NBER WORKING PAPER SERIES

DIVERSITY AND CONFLICT

Cemal Eren Arbatli
Quamrul H. Ashraf
Oded Galor
Marc Klemp

Working Paper 21079
http://www.nber.org/papers/w21079

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2015, Revised November 2018

Previously circulated as "The Nature of Conflict." 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.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). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Diversity and Conflict
Cemal Eren Arbatli, Quamrul H. Ashraf, Oded Galor, and Marc Klemp
NBER Working Paper No. 21079
April 2015
JEL No. D74,N30,N40,O11,O43,Z13

ABSTRACT

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

Cemal Eren Arbatli
Faculty of Economic Sciences
National Research University
Higher School of Economics
26 Shabolovka St. Building 3
3116A Moscow, Russia
earbatli@gmail.com

Quamrul H. Ashraf
Williams College
Department of Economics
24 Hopkins Hall Drive
Williamstown, MA 01267
Quamrul.H.Ashraf@williams.edu

Oded Galor
Department of Economics
Brown University
Box B
Providence, RI 02912
and CEPR
and also NBER
oded_galor@brown.edu

Marc Klemp
Økonomisk Institut
University of Copenhagen
Øster Farimagsgade 5, bygning 26
1353 København K
and Brown University
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 horizon.

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 that interpersonal diversity is pivotal for the understanding of civil conflict, as illustrated in Figure 1. Exploiting exogenous variations in population diversity across nations and ethnic groups, the study establishes that interpersonal population diversity, as determined predominantly during the exodus of humans from Africa tens of thousands of years ago, has contributed significantly to conflicts in the course of human history. Furthermore, the study suggests that the adverse effect of interpersonal population diversity on interpersonal trust and cooperation, its contribution to divergence in preferences for public goods and redistributive policies, and its impact on the degree of fractionalization and polarization across ethnic, linguistic, and religious groups have fostered social, political, and economic disorder and have, thus, 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 cooperation, 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 prevented the culmination of such grievances to 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 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.

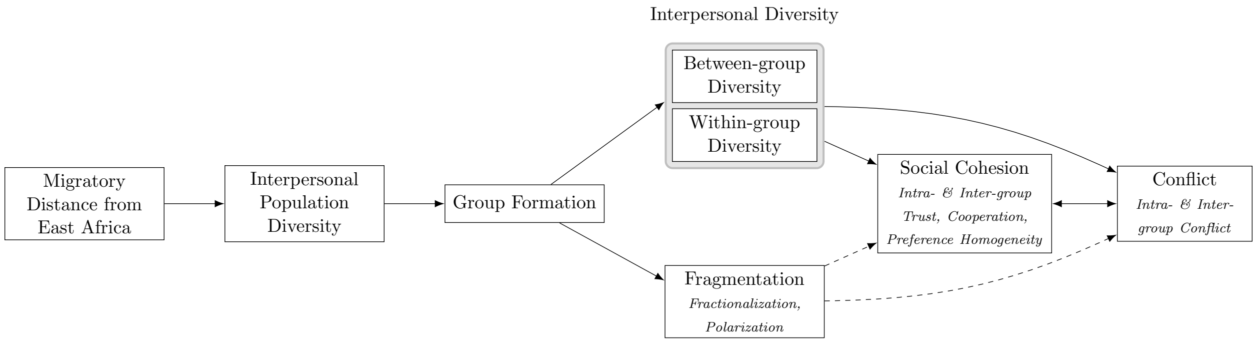


Figure 1: The evolution of population diversity in a given location and its impact on conflict.

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.

The exploration of the contribution of interpersonal population diversity to conflict within nations and ethnic groups relies on a measure that encompasses various dimensions of population diversity (i.e., proportional representation of ethnic groups, interpersonal diversity between groups, and interpersonal diversity within groups). While the level of population diversity at the national level can be partly captured by indexes of ethnolinguistic fractionalization and polarization, these measures predominantly reflect the proportional representation of ethnic groups in the population (and rarely pairwise distances amongst ethnic groups), abstracting from the importance of the degree of interpersonal diversity *within* each ethnic group to the overall level of diversity in the national population. These measures therefore cannot capture the role of population diversity within an ethnic group on intra-group conflict. In contrast, population diversity within ethnic groups, as determined in the course of the exodus of modern humans from Africa, can shed light on the effect of interpersonal population diversity on intra-group and inter-group conflicts. This measure captures the probability that two individuals, selected at random from a given population differ from one another in a certain spectrum of their genome. Moreover, the measures of population diversity for multi-ethnic populations, as constructed by Ashraf and Galor (2013a), permits the exploration various dimensions of population diversity on conflict at the country level.

Exploiting variations across countries and ethnic homelands, the analysis demonstrates that interpersonal population diversity within and between ethnic groups, rather than ethnolinguistic fragmentation, has contributed fundamentally 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 via the number of ethnic groups in the population, the prevalence of generalized interpersonal trust, 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 boarders (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 location, at least since the precolonial era, 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 exploits 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 migratory distance from East Africa to the location of each ethnic group to predict population diversity for a globally representative sample of more than 1,200 ethnic groups, reflecting that observed population diversity of an indigenous contemporary ethnic group decreases with distance along ancient migratory paths from East Africa due to the serial founder effect (e.g., Harpending and Rogers, 2000; Ramachandran et al., 2005; Ashraf and Galor, 2013a).[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 necessitated, 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

[1] The contemporary worldwide distribution of genetic diversity across prehistorically 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 genetic diversity of their parental colony, the genetic diversity of a prehistorically indigenous ethnic group as observed today decreases with the distance along ancient human migratory paths from East Africa.

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 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 the population diversity of an ethnic group from the 10th percentile (e.g., the Huave people of Mesoamerica) to the 90th percentile (e.g., the Nuer people of Central Africa) of its global distribution is associated with an increase in the spatiotemporal prevalence of conflict by 1.2 percentage points *per year* and *per unit area* of the *ethnic homeland* during the 1989–2008 time period (relative to a sample mean of 0.13 and a standard deviation of 0.25). Further, this change in ethnic population diversity is also associated with an increase of about 260 in the total number of conflict events and an increase of over 6,390 in the total number of deaths across all conflict events during the same time horizon.

2 Related Literature

This study is related to several well-established lines of inquiry. First, the paper contributes to the vast literature on the determinants of civil conflict. The origins of civil conflict have been the focus of intensive research over the past two decades highlighting the role of social, political, and economic grievances as determinants of the risk of civil conflict, in view of 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 (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, highlighting 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, since 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, as proxied by genetic 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, notwithstanding some notable exceptions (Fearon, 2003; Desmet et al., 2009; Esteban et al., 2012), the commonly used measures of ethnolinguistic fragmentation typically do not exploit

[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.

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 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, the fact that the endowment of population diversity in a given location was overwhelmingly determined during the prehistoric "out of Africa" expansion of humans permits the analysis 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, 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).

Second, the study 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, as proxied by genetic diversity, captures 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.

Third, 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

[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 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.

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). In particular, since the initial endowment of interpersonal population diversity at a 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.[6]

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 effect of population diversity on conflict, the key variables used, and the underlying strategies in identifying the effect of population diversity on conflict.

The analysis focuses initially on contemporary conflicts, exploiting variations in either cross-country or repeated cross-country data. First, it explores, the explanatory power of interpersonal population diversity for observed variations in three different dimensions of civil conflicts: (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. Second, it examines the impact of interpersonal population diversity on the intensive margin of conflict, exploiting cross-country variations over time in the *severity* of social unrest, as reflected by two alternative measures of the intensity of intrastate conflict. Third, it analyzes the impact of interpersonal diversity on *intra-group* factional conflicts within a national population. Fourth 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 gained independence. This time horizon permits an assessment of the correlates of civil conflict at the national level, independently of their interactions with the influence of the colonial powers on the potential for internal conflict. The baseline sample for contemporary analysis of conflict contains information on 150 countries for the period 1960–2017, of which 123 are in the Old World.

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

The main dependent variable at the country level is the occurrence of civil conflict (i.e., an internal armed conflict between the government of a state and internal opposition groups), as reported in the most comprehensive conflict coding - PRIO25 - which encompasses all conflict events that resulted

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

[6]The modernist viewpoint (Bates, 1983; Gellner, 1983; Wimmer, 2002) stress that interethnic conflict arises from increased competition over scarce resources, especially when previously marginalized groups that were excluded from the nation-building process experience socioeconomic modernization and, thus, begin to challenge the status quo.

in at least 25 battle-related deaths in a given year.[7] In particular, the onset of conflicts, over the sample period, reflects the number of new conflicts between state actors and armed opposition groups that have exceeded the battle-related death threshold. Furthermore, the analysis exploits the temporal dimension of civil conflict to conduct a repeated cross section analysis of the incidence (prevalence) of conflicts, where the outcome variable is an indicator, coded 1 for each country-period (a period typically being a 5-year time interval) in which there is at least one active conflict is observed, and 0 otherwise. In addition, an annually repeated cross-country data is employed to examine the predictive power of population diversity for the *onset* of a new conflict episode.[8]

3.1.2 Population Diversity: Measurement and Identification Strategy

Interpersonal population diversity in each country is captured by the measure of predicted genetic diversity developed by (Ashraf and Galor, 2013a). It is derived based on the proportional representation of each ancestral population in contemporary nations, the genetic diversity of each of these ancestral populations, as predicted by their migratory distance from Africa, and the pairwise genetic distances between each pair of these ancestral population, 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 genetically 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).[9] Expected heterozygosity, H_{exp}, is taking 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.

