ancestral)	0.41	0.21	0.12	0.67	2,294
Absolute latitude (ancestral)	46.07	11.82	23.00	60.00	2,294
Ruggedness (ancestral)	131.80	94.05	30.64	237.76	2,294
Mean elevation (ancestral)	436.42	339.34	105.77	1015.28	2,294
Mean land suitability (ancestral)	0.48	0.21	0.10	0.75	2,294
Range of land suitability (ancestral)	0.92	0.12	0.82	1.00	2,294
Distance to nearest waterway (ancestral)	223.00	496.37	29.43	332.58	2,294

Supplemental References

Ahlerup, P. and O. Olsson (2012): "The Roots of Ethnic Diversity," *Journal of Economic Growth*, 17, 71–102.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003): "Fractionalization," *Journal of Economic Growth*, 8, 155–194.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016): "Ethnic Inequality," *Journal of Political Economy*, 124, 428–488.

Ashraf, Q. and O. Galor (2013a): "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103, 1–48.

Banks, A. S. and K. A. Wilson (2018): "Cross-National Time-Series Data Archive [Data file]," Databanks International, Jerusalem, Israel. https://www.cntsdata.com/.

Bazzi, S. and C. Blattman (2014): "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, 6, 1–38.

Birnir, J. K., D. D. Laitin, J. Wilkenfeld, D. M. Waguespack, A. S. Hultquist, and T. R. Gurr (2018): "Introducing the AMAR (All Minorities at Risk) Data," *Journal of Conflict Resolution*, 62, 203–226.

Birnir, J. K., J. Wilkenfeld, J. D. Fearon, D. D. Laitin, T. R. Gurr, D. Brancati, S. M. Saideman, A. Pate, and A. S. Hultquist (2015): "Socially Relevant Ethnic Groups, Ethnic Structure, and AMAR," *Journal of Peace Research*, 52, 110–115.

Bockstette, V., A. Chanda, and L. Putterman (2002): "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth*, 7, 347–369.

Brecke, P. (1999): "Violent Conflicts 1400 A.D. to the Present in Different Regions of the World," Paper presented at the 1999 Annual Meeting of the Peace Science Society, October 8–10.

Burke, M., S. M. Hsiang, and E. Miguel (2015): "Climate and Conflict," *Annual Review of Economics*, 7, 577–617.

Burke, M. B., E. Miguel, S. Satyanath, J. A. Dykema, and D. B. Lobell (2009): "Warming Increases the Risk of Civil War in Africa," *Proceedings of the National Academy of Sciences*, 106, 20670–20674.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference With Multiway Clustering," *Journal of Business & Economic Statistics*, 29, 238–249.

Central Intelligence Agency (2018): "The World Factbook," The Central Intelligence Agency, Washington, DC. Data retrieved at https://www.cia.gov/library/publications/the-world-factbook/.

Cervellati, M., U. Sunde, and S. Valmori (2017): "Pathogens, Weather Shocks and Civil Conflicts," *Economic Journal*, 127, 2581–2616.

Chandler, T. (1987): *Four Thousand Years of Urban Growth: An Historical Census*, Lewiston, NY: The Edwin Mellen Press.

Cioffi-Revilla, C. (1996): "Origins and Evolution of War and Politics," *International Studies Quarterly*, 40, 1–22.

Collier, P. and A. Hoeffler (2004): "Greed and Grievance in Civil War," *Oxford Economic Papers*, 56, 563–595.

Conley, T. G. (1999): "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92, 1–45.

Croicu, M. and R. Sundberg (2015): "UCDP Georeferenced Event Dataset Codebook version 4.0," Department of Peace and Conflict Research, Uppsala University. http://ucdp.uu.se/downloads/ged/ucdp-ged-40-codebook.pdf.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012): "The Political Economy of Linguistic Cleavages," *Journal of Development Economics*, 97, 322–338.

Dincecco, M., J. Fenske, and M. G. Onorato (2015): "Is Africa Different? Historical Conflict and State Development," IMT Lucca EIC Working Paper No. 08/2015, IMT Institute for Advance Studies Lucca.

Esteban, J., L. Mayoral, and D. Ray (2012): "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, 102, 1310–1342.

Fearon, J. D. (2003): "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth*, 8, 195–222.

Fearon, J. D. and D. D. Laitin (2003): "Ethnicity, Insurgency, and Civil War," *American Political Science Review*, 97, 75–90.