In particular, population geneticists constructs a measure of expected heterozygosity as well as pairwise genetic distances, in a sample of 53 globally representative ethnic groups from the Human Genome Diversity Cell Line Panel.[10] These ethnic groups have been prehistorically native to their current geographical locations, and 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. 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. Moreover, observed socioeconomic influence of expected heterozygosity in microsattelites reflects the mounting evidence from the fields of physical and cognitive anthropology that establishes serial founder effects (associated with the prehistoric "out of Africa" migration process) on worldwide spatial patterns

[7] See the UCDP/PRIO Armed Conflict Dataset, Version 4-2012 (Gleditsch et al., 2002; Themnér and Wallensteen, 2012).

[8] The onset year of a PRIO25 civil conflict corresponds to either the first year of a new PRIO25 conflict or the first year of a recurring conflict after more than one year of inactivity.

[9] See Ashraf and Galor (2018).

[10] 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).

in various forms of intra-group phenotypic and cognitive diversity, including phonemic diversity as well as interpersonal diversity in skeletal features pertaining to cranial characteristics, dental attributes, and pelvic traits, as surveyed by Ashraf and Galor (2018).

Nevertheless, *observed* genetic diversity, like measures of ethnolinguistic fragmentation, based on fractionalization or polarization indices, 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 measure of predicted genetic diversity introduced to the literature by Ashraf and Galor (2013a). Exploiting the explanatory power of a serial founder effect associated with the out of Africa' migration process, population 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 a 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.

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 our entire analysis to the Old World, a measure of ancestry-adjusted genetic diversity from Ashraf and Galor (2013a) 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.[11]

Yet, 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. These migrations may have been spurred by historically persistent spatial patterns of conflict. Two alternative strategies address this issue. The first strategy iexploit 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.[12] This strategy exploits the view that the great human migrations of the post-

[11] 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).

[12] This is a pattern that primarily arises from the view that historical cross-country migrations in the Old World did not result in significant admixture of populations that are genetically distant from one another.

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, the study effectively exploit the spatial variation in contemporary population diversity that largely coincides with the variation in diversity of prehistorically indigenous populations. This prehistoric diversity was overwhelmingly determined 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 interpersonal 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 continues to be seen in the worldwide pattern of genetic diversity across contemporary national populations, as reflected by the sizable correlation of 0.750 between the proxies for precolonial and contemporary population diversity in the 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 is has no direct effect on civil conflict faced by its modern national population, conditional on our 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 the determinants of civil conflict has emphasized a large number of potentially contributing factors. Drawing on this literature, a large set of control variables are included in our baseline specifications. All other control variables used in robustness checks are discussed in corresponding Appendices.[13]

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. The analysis also account for total land area of a country. Larger territories may mechanically experience more conflict while also hosting a larger and more diverse national population.[14] 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). This in turn may increase the potential for inter-group conflict. Finally, 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

[13] The definition and data sources for all confounding characteristics are listed in the Supplemental Material.

[14] Land area can also account for any bias that might arise if the measure of population diversity, by virtue of being based on migratory distance from East Africa to the modern-day capital city of a country, is less comparable across countries of different geographical size.

propensity through the ethnolinguistic fragmentation of the population.[15] To account for these factors, the baseline analysis consists of terrain ruggedness, as well as the mean and range of both agricultural land suitability and elevation. 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.[16]

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 the potential for conflict in society.These two different sets of covariates are part of the baseline specifications, accounting 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 our 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.[17]

Ethnolinguistic Fragmentation Previous empirical findings regarding the role of ethnic fragmentation have generally 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 included in the baseline model 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

[15] 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 rinfluence on the potential for conflict in society.

[16] In addition to "soaking up" the possibility of omitted-variable bias from unobserved time-invariant characteristics at the continent level, the need to account for continent fixed effects is perhaps even more binding for observed non-geographical factors, given the potential for systematic measurement error at the continent level in covariates reflecting cultural and institutional characteristics.

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

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 we employ 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.[18]

Development Outcomes In the light of of the potential impact of the process of development on conflict, the baseline specifications accounts for development outcomes such as population size and GDP per capita, as reported by Maddison (2010), as well as per-capita value of oil production, as reported by Ross (2013). In particular, as established by Ashraf and Galor (2013a), interpersonal population diversity, as proxied by genetic diversity, confers a hump-shaped influence on productivity at the country level. Furthermore, scholars of civil conflict have linked various factors, codetermined with the level of development, such as natural resource revenues, the size of a country's population, and income per capita to conflict risk. For example, 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 vulnerability of political elite to terms-of-trade shocks (e.g., Humphreys, 2005) and raising the return to regional secession (e.g., Ross, 2006).

Population size is also a standard covariate in the empirical literature. 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 larger recruitment pool for rebels (Fearon and Laitin, 2003). Therefore to the extent more populous countries exhibit greater intrapopulation heterogeneity, they could also harbor stronger motives for secessionist conflicts (Alesina and Spolaore, 2003; Desmet et al., 2011).

Average living standards can influence civil conflict potential in a country through several channals. 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).

Yet, since 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 interpersonal diversity on conflict potential can be attributed to more conventional explanations in the literature.

[18] 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) and Esteban et al. (2012).

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

3.2 Empirical Results

This section presents our 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. We commence with the results of our baseline cross-country regressions that explain the annual frequency of civil conflict outbreaks in the post-1960 time period. We then discuss the results of our 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. We conclude this section 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. For our analysis of each conflict outcome, we conduct 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

Our 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 our 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 1 presents the results from our baseline cross-country analysis. We start 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, we progressively include an expanding set of covariates to the specification. We first incorporate exogenous geographical characteristics and then additionally account for measures of ethnolinguistic fragmentation, before controlling for semi-endogenous institutional factors and more endogenous outcomes of economic development in our 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,

Table 1: 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)
Small 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)
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)
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	0.044	0.040		0.051	0.068	0.088		
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)											−0.007***	−0.006***
											(0.001)	(0.001)
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.

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.[19] Based on our 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 our 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 we condition the specification to only exploit intra-continental cross-country variations. However, even after including a complete set of continent dummies, our 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.[20] 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 our baseline set of geographical covariates.[21]

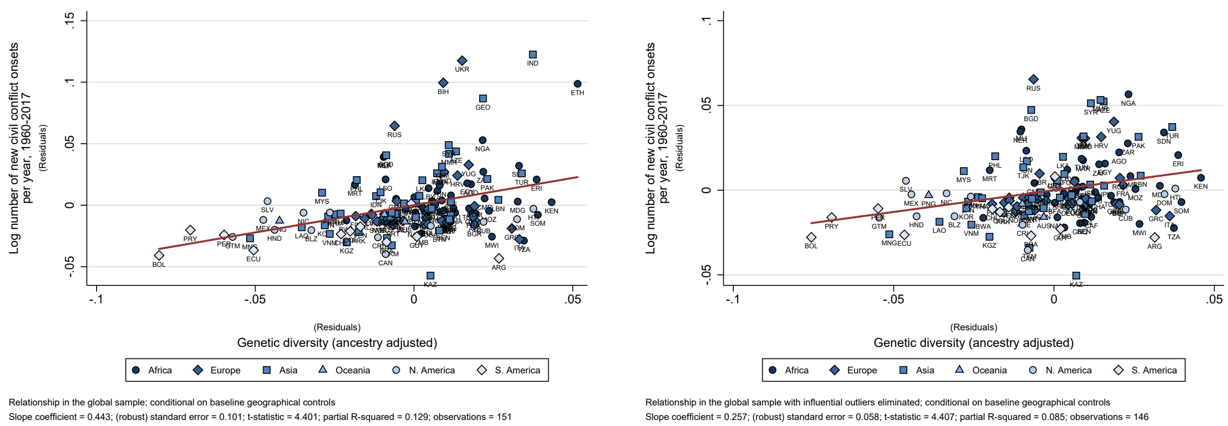
In Column 7, we replicate the specification from Column 6 except that now, we decompose our measure of overall interpersonal diversity of the national population into its two components and jointly examine their conditional associations with conflict. The two components of overall diversity capture the 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.[22] 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.

[19] 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).

[20] By restricting both fractionalization and polarization measures to enter our regressions linearly, our baseline approach follows Esteban et al. (2012), but we nevertheless checked the robustness of our main finding to employing alternative specifications that allow for both a linear and a quadratic term in ethnic fractionalization, and we found qualitatively similar results (not reported).

[21] The analysis in Table SA.9 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.

[22] 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 1. 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 1, the regression coefficients reported in this panel are marginally different from those presented in Column 2 of Table 1. 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).

Our 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 our 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 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 our 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 Our 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 our 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 our covariates. Although the stability of our coefficient of interest across specifications suggests that selection on unobservables needs to be unreasonably strong to fully explain away our results, we cannot rely entirely on OLS point estimates to assess causality.[23] Thus, as discussed previously in Section 3.1, we exploit two alternative identification strategies to address this issue. In Columns 9–10, we implement our 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, we run 2SLS regressions that employ the migratory distance of the prehistorically native population in each country from East Africa as an instrument for the country's contemporary population diversity. Our 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 our rich set 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 our global-sample OLS regressions (from either Column 3 or Column 8). In particular, our coefficient of interest 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 global the cross-country conflict frequency distribution.

In our view, there are three distinct rationales – perhaps operating in tandem – for why our 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, our ancestry-adjusted measure of interpersonal diversity – which incorporates the diversity of both native and non-native

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

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 our 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, our better-identified estimates mitigate all the aforementioned sources of negative bias.

Robustness Checks In Appendix A.2, we establish 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.2). We also show the robustness of baseline cross-country findings to accounting for: spatial dependence across observations by estimating spatial regressions (Table A.3); and the property of our measure of population diversity as a generated regressor by bootstrapping our standard errors (Table A.4).

Further, in Section A.1 of the Supplemental Material, we present several robustness checks for our cross-country analysis of the influence of population diversity on the temporal frequency of civil conflict outbreaks in the post-1960 time horizon. We demonstrate that our 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.1); **(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.2); **(3)** accounting for inequality across ethnic homelands as well as overall spatial inequality in nighttime luminosity within a country (Table SA.3); **(4)** accounting for linguistic rather than ethnic fractionalization as a baseline covariate (Table SA.4); **(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.5); **(6)** accounting for the initial-year values of time-varying baseline covariates rather than their temporal means over the sample period (Table SA.6); **(7)** accounting for spatial autocorrelation in unobserved heterogeneity (Table SA.7); and **(8)** the elimination of world regions from the estimation sample that could, a priori, be considered as statistically influential for generating the key empirical pattern (Table SA.8).

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

The second dimension of civil conflict that we examine is its temporal prevalence. Specifically, exploiting the time structure of quinquennially repeated cross-country data, we investigate 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. We therefore estimate the following probit model using maximum-likelihood estimation.

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

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

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

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 $\widehat{GD}_i$, 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.[24] By specifying each of our time-varying controls to enter the model with a one-period lag, we aim to mitigate the concern that the use of contemporaneous measures of these covariates may exacerbate reverse-causality bias in their estimated coefficients.[25] Finally, we assume 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.[26] 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, $\eta_{i,t}$.

We present our main results for the temporal prevalence (or incidence) of PRIO25 civil conflict episodes in Columns 1–4 of Table 2. In the interest of brevity, we only report our 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.[27] For each of these two identification strategies, we estimate two distinct specifications; 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 our 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 our coefficient of interest, the IV probit regression presented in Column 4 suggests that conditional on our 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,

[24] In Appendix A.2, we confirm the robustness of our analysis of conflict incidence to exploiting variations in annually (rather than quinquennially) repeated cross-country data. 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.

[25] 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.

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

[27] The results of probit regressions in the global sample of countries are available upon request.

Table 2: 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***	6.167**	6.380**	7.057***	8.811***
	(3.700)	(3.787)	(3.652)	(4.210)	(2.573)	(2.637)	(2.592)	(3.167)
Ethnic fractionalization		−0.399		−0.519		−0.087		−0.324
		(0.353)		(0.332)		(0.252)		(0.280)
Ethnolinguistic polarization		0.049		0.322		0.173		0.335
		(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***	0.323**	0.333**	0.335**	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)			−0.007***	−0.007***			−0.007***	−0.007***
			(0.001)	(0.001)			(0.001)	(0.001)
First-stage F-statistic			145.394	99.877			151.512	102.677

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.

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.

Robustness Checks In Appendix A.2, we show 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.5); exploiting variations in annually rather than quinquennially repeated cross-country data (Table A.6, 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.6, Columns 5–8).

Further, in Section A.2 of the Supplemental Material, we establish that our 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.10, 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.11, Columns 1–4); **(3)** accounting for inequality in nighttime luminosity across gridded space and across ethnic homelands within a country (Table SA.12, 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.13); **(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.14, Columns 1–4); and **(6)** allowing for spatiotemporal dependence across country-time observations by exploiting two-dimensional clustering of standard errors (Table SA.15, Columns 1–4).

Finally, akin to our 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

The third dimension of conflict examined by our analysis is 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.[28] 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 our 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, we now (i) exploit variations in annually repeated cross-country data, with our binary outcome variable assuming a value of 1 if a country-year observation coincides with

[28] 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.

the first year of a new civil conflict, and 0 otherwise; and (ii) control 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, we implement the same two strategies followed by our analyses of conflict frequency and conflict incidence.

Our main results for the onset of new PRIO25 civil conflict episodes are presented in Columns 5–8 of Table 2. 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 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 our 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 2 to 3.5 percent in the Old-World sample of countries, and from about 1 to 4 percent amongst countries worldwide.

Robustness Checks In Section A.2 of the Supplemental Material, we demonstrate that our 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.10, 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.11, Columns 5–8); **(3)** accounting for inequality in nighttime luminosity across gridded space and across ethnic homelands within a country (Table SA.12, 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.14, Columns 5–8); **(5)** allowing for spatiotemporal dependence across country-year observations by exploiting two-dimensional clustering of standard errors (Table SA.15, 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.16); 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.17).

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

One crucial dimension in which our measure of population diversity adds value beyond standard indices of ethnolinguistic fragmentation is that our index 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, our 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 our 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 our 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 dataset, combining information on 291 subnational groups originally included 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).[29] For each subnational group in the AMAR Sample Data, we observe whether the group experienced one or more intra-group conflicts in each year during the 1985–2006 time horizon.

Table 3 presents our analyses of intra-group conflict incidence. The outcome variable for our 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 our analysis based on annually repeated cross-country data (Panel B), we code 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 our 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), we employ 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 our two identification strategies, we present the results from estimating three alternative specifications. The first two of these follow from our 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 our 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 our 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 our methodology from earlier sections, our 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.

Turning to our findings, the results in Table 3 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

[29] 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 3: 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 group-years with intra-group conflict, 1985-2006					
Population diversity (ancestry adjusted)	4.655***	4.267**	3.580**	5.861***	5.606***	5.124***
	(1.659)	(1.711)	(1.694)	(1.705)	(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	94	91	91	118	115	115
Adjusted R^2	0.078	0.187	0.231			
Effect of 10th–90th %ile move in diversity	0.237***	0.209**	0.175**	0.402***	0.384***	0.351***
	(0.085)	(0.084)	(0.083)	(0.117)	(0.129)	(0.130)
FIRST STAGE				Population diversity (ancestry adjusted)		
Migratory distance from East Africa (in 10,000 km)				−0.006***	−0.006***	−0.006***
				(0.001)	(0.001)	(0.001)
First-stage F-statistic				94.727	47.887	47.107
PANEL B	Probit	Probit	Probit	IV Probit	IV Probit	IV Probit
	Annual intra-group conflict incidence, 1985–2006					
Population diversity (ancestry adjusted)	31.708***	37.535***	31.687***	35.942***	40.579***	38.375***
	(9.027)	(9.792)	(10.547)	(6.757)	(7.261)	(7.973)
Controls as in same column of Panel A	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×
Observations	1,905	1,658	1,658	2,433	2,179	2,179
Countries	93	90	90	117	114	114
Pseudo R^2	0.201	0.338	0.390			
Marginal effect of diversity	9.573***	9.107***	7.067***	10.109***	10.318***	9.402***
	(2.377)	(2.301)	(2.428)	(1.777)	(2.008)	(2.212)
FIRST STAGE				Population diversity (ancestry adjusted)		
Migratory distance from East Africa (in 10,000 km)				−0.006***	−0.006***	−0.006***
				(0.001)	(0.001)	(0.001)
First-stage F-statistic				97.016	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.

significance as well. For instance, the regression in Column 5 of Panel A suggests that conditional on our 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-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 53 percent amongst countries in the Old World, and from about 6 to 60 percent in the global sample of countries.

3.2.5 Analysis of Historical Conflict Outcomes in Cross-Country Data

Thus far, our analysis has been confined to examining intrastate conflict events in the last half-century. This restriction permitted us 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, a priori, there is no particular reason why the conflict-promoting role of population diversity should not extend to the distant past.

In this section, we investigate whether population diversity predicts historical conflict events in a cross-section of countries. Specifically, we exploit 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). We employ the geocoding of conflict locations to map these historical conflicts to territories, as defined by their contemporary national borders. The time period we examine 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, we do not expect them to be related to local population diversity in a meaningful way.

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.[30] Further, although the dataset 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 our analysis of modern conflicts, our variable of interest in the current analysis is the pre-colonial 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

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

Table 4: 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	12.809***	9.911***	6.662**	7.498***	6.819***	16.193***	32.531***	10.705**	11.949**	11.451**
	(3.590)	(2.102)	(2.819)	(2.685)	(2.598)	(4.718)	(5.728)	(4.934)	(4.740)	(4.832)
Region dummies	×	×	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×	×	×
Observations	155	155	155	155	155	155	155	155	155	155
Partial R^2 of population diversity	0.091	0.103	0.046	0.052	0.060					
Adjusted R^2	0.347	0.348	0.330	0.246	0.236					
Pseudo R^2						0.245	0.370	0.262	0.206	0.212
Effect of 10th–90th %ile move in diversity	24.875***	6.104***	4.638**	4.487***	2.985***	0.504***	0.621***	0.383***	0.455***	0.406***
	(6.972)	(1.295)	(1.963)	(1.607)	(1.137)	(0.097)	(0.045)	(0.137)	(0.126)	(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 three indicators for Sub-Saharan Africa, Middle East and North Africa, and Europe and Central 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.

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 our measure of pre-colonial 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, our measure should possess some explanatory power for the onset of such conflicts. Because we examine conflicts during a time period when long-distance campaigns were uncommon, due to the constraints imposed by historical transportation and warfare technologies, our measure 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.