Fenske, J. (2014): "Ecology, Trade and States in Pre-Colonial Africa," *Journal of the European Economic Association*, 12, 612–640.

Fincher, C. L. and R. Thornhill (2008): "Assortative Sociality, Limited Dispersal, Infectious Disease and the Genesis of the Global Pattern of Religion Diversity," *Proceedings of the Royal Society B: Biological Sciences*, 275, 2587–2594.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1999): "Geography and Economic Development," *International Regional Science Review*, 22, 179–232.

Galor, O. and Ö. Özak (2016): "The Agricultural Origins of Time Preference," *American Economic Review*, 106, 3064–3103.

Gething, P. W., A. P. Patil, D. L. Smith, C. A. Guerra, I. R. Elyazar, G. L. Johnston, A. J. Tatem, and S. I. Hay (2011): "A New World Malaria Map: *Plasmodium falciparum* Endemicity in 2010," *Malaria Journal*, 10.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002): "Armed Conflict 1946-2001: A New Dataset," *Journal of Peace Research*, 39, 615–637.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister (2014): "Updated High-Resolution Grids of Monthly Climatic Observations – The CRU TS3.10 Dataset," *International Journal of Climatology*, 34, 623–642.

Harris, I. C. and P. D. Jones (2013): "CRU TS3.21: Climatic Research Unit (CRU) Time-Series (TS) Version 3.21 of High Resolution Gridded Data of Month-by-Month Variation in Climate (Jan. 1901 - Dec. 2012) [Data file]," University of East Anglia Climatic Research Unit, NCAS British Atmospheric Data Centre, 24 September 2013. doi:10.5285/D0E1585D-3417-485F-87AE-4FCECF10A992.

——— (2017): "CRU CY4.01: Climatic Research Unit (CRU) Year-by-Year Variation of Selected Climate Variables by Country (CY) version 4.01 (Jan. 1901 - Dec. 2016) [Data file]," University of East Anglia Climatic Research Unit, Centre for Environmental Data Analysis, 4 December 2017. doi:10.5285/d4e823f0172947c5ae6e6b265656c273.

Hastings, D. A., P. K. Dunbar, G. M. Elphingstone, M. Bootz, H. Murakami, H. Maruyama, H. Masaharu, P. Holland, J. Payne, N. A. Bryant, et al. (1999): "The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0," National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO. Data retrieved at https://www.ngdc.noaa.gov/mgg/topo/globe.html.

Hegre, H. and N. Sambanis (2006): "Sensitivity Analysis of Empirical Results on Civil War Onset," *Journal of Conflict Resolution*, 50, 508–535.

Heizer, R. F. (1978): *Handbook of North American Indians, Vol. 8: California*, Washington, DC: Smithsonian Institution.

Hsiang, S. M., M. Burke, and E. Miguel (2013): "Quantifying the Influence of Climate on Human Conflict," *Science*, 341, 1235367/1–14.

King, G. and L. Zeng (2001): "Logistic Regression in Rare Events Data," *Political Analysis*, 9, 137–163.

Klein Goldewijk, K., A. Beusen, and P. Janssen (2010): "Long-Term Dynamic Modeling of Global Population and Built-Up Area in a Spatially Explicit Way: HYDE 3.1," *The Holocene*, 20, 565–573.

Klein Goldewijk, K., A. Beusen, G. van Drecht, and M. de Vos (2011): "The HYDE 3.1 Spatially Explicit Database of Human-Induced Global Land-Use Change Over the Past 12,000 Years," *Global Ecology and Biogeography*, 20, 73–86.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny (1999): "The Quality of Government," *Journal of Law, Economics, and Organization*, 15, 222–279.

Lujala, P., J. Ketil Rod, and N. Thieme (2007): "Fighting over Oil: Introducing a New Dataset," *Conflict Management and Peace Science*, 24, 239–256.

Marshall, M. G. (1999): *Third World War*, Lanham, MD: Rowman & Littlefield Publishers.

——— (2002): "Measuring the Societal Impact of War," in *From Reaction to Conflict Prevention: Opportunities for the UN System*, ed. by F. O. Hampson and D. M. Malone, Boulder, CO: Lynne Rienner, 63–105.

——— (2017): "Major Episodes of Political Violence (MEPV) and Conflict Regions, 1946–2017," Center for Systemic Peace, Vienna, VA. Data retrieved at http://www.systemicpeace.org/inscrdata.html.