Table 4 presents our 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.[31] In Column 1, we examine conflicts during the entire time horizon of four centuries, whereas the specifications in Columns 2–5 focus on the conflicts recorded for each individual century of the time horizon. Indeed, we expect the data on conflicts that occurred prior to the discovery of the New World 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 our 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 our analysis of modern conflicts. In addition, we include regional dummies in all regressions to mitigate the concern that Brecke's conflict data could suffer from a regional bias in coverage, due

[31]For each observation, we add one to the total number of conflicts before performing the log transformation, in order to retain observations without any recorded conflict.

to differences across world regions in the quality of primary sources and in the nature and scale of historical conflict events.[32]

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

To interpret the impact of population diversity on historical conflicts, 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 6.1 more conflicts during the 15th century. This impact is somewhat larger than those implied by our comparable specifications for modern civil conflicts.[33] This could potentially reflect a waning, albeit significant, influence of population diversity as we move closer to more contemporary time periods. However, it could also be a mechanical consequence of measurement issues associated with the fact that in contrast to our analysis of modern civil conflicts, our 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 62 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 our earlier analyses of civil conflict in the contemporary era, our 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 origins of the prevailing variation in the prevalence and severity of intra-societal conflicts within ethnic groups. This ethnic level analysis mitigates potential concerns regarding the endogeneity of contemporary national borders to population diversity and the intensity of conflict. Moreover, since populations within ethnic homelands have been largely native to these locations at least since the pre-colonial era, the ethnic level analysis further circumvents potential concerns about the effect of conflict on migrations across countries and its potential impact on population diversity. In addition, the exploration of the effect of interpersonal population diversity on conflict within ethnic homelands permits the analysis to disentangle the impact of population diversity within an ethnic group, from the impact of ethnic diversity across groups. Furthermore, the analysis establishes that the effect of population diversity on the risk of conflict is independent of the population scale.

[32] 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.

[33] For instance, in Column 3 of Table 1, 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.

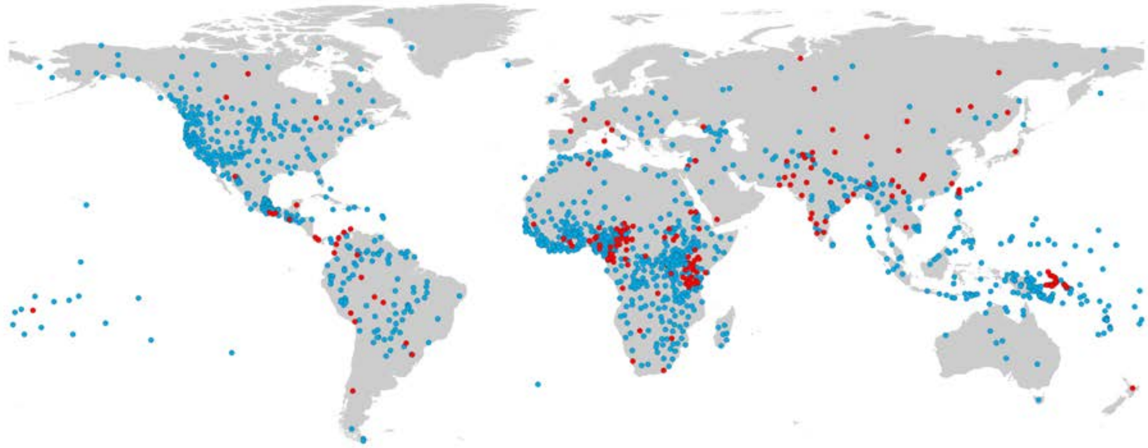


FIGURE 3: The Spatial Distribution of Ethnic Homelands

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 groups for which population diversity is observed, whereas blue points depict groups for which population diversity is predicted.

4.1 Data

The ethnic level analysis is conducted, based on a novel geo-referenced dataset consisting of ethnic groups, for which genetic diversity is either observed, or can be predicted. In particular, the geo-referenced dataset exploits the recently assembled data on observed genetic diversity within indigenous ethnic groups across the globe that have been largely isolated and shielded from genetic admixture (Pemberton et al., 2013).[34] The distribution of these ethnic groups across the globe is depicted in Figure 3 and the summary statistics of this measure of genetic diversity, as documented in Table **??**, establishes that observed diversity ranges from 0.58 among ethnic groups in South America to 0.77 among those in Africa. The geo-referenced dataset maps the genetic diversity of each ethnic group, as reported in Pemberton et al. (2013), to the geographical characteristics of its ethnic homeland.[35] The data consists of 230 ethnic groups for which genetic diversity is observed.[36] In addition, using the methodology of predicting genetic diversity based on migratory distance from Africa, the ethnic-homeland geo-referenced dataset is further augmented, beyond the set of ethnic groups covered by Pemberton et al. (2013), and it encompasses the entire set of ethnic groups in the *Ethnographic Atlas*.

It should be noted that the territory of the historical homeland of each ethnic group captures the area of the globe in which each ethnic group is predominantly residing. Nevertheless, only 277 of 1238 ethnic homelands in the sample appear linguistically homogeneous, whereas the remaining 962 ethnic homelands are inhabited by multiple linguistic groups. Hence, the analysis of the impact

[34]This dataset 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).

[35]Further details on the construction of the data set are presented in Section B.2 of the Supplemental Material.

[36]The analysis includes all observations on ethnic groups in Pemberton et al. (2013), excluding two ethnicities (the Surui and the Ache of South America) that are largely viewed by population geneticists as extreme outliers in terms of genetic diversity (e.g. Wang et al., 2007). In particular, Ramachandran et al. (2005) omit the Surui, as "an extreme outlier in a variety of previous analyses", and do not include the Ache either. Furthermore, these ethnicities have the lowest levels of genetic diversity in the sample and the largest residuals of an OLS regression of genetic diversity on migratory distance from Addis Ababa. Including these observations, nevertheless, does not affect the qualitative analysis.

of interpersonal population diversity of conflict is conducted in two layers. First, it is restricted to the sample of 277 linguistically homogenous homelands, for which conflicts arguably capture intra-group conflict. Second, it employs the entire sample, where conflicts may partly reflect the degree of fragmentation of these ethnic homelands, while accounting for the potentially confounding effects of the degree of linguistic fractionalization and polarization within non-homogenous homelands.[37]

The main measure of conflicts at the ethnic-group level, in line with the country-level analysis, is based on 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 that was within the boundaries of internal armed conflict events occurring between the government of a state and internal opposition groups over the period 1989–2008. This variable is denoted as the *spatio-temporal prevalence of UCDP/PRIO conflicts*.[38] Furthermore, a secondary measure is based on the number of events recorded within each ethnic homeland in the UCDP geo-referenced Event Dataset (Sundberg et al., 2012; Croicu and Sundberg, 2015).

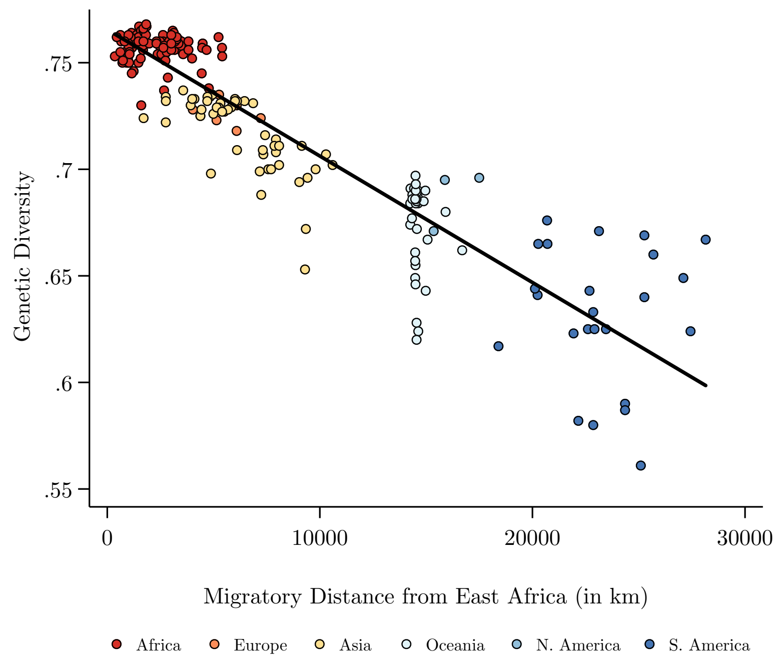
4.2 Empirical Strategy

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 groups. In particular, the positive associations between the extent of the observed population diversity within an ethnic group 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 a 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 exploits *predicted* population diversity rather than observed diversity to explore the effect of diversity of civil conflict. In particular, as established 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 *Ethnographic Atlas*.

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

[37]As elaborated in Section B.1 of the Supplemental Material, the measures of the degree of ethnolinguistic fractionalization and polarization in non-homogeneous ethnic homelands is constructed based on the population that resides in each of the intersections between each of the linguistic homelands and the ethnic homeland.

[38]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.



Relationship in the global sample

Slope coefficient = -0.000; (robust) standard error = 0.000; t-statistic = -35.500; partial R-squared = 0.847; observations = 230

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

Notes: This figure depicts the relationship between prehistoric migratory distance from East Africa and observed population diversity in a sample of 230 ethnic groups based on Pemberton et al. (2013). The negative relationship reflects the serial founder effect associated with expansion of humans from East Africa to the rest of the world.

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, a highly implausible 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 necessitated 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 migration process (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 roots. Nevertheless, in order to further mitigate this highly implausible 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 a geographical region 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 implausible 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 necessitated 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, and diminishing symmetrically along these diverging migratory roots. Nevertheless, to further mitigate this highly implausible 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 group on intra-societal conflicts within the ethnic homeland. The analyzes explores the effect of population diversity within ethnic groups on the prevalence of conflict, as well as on the extensive and the intensive margins of conflicts 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 5 and 6 presents the results of the baseline analysis of the influence of population diversity within an ethnic group on the log spatio-temporal prevalence of UCDP/PRIO conflicts during 1989–2008. Table 8 conducts the analysis for the observed-diversity sample. In particular, column 1 establishes a highly significant relation between observed diversity across the 230 ethnic groups and the conflict measure, conditional on the world-region fixed effects. Column 2, demonstrates that — as depicted in Figure 5 — the association 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 and 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., mean and standard deviation of temperature and precipitation), the association between observed diversity and conflict remains highly significant. The coefficient estimate suggests that an increase in the population diversity from the 10th percentile of the observed level of diversity (e.g., the Guaraní people of South America) to the 90th percentile (e.g., the Bulu people of South Africa) corresponds to an increase in the prevalence of spatio-temporal conflict by 0.38 percentage points (compared to a sample mean of 0.13 and a standard deviation of 0.25).[39] Columns 4 and 5 establishes that the qualitative results are unaffected by accounting for the potentially confounding effects of linguistic fractionalization and

[39]See Appendix A.6.

Table 5: Population Diversity and the Conflict across Ethnic Homelands – Observed Diversity

	Log spatio-temporal prevalence of UCDP GED conflicts, 1989–2008						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS
Genetic diversity (observed)	31.166***	28.348***	25.675***	24.975**	25.838***	29.874***	30.698***
	[9.185]	[9.374]	[9.773]	[9.713]	[9.772]	[10.776]	[10.839]
Linguistic fractionalization				0.471		0.197	
				[0.537]		[0.538]	
Linguistic polarization					-0.367		-0.468
					[0.480]		[0.502]
Regional dummies	×	×	×	×	×	×	×
Geographical controls		×	×	×	×	×	×
Climatic controls			×	×	×	×	×
Development outcomes						×	×
Disease environment controls						×	×
Sample	Observed	Observed	Observed	Observed	Observed	Observed	Observed
Observations	230	230	230	230	230	230	230
Effect of 10th90th %ile move in diversity	0.461***	0.419***	0.380***	0.369**	0.382***	0.442***	0.454***
meserrest	[0.136]	[0.139]	[0.145]	[0.144]	[0.145]	[0.159]	[0.160]
Adjusted R^2	0.159	0.327	0.434	0.434	0.433	0.450	0.452
β^*		26.660	22.997	21.965	23.245	29.272	30.481

Notes: This table exploits cross-ethnicity variations to establish a significant positive reduced-form impact of contemporary population diversity on the log spatio-temporal prevalence of UCDP/PRIO conflicts during the 1989–2008 period, conditional on the potentially confounding effects of geographic, climatic, and development-related characteristics, as well as the disease environment. Regional dummies are indicators for 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 land suitability, distance from waterway, and an island dummy. Climatic controls are the mean and standard deviation of temperature and precipitation. Development outcomes are log luminosity, time since settlement, and population density, and 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 average yearly share of the area of each ethnic homeland that was within the boundaries of internal armed conflict over the period 1989–2008. The β^* statistic is the estimated effect of population diversity, if the proportion of selection on observables and unobservables is equal, and the maximal R^2 is equal to 1.3 times the observed R^2 (Oster, 2017). Cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent.

polarization, each of which are insignificantly associated with the conflict measure. Finally, columns 6 and 7, demonstrates that the estimates remain highly significant and stable if one accounts for a set of potentially endogenous confounders (i.e., log luminosity, malaria endemicity, time since settlement, and population density).

In light of the potential endogeneity of observed population diversity, Table 9 explore the effect of predicted population diversity, based on migratory distance from East Africa, on the prevalence of spatio-temporal conflicts in a sample of 1,238 ethnic groups. In particular, column 1 establishes a highly significant effect of predicted diversity across the 1,238 ethnic groups on the conflict measure, conditional on the 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 and 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., mean and standard deviation of temperature and precipitation). In particular, the estimated coefficient of interest in Column 2 suggests that an increase in the predicted diversity from the 10th percentile (e.g., the Zapotec people of Mesoamerica) to the 90th percentile (e.g., the Dinka of Africa) corresponds to an increase in the prevalence of spatio-temporal conflict by roughly 127 percentage points (compared to a sample mean of 0.15 and a standard deviation of 0.29). The larger coefficient on predicted diversity, relative to that on observed diversity —which also holds in

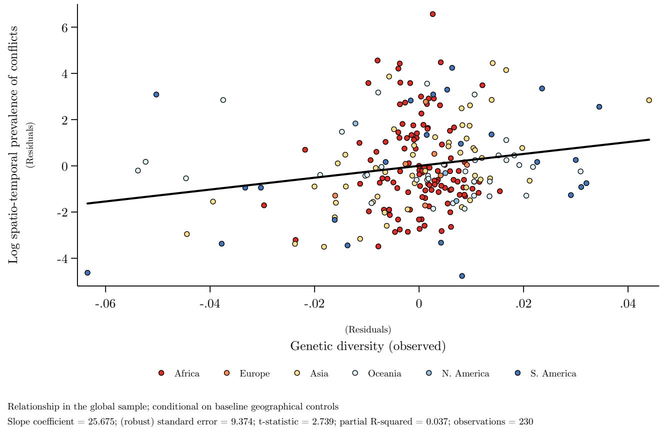


FIGURE 5: Observed Population Diversity and Conflict across Ethnic Homelands

Notes: This figure depicts the global cross-ethnic homeland relationship between contemporary observed population diversity and log spatio-temporal prevalence of UCDP/PRIO conflicts, 1989–2008, conditional on regional dummies, and potential geographic and climatic confounders, as reported in Column 3 of Table 5.

the observed diversity sample— suggests that indeed the use of predicted diversity mitigates the potential reduction in the diversity of ethnic populations due to conflict.

Columns 3 and 4 of Table 9 establish that the qualitative results are unaffected by accounting for the potentially confounding effects of linguistic fractionalization and polarization (where linguistic fractionalization is significantly associated with the conflict measure), accounting for a set of potentially endogenous confounders (i.e., log luminosity, malaria endemicity, time since settlement, and population density). Importantly, restricting the analysis to the sample of 276 linguistically homogeneous ethnic homelands, for which arguably the measure of conflicts capture primarily intragroup conflict, Columns 5 and 6 suggest that the effect of predicted diversity on conflict remains highly significant and stable.

Finally, using the migratory distance from Africa as an instrumental variable for observed population diversity, the 2SLS regression analysis reported in Columns 7–9, suggests that there exists a highly significant reduced-form impact of population diversity on conflict, accounting for the potentially confounding effects of linguistic fractionalization and polarization in the ethnic homelands, as well as geographical and climatic characteristics, regional fixed-effects, development outcomes and the disease environment.[40] In line with the results based on predicted diversity, once the potential reduction in the 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 of the observed level of diversity (e.g., the Guaraní people of South America) to the 90th percentile (e.g., the Bulu people of South Africa) corresponds to an increase in the prevalence of spatio-temporal conflict by 154 percentage points (compared to a sample mean of 0.13 and a standard deviation of 0.25).

In Appendix A.5, the analysis in Table A.8 establishes population diversity as a significant contributor to both the total number of conflict events and the total number of conflict-related deaths within an ethnic homeland during the 1989–2008 time period. In addition, Table A.9 explores

[40] The first-stage F-statistic indicates that the migratory distance is not a weak instrument.

TABLE 6: Population Diversity and the Spatiotemporal Prevalence of Conflict across Ethnic Homelands – Predicted Diversity

	Log spatio-temporal prevalence of UCDP GED conflicts, 1989–2008								
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS	(8) 2SLS	(9) 2SLS
Genetic diversity (predicted)	80.672***	77.583***	73.810***	75.385***	79.380***	74.089***			
	[5.738]	[7.142]	[7.513]	[7.488]	[10.765]	[11.029]			
Genetic diversity (observed)							104.544***	94.221***	97.170***
							[33.642]	[28.734]	[28.600]
Linguistic fractionalization			0.671**					-0.091	
			[0.272]					[0.522]	
Linguistic polarization				0.255					-0.646
				[0.211]					[0.506]
Regional dummies	×	×	×	×	×	×	×	×	×
Geographical controls		×	×	×	×	×	×	×	×
Climatic controls		×	×	×	×	×	×	×	×
Development outcomes			×	×		×		×	×
Disease environment controls			×	×		×		×	×
Sample	Predicted	Predicted	Predicted	Predicted	Homogenous	Homogenous	Observed	Observed	Observed
Observations	1238	1238	1238	1238	276	276	230	230	230
Effect of 10th90th %ile move in diversity	1.328***	1.277***	1.215***	1.241***	1.116***	1.042***	1.546***	1.394***	1.437***
meserrest	[0.094]	[0.118]	[0.124]	[0.123]	[0.151]	[0.155]	[0.498]	[0.425]	[0.423]
Adjusted R^2	0.372	0.409	0.438	0.435	0.441	0.491			
β^*		68.476	61.569	65.659	71.240	62.332			
First stage coefficient							0.653	0.653	0.653
							(0.203)	(0.203)	(0.203)
First-stage F-statistic							10.312	17.877	17.632

Notes: This table exploits cross-ethnicity variations to establish a significant positive reduced-form impact of contemporary population diversity on the log spatio-temporal prevalence of UCDP/PRIO conflicts during the 1989–2008 period, conditional on the potentially confounding effects of geographic, climatic, and development-related characteristics, as well as the disease environment. Regional dummies are indicators for 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 land suitability, distance from waterway, and an island dummy. Climatic controls are the mean and standard deviation of temperature and precipitation. Development outcomes are log luminosity, time since settlement, and population density, and 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 average yearly share of the area of each ethnic homeland that was within the boundaries of internal armed conflict over the period 1989–2008. The 2SLS regressions exploit 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 the proportion of selection on observables and unobservables is equal, and the maximal R^2 is equal to 1.3 times the observed R^2 (Oster, 2017). Cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent.

the impact of predicted population diversity on the extensive and intensive margins of conflict.[41] Further, the baseline results with respect to the spatiotemporal prevalence of conflicts across ethnic homelands are shown to be robust to accounting for: (i) spatial dependence across observations (Tables A.10 and A.11), and (ii) the use of predicted population diversity as a generated regressor (Table A.12).