Marshall, M. G., T. R. Gurr, and K. Jaggers (2017): "Polity IV Project: Political Regime Characteristics and Transitions, 1800–2017," Center for Systemic Peace, Vienna, VA. Data retrieved at http://www.systemicpeace.org/inscrdata.html.

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

Modelski, G. (2003): *World Cities: -3000 to 2000*, Washington, DC: FAROS 2000.

Montalvo, J. G. and M. Reynal-Querol (2005): "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 95, 796–816.

Murdock, G. P. (1959): *Africa: Its Peoples and Their Culture History*, New York, NY: McGraw-Hill Book Co., Inc.

Nei, M. (1973): "Analysis of Gene Diversity in Subdivided Populations," *Proceedings of the National Academy of Sciences*, 70, 3321–3323.

New, M., D. Lister, M. Hulme, and I. Makin (2002): "A High-Resolution Data Set of Surface Climate Over Global Land Areas," *Climate Research*, 21, 1–25.

Nordhaus, W. D. (2006): "Geography and Macroeconomics: New Data and New Findings," *Proceedings of the National Academy of Sciences*, 103, 3510–3517.

Nunn, N. (2008): "The Long-term Effects of Africa's Slave Trades," *Quarterly Journal of Economics*, 123, 139–176.

Nunn, N. and D. Puga (2012): "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 94, 20–36.

Nunn, N. and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101, 3221–3252.

Özak, Ö. (2010): "The Voyage of Homo-Economicus: Some Economic Measures of Distance," Unpublished manuscript. Department of Economics, Southern Methodist University.

Pemberton, T. J., M. DeGiorgio, and N. A. Rosenberg (2013): "Population Structure in a Comprehensive Genomic Data Set on Human Microsatellite Variation," *G3: Genes, Genomes, and Genetics*, 3, 891–907.

Pettersson, T. and K. Eck (2018): "Organized Violence, 1989–2017," *Journal of Peace Research*, 55, 535–547.

Putterman, L. (2008): "Agriculture, Diffusion, and Development: Ripple Effects of the Neolithic Revolution," *Economica*, 75, 729–748.

Putterman, L. and D. N. Weil (2010): "Post-1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality," *Quarterly Journal of Economics*, 125, 1627–1682.

Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman, and L. L. Cavalli-Sforza (2005): "Support from the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa," *Proceedings of the National Academy of Sciences*, 102, 15942–15947.

Ramankutty, N., J. A. Foley, J. Norman, and K. McSweeney (2002): "The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change," *Global Ecology and Biogeography*, 11, 377–392.

Reynal-Querol, M. (2002): "Ethnicity, Political Systems, and Civil Wars," *Journal of Conflict Resolution*, 46, 29–54.

Riley, S. J., S. D. DeGloria, and R. Elliot (1999): "A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity," *Intermountain Journal of Sciences*, 5, 23–27.

Smith, T. W., M. Davern, J. Freese, and S. L. Morgan (2018): "General Social Surveys, 1972–2018 [Data file]," National Data Program for the Social Sciences, Chicago, IL. Data retrieved at gss.norc.org.

Sundberg, R., K. Eck, and J. Kreutz (2012): "Introducing the UCDP Non-State Conflict Dataset," *Journal of Peace Research*, 49, 351–362.

Tollefsen, A. F., H. Strand, and H. Buhaug (2012): "PRIO-GRID: A Unified Spatial Data Structure," *Journal of Peace Research*, 49, 363–374.

Weidmann, N. B., J. K. Rød, and L.-E. Cederman (2010): "Representing Ethnic Groups in Space: A New Dataset," *Journal of Peace Research*, 47, 491–499.

World Bank (2018): "World Development Indicators," The World Bank, Washington, DC. Data retrieved at https://datacatalog.worldbank.org/dataset/world-development-indicators.

World Values Survey (2006): "European and World Values Surveys, Four-Wave Integrated Data File, 1981–2004, version 20060423," The World Values Survey Association, Stockholm, Sweden. Data retrieved at http://www.worldvaluessurvey.org.

——— (2009): "World Values Survey, 1981–2008 Official Aggregate, version 20090914," The World Values Survey Association, Stockholm, Sweden. Data retrieved at http://www.worldvaluessurvey.org.