Finally, in Section B.3 of the Supplemental Material, the baseline results are shown to be qualitatively insensitive to introducing controls for: (a) migratory distances from historical technological frontiers (Table SB.1), and (b) ecological diversity and ecological polarization (Tables SB.2 and SB.3).

[41]In particular, Column 1, 2, 4, 5, 7 and 8 establish that observed and predicted diversity is positively correlated with the number of conflicts, conflict-related deaths, and deaths per conflict, accounting for regional fixed effects. Columns 3, 6 and 9 show that the estimated influence of predicted diversity is robust to controlling for geographic and climatic controls as well as linguistic fractionalization and polarization.

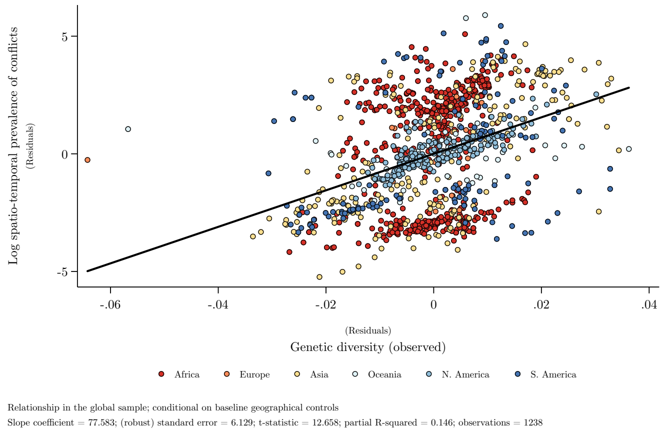


Figure 6: Predicted Population Diversity and Conflict across Ethnic Homelands

Notes: This figure depicts the global cross-ethnic homeland relationship between contemporary predicted population diversity and log spatio-temporal prevalence of UCDP/PRIO conflicts, 1989–2008, conditional on regional dummies, and potential geographic and climatic confounders, as reported in Column 2 of Table 6.

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 of our 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, we provide evidence that our 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 number of ethnic groups in a national population Fearon (2003);[42] (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;[43] 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

[42] 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.

[43] 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?"

in self-reported individual political positions on a politically "left"–"right" categorical scale, based on data from the WVS.[44]

Table 7 reports the findings from our 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, we present 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, throughout our analysis, we restrict our specifications to partialling out the influence of only our baseline set of geographical covariates (including continent or regional fixed effects). Specifically, we do not include potentially endogenous control variables, many of which (like GDP per capita) may well be afflicted 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.

For our analysis of each mechanism, we proceed 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 83.6 percent of a standard deviation from the cross-country distribution of this particular measure.[45]

The latter two regressions in our 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

[44] 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, we standardize the cross-country distribution of this variable prior to conducting our regressions.

[45] The three scatter plots presented in Figure A.1 of Appendix A.3 depict these statistically significant cross-country relationships, conditional on our 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)).

Table 7: 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	Frequency of new PRIO25 civil conflict onsets, 1960–2017		Prevalence of interpersonal trust	Frequency of new PRIO25 civil conflict onsets, 1960–2017		Variation in political attitudes	Frequency of new PRIO25 civil conflict onsets, 1960–2017	
Population diversity (ancestry adjusted)	5.187***	0.316***	0.294***	−1.817**	0.488**	0.447*	14.552**	0.451**	0.375
	(1.887)	(0.114)	(0.109)	(0.848)	(0.221)	(0.236)	(6.772)	(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*	0.836**	0.027**	0.022
	(0.777)	(0.008)	(0.007)	(0.049)	(0.013)	(0.014)	(0.389)	(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.

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 our findings in Table 7 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, we do not interpret our findings from this analysis as conclusive evidence per se 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. This is a task that we leave 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 and interpersonal trust, using individual data. 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 Ethnic-Homeland Population Diversity and Individual Trust 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) the homeland country fixed effects, (ii) the host-country fixed effects, (iii) host district characteristics (i.e., presence of school, electricity, piped water, sewage, health clinic, and urban status), (iv) individuals controls (i.e., age, gender, living condition, education, occupation, and religion), and (v) slavery. 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 8 presents the regression analysis of various measures of trust (across individuals in Africa) on interpersonal population diversity in their ethnic homelands.[46] Panel A establishes the association between observed population diversity in the individual's ethnic homeland and intra-group trust. The coefficient suggests that an increase in observed population diversity from the 10th percentile of the distribution (e.g., individuals belonging to the Mandinka people) to the 90th percentile (e.g., individuals belonging to the Turu people) corresponds to a 0.26 points decrease in intra-group trust (compared to a sample mean of 1.47, a minimum of 0, a maximum of 3, and a standard deviation of 1.00). Panel B establishes the negative association between observed population diversity and trust in relatives. Finally, Panel C demonstrates the negative association between observed population diversity and trust in neighbors.

The analysis further suggests that the linguistic polarization in the ethnic homeland has an adverse effect on intra-group trust but linguistic fractionalization appears uncorrelated with it. Moreover, conditional on fragmentation in the ethnic homeland, fractionalization in the host district appears uncorrelated with intra-group trust. However it appears that all measures of fragmentation

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

Table 8: Ethnic-Homeland Population Diversity and Individual-Level Trust in Africa

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) OLS	(12) OLS
Panel A	Intra-group trust											
Population diversity (observed)	-23.010**	-21.851**	-28.775**	-16.674***	-25.081***	-20.255***	-20.140***	-20.140***	-19.951***	-20.120***	-20.308***	-26.424***
	(10.472)	(10.148)	(11.959)	(5.556)	(7.096)	(5.180)	(7.216)	(7.216)	(7.452)	(7.335)	(6.462)	(8.901)
Linguistic fractionalization					-0.474		0.008	0.008	0.004	0.005	0.008	-0.296
					(0.297)		(0.239)	(0.239)	(0.234)	(0.229)	(0.240)	(0.210)
Linguistic polarization						-1.265***	-1.276***	-1.276***	-1.275***	-1.273***	-1.272***	-1.108*
						(0.451)	(0.484)	(0.484)	(0.484)	(0.490)	(0.494)	(0.577)
District-level ethnic fractionalization									-0.029		0.031	-0.025
									(0.042)		(0.210)	(0.204)
Proportion of ethnic group in district										0.040	0.062	-0.007
										(0.085)	(0.224)	(0.204)
Country-fixed effects	×	×	×	×	×	×	×	×	×	×	×	×
Baseline individual controls		×	×	×	×	×	×	×	×	×	×	×
Slave export control			×	×	×	×	×	×	×	×	×	×
Development controls				×	×	×	×	×	×	×	×	×
Urban area dummy				×	×	×	×	×	×	×	×	×
Living conditions dummies				×	×	×	×	×	×	×	×	×
Education dummies				×	×	×	×	×	×	×	×	×
Occupation dummies				×	×	×	×	×	×	×	×	×
Religion dummies				×	×	×	×	×	×	×	×	×
Homeland country-fixed effects												×
Number of Observations	3448	3448	3448	3438	3438	3438	3438	3438	3438	3438	3438	3438
Adjusted R2	0.220	0.227	0.236	0.266	0.270	0.275	0.275	0.275	0.275	0.275	0.275	0.280
Effect of 10th–90th %ile move in diversity	-0.230**	-0.219**	-0.288**	-0.167***	-0.251***	-0.203***	-0.201***	-0.201***	-0.200***	-0.201***	-0.203***	-0.264***
	(0.105)	(0.101)	(0.120)	(0.056)	(0.071)	(0.052)	(0.072)	(0.072)	(0.075)	(0.073)	(0.065)	(0.089)
Panel B	Trust in relatives											
Population diversity (observed)	-35.552**	-34.369**	-33.129**	-23.873*	-36.412*	-27.907***	-32.248***	-32.248***	-31.141***	-32.119***	-32.602***	-25.606***
	(13.839)	(13.492)	(14.179)	(13.689)	(18.630)	(8.074)	(11.388)	(11.388)	(10.780)	(11.537)	(11.874)	(9.802)
Linguistic fractionalization					-0.711***		-0.307	-0.307	-0.334	-0.324	-0.316	-0.535**
					(0.215)		(0.269)	(0.269)	(0.253)	(0.239)	(0.247)	(0.266)
Linguistic polarization						-1.459***	-1.068*	-1.068*	-1.063*	-1.055*	-1.054*	-0.719
						(0.509)	(0.616)	(0.616)	(0.600)	(0.614)	(0.621)	(0.697)
District-level ethnic fractionalization									-0.176**		0.082	0.100
									(0.085)		(0.112)	(0.126)
Proportion of ethnic group in district										0.208***	0.267***	0.270**
										(0.077)	(0.103)	(0.114)
Controls as in same column of Panel A	×	×	×	×	×	×	×	×	×	×	×	×
Number of Observations	3460	3460	3460	3450	3450	3450	3450	3450	3450	3450	3450	3450
Adjusted R2	0.062	0.070	0.070	0.114	0.125	0.127	0.128	0.128	0.130	0.132	0.131	0.132
Effect of 10th–90th %ile move in diversity	-0.356**	-0.344**	-0.331**	-0.239*	-0.364*	-0.279***	-0.322***	-0.322***	-0.311***	-0.321***	-0.326***	-0.256***
	(0.138)	(0.135)	(0.142)	(0.137)	(0.186)	(0.081)	(0.114)	(0.114)	(0.108)	(0.115)	(0.119)	(0.098)
Panel C	Trust in neighbours											
Population diversity (observed)	-27.106**	-25.428**	-28.361**	-20.127***	-31.392***	-24.245***	-26.522***	-26.522***	-25.629***	-26.458***	-26.487***	-32.532**
	(11.228)	(10.655)	(12.161)	(6.708)	(11.397)	(7.332)	(8.814)	(8.814)	(9.171)	(9.107)	(8.876)	(13.277)
Linguistic fractionalization					-0.636***		-0.161	-0.161	-0.182	-0.173	-0.173	-0.339*
					(0.246)		(0.194)	(0.194)	(0.175)	(0.173)	(0.175)	(0.200)
Linguistic polarization						-1.461***	-1.256***	-1.256***	-1.253***	-1.246***	-1.246***	-1.194**
						(0.426)	(0.467)	(0.467)	(0.465)	(0.476)	(0.478)	(0.555)
District-level ethnic fractionalization									-0.140***		0.005	-0.037
									(0.037)		(0.174)	(0.175)
Proportion of ethnic group in district										0.146***	0.150	0.105
										(0.029)	(0.145)	(0.143)
Controls as in same column of Panel A	×	×	×	×	×	×	×	×	×	×	×	×
Number of Observations	3452	3452	3452	3442	3442	3442	3442	3442	3442	3442	3442	3442
Adjusted R2	0.131	0.145	0.147	0.187	0.195	0.200	0.200	0.200	0.201	0.201	0.201	0.203
Effect of 10th–90th %ile move in diversity	-0.271**	-0.254**	-0.284**	-0.201***	-0.314***	-0.242***	-0.265***	-0.265***	-0.256***	-0.265***	-0.265***	-0.325**
	(0.112)	(0.107)	(0.122)	(0.067)	(0.114)	(0.073)	(0.088)	(0.088)	(0.092)	(0.091)	(0.089)	(0.133)

Notes: This table presents the results of an individual-level OLS regression analysis of interpersonal trust towards individuals of the same ethnicity (as reported recorded 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 country in which the homeland of the individual's ethnicity is located. Two-way cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent.

both at the ethnic homeland and the host districts are negative associated with trust in relatives and neighbors.

Table 9: Country-of-Origin Population Diversity and Individual-Level Trust among Second-Generation U.S. Immigrants

	Trust							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Population diversity (ancestry adjusted)	-8.862***	-8.967***	-8.308***	-7.632***	-7.519***	-7.607***	-7.068***	-7.138***
	(2.135)	(2.164)	(2.062)	(2.061)	(2.096)	(2.165)	(1.572)	(1.985)
Linguistic fragmentation					0.012		0.040	0.045
					(0.117)		(0.112)	(0.108)
Linguistic polarization						-0.008	-0.061	-0.006
						(0.237)	(0.264)	(0.262)
Regional dummies	×	×	×	×	×	×	×	×
GSS year		×	×	×	×	×	×	×
Baseline individual controls			×	×	×	×	×	×
Living conditions dummies				×	×	×	×	×
Religion dummies				×	×	×	×	×
Education dummies				×	×	×	×	×
Region of interview dummies								×
Number of observations	2297	2297	2297	1789	1789	1789	1789	1789
Adjusted R^2	0.047	0.065	0.108	0.168	0.168	0.168	0.168	0.178
Effect of 10th–90th %ile move in diversity	-0.615***	-0.623***	-0.577***	-0.530***	-0.522***	-0.528***	-0.491***	-0.496***
	(0.148)	(0.150)	(0.143)	(0.143)	(0.146)	(0.150)	(0.109)	(0.138)

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 host region fixed-effect in the US. The trust variable takes the value 1 if the respondent replies that "people cannot be trusted"; 2 if the respondent replies that "it depends", and 3 if the respondent replies that "people can be trusted". Two-way cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent.

5.2.2 Country-of-Origin Population Diversity and Individual Trust amongst Second-Generation Immigrants to the U.S.

This subsection explores the effect of the level of population diversity in the parental country of origin of second-generation migrants in the United States on their level of trust (as reported in the General Social Survey, GSS). The analysis accounts for time-invariant unobserved heterogeneity in the host country (e.g., geographical, cultural, and institutional characteristics), a range of individual controls, and the degree of linguistic fractionalization in the parental country of origins, mitigating possible concerns about the confounding effect of host country-specific characteristics.

Table 9 explores the association between the trust of second-generation migrants and the degree of population diversity in their parental country of origin, following Galor and Özak (2016). 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 effect associated with the parental country of origin. This highly significant negative association remains largely stable if one accounts for interview-year fixed effects (Column 2), the respondent's age-specific and sex-specific fixed effects (Column 3), living condition, relative income of the family at the age of 16, religion , and educational attainment (Column 4). Moreover, the results are robust to controlling for region-specific fixed effects within the United States. The estimate of the coefficient of interest in Column 7 suggests that an increase in population diversity in the parental country of origin from the 10th percentile of the predicted contemporary level of diversity (e.g., individuals of Mexican descent) to the 90th percentile (e.g., individuals of Hungarian descent) corresponds to a decrease in trust by 0.50 points (compared to a sample mean of 1.87, a minimum of 0, a maximum of 3, and a standard deviation of 0.97). The analysis further suggests that the linguistic polarization and polarization in the parental country of origin has no association with trust.

6 Concluding Remarks

This research 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 contributed significantly to the emergence, prevalence, recurrence, and severity of intrasocietal conflicts. Exploiting an exogenous source of variations in population diversity across nations and ethnic groups, it demonstrates that interpersonal population diversity has contributed significantly to the risk and intensity of historical and contemporary internal conflicts, accounting for the confounding effects of geographical, institutional, and cultural characteristics, as well as for the level of economic development. These findings arguably reflect the adverse effect of population diversity on interpersonal trust, its contribution to divergence in preferences for public goods and redistributive policies, and its impact on the degree of fractionalization and polarization across ethnic, linguistic, and religious groups.

Appendix

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

Our 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, our earlier findings reflect the influence of interpersonal population diversity on the *extensive* margin of conflict. In this appendix section, we explore the influence of population diversity on the *intensive* margin of conflict. In particular, we employ 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.

Our 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, 2010).[47] 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 that we employ 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.[48] 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.

Our 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, 2010). 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.[49]

[47]The specific version of the MEPV data set that we employ 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 our measure of conflict intensity from the MEPV data set.

[48]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.

[49]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.1: 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)	4.675***	3.992**	4.325***	3.780**	3.664**	2.372***	3.543**	2.461***
	(1.566)	(1.820)	(1.554)	(1.910)	(1.561)	(0.852)	(1.568)	(0.919)
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	1,147	1,001	1,427	1,265	1,018	880	1,272	1,119
Countries	123	121	149	147	123	120	149	146
Partial R^2 of population diversity	0.025	0.019			0.009	0.006		
Adjusted R^2	0.640	0.627			0.082	0.104		
Effect of 10th–90th %ile move in diversity	0.251***	0.188**	0.291***	0.258**	0.197**	0.114***	0.238**	0.168***
	(0.084)	(0.086)	(0.104)	(0.130)	(0.084)	(0.041)	(0.105)	(0.063)
First-stage F statistic			148.488	98.667			145.650	90.298

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 1960–2008 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 our analysis of conflict severity follows Esteban et al. (2012), in terms of exploiting variations in quinquennially repeated cross-country data, for each country in our sample, we collapse the annual data on both measures of conflict intensity to a quinquennial time series, by assigning to any given 5-year interval in our post-1960 sample period, the maximum level of conflict intensity reflected by that measure across all years in that 5-year interval. As in our analysis of civil conflict incidence or onset, we focus on better-identified specifications that either (i) exploit variations in a sample of countries belonging only to the Old World, or (ii) exploit migratory

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/.

distance from East Africa as an instrument for contemporary population diversity in a global sample of countries. In all regressions, we 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 either of our proxies for conflict intensity are measured in the data have no natural interpretation, we standardized the outcome variables prior to running the regressions.

Table A.1 presents the results from our 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).[50] Regardless of the measure for conflict intensity examined, the identification strategy exploited, or the set of covariates considered by the specification, the results from our analysis of conflict severity in Table A.1 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 19 to 29 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 11 to 24 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), we exploit 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 due to 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.[51] Specifically, we compare 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), and we examine 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.

In addition, we compute the δ and β^* statistics suggested by Oster (2017). The δ statistic reflects how strongly correlated the unobservables need to be with population diversity, relative to

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

[51] 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.2: 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 1. 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 1. The reader is therefore referred to Table 1 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.

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 (2017) shows that if zero does not belong to the interval between the estimated coefficient on population diversity and β^*, then we can reject the null hypothesis that our coefficient of interest is exclusively driven by unobservables.

We consider the specification from Column 3 of Table 1 as our restricted model. This specification includes, besides population diversity, our baseline geographical controls and continent fixed effects. We assess coefficient stability relative to the augmented specification presented in Column 8 that includes our 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.[52] 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 our estimate down to zero. Assuming that the unobservables are equally correlated with population diversity as are the observables, and that these correlations have

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

Table A.3: 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								
Popopulation 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***	0.027***	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 1. 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 1. The reader is therefore referred to Table 1 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 we were able to control for all unobservables, would be $\beta^* = 1.15$. Thus, the interval between the actual coefficient estimate from our full specification (0.309) and β^* excludes zero.[53] We therefore conclude that it is highly unlikely that our results could be explained away by omitted variables.

Robustness to Examining the Count of Civil Conflict Onset across Countries Given that our baseline cross-country regressions employ least-squares estimation, we apply a log transformation to our 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. In Table A.2, we consider 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 our instrument-based identification strategy in the global sample of countries, Columns 8–9 examine robustness to employing the Poisson rather than the negative binomial-

[53] 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 our model if we could include all unobservables correlated with population diversity to the model.

Table A.4: 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***
	(0.068)	(0.108)	(0.123)	(0.129)	(0.137)	(0.188)	(0.202)	(0.186)	(0.229)
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)	(0.008)	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)	(0.013)	(0.017)

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 1. 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 1. The reader is therefore referred to Table 1 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. Bootstrap 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 our 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, in Table A.3, we replicate our key specifications from Table 1 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 our main cross-country findings remain qualitatively intact. Thus, as far as our cross-country analysis of the influence of population diversity on conflict frequency is concerned, spatial dependence across observations does not appear to be a confounding issue.

Robustness to Accounting for Population Diversity as a Generated Regressor Our proxy measure of contemporary population diversity is a generated regressor in our empirical 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

Table A.5: 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***	24.262***	24.692***	22.312***	24.442***
	(4.285)	(5.482)	(4.256)	(5.808)	(5.356)	(6.414)	(4.951)	(5.557)
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			0.433	0.457		
Marginal effect of diversity	1.850***	1.212**	2.005***	1.786**	3.837***	3.521***	3.797***	3.826***
	(0.540)	(0.617)	(0.631)	(0.777)	(0.837)	(0.930)	(0.910)	(1.013)

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 2. 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 1990–2008 time period (Columns 5–8). The specifications examined in this table are otherwise identical to corresponding ones reported in Columns 1–4 of Table 2. The reader is therefore referred to Table 2 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.

all pairs of ethnic groups in this sample. Table A.4 therefore checks the robustness of our standard-error estimates to accounting for potential bias due to the use of a generated regressor. To perform this robustness check, the analysis replicates our key specifications from Table 1, 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 1, 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 2, 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.5 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 1990–2008 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

Table A.6: Population Diversity and the Incidence of Civil Conflict in Repeated Cross-Country Data – Robustness to Examining the Annual Incidence or Quinquennial Prevalence of Conflict

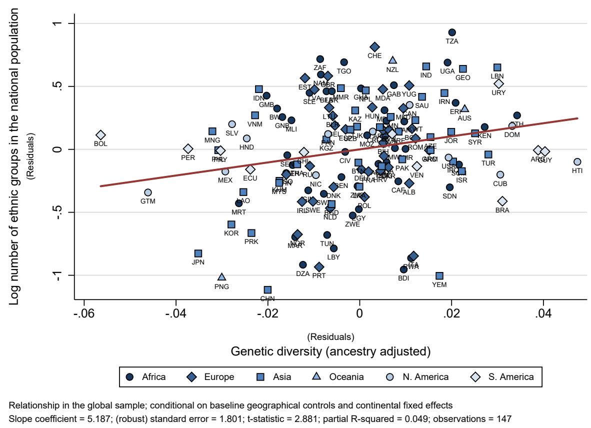
Cross-country sample:	Old World		Global		Old World		Global	
	(1) Probit	(2) Probit	(3) IV Probit	(4) IV Probit	(5) OLS	(6) OLS	(7) 2SLS	(8) 2SLS
	Annual PRIO25 civil conflict incidence, 1960–2017				Quinquennial PRIO25 civil conflict prevalence, 1960-2017			
Population diversity (ancestry adjusted)	9.301***	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					0.096***	0.087***	0.136***	0.154***
					(0.031)	(0.032)	(0.043)	(0.055)

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 2. 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 2, 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 2 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.

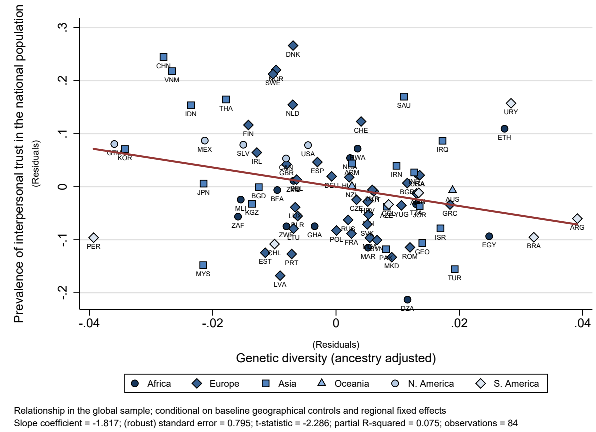
aforementioned types of intrastate conflict. To interpret our coefficient of interest, the IV probit regressions presented in Columns 4 and 8 suggest that conditional on our full set of control variables, a 1 percentage point increase in population diversity increases the quinquennial likelihoods of 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 Conflict In Table A.6, we check the robustness of our baseline results for the temporal incidence of civil conflict in repeated cross-country data, as shown in Columns 1–4 of Table 2, 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 2, 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.6, 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 our 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 15 percentage points in the prevalence of years with an active conflict in any given 5-year interval.

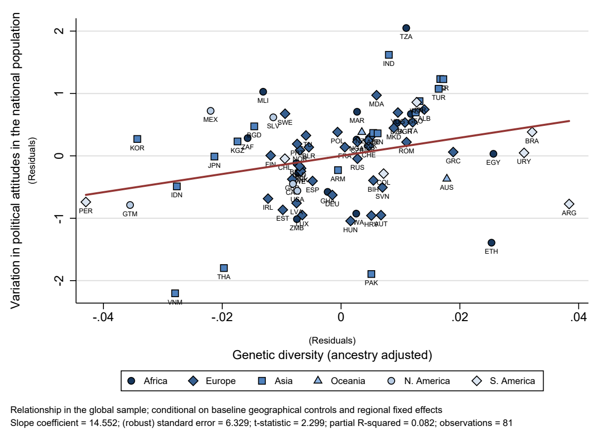
A.3 Appendix Figures for the Country-Level Analyses



(a) Number of ethnic groups



(b) Prevalence of interpersonal trust



(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 7. 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 7.

A.4 Descriptive Statistics at the Country Level

Table A.7: Summary Statistics of the 147-Country Sample for Explaining the Frequency of Civil Conflict Onsets

		Mean	SD	Min	Max	Percentile	
						10th	90th
(1)	Log new PRIO25 civil conflict onsets per year, 1960–2017	0.021	0.029	0.000	0.174	0.000	0.062
(2)	Population diversity (ancestry adjusted)	0.728	0.027	0.628	0.774	0.685	0.752
(3)	Ethnic fractionalization	0.467	0.254	0.002	0.930	0.115	0.792
(4)	Ethnolinguistic polarization	0.452	0.241	0.000	0.957	0.097	0.747
(5)	Absolute latitude	0.027	0.017	0.001	0.064	0.006	0.051
(6)	Ruggedness	0.125	0.126	0.004	0.747	0.018	0.278
(7)	Mean elevation	0.594	0.552	0.001	2.837	0.104	1.250
(8)	Range of elevation	1.701	1.389	0.040	6.176	0.283	3.752
(9)	Mean land suitability	0.386	0.246	0.003	0.951	0.046	0.718
(10)	Range of land suitability	0.715	0.264	0.000	0.999	0.317	0.994
(11)	Small island nation dummy	0.048	0.214	0.000	1.000	0.000	0.000
(12)	Distance to nearest waterway	0.353	0.458	0.014	2.386	0.036	1.010
(13)	Ever a U.K. colony dummy	0.259	0.439	0.000	1.000	0.000	1.000
(14)	Ever a French colony dummy	0.190	0.394	0.000	1.000	0.000	1.000
(15)	Ever a non-U.K./non-French colony dummy	0.320	0.468	0.000	1.000	0.000	1.000
(16)	British legal origin dummy	0.252	0.435	0.000	1.000	0.000	1.000
(17)	French legal origin dummy	0.463	0.500	0.000	1.000	0.000	1.000
(18)	Executive constraints, 1960–2017 average	4.145	1.827	1.000	7.000	1.839	7.000
(19)	Fraction of years under democracy, 1960–2017	0.408	0.377	0.000	1.000	0.000	1.000
(20)	Fraction of years under autocracy, 1960–2017	0.352	0.323	0.000	1.000	0.000	0.879
(21)	Oil or gas reserve discovery	0.673	0.471	0.000	1.000	0.000	1.000
(22)	Log population, 1960–2017 average	16.087	1.431	12.916	20.798	14.423	17.877
(23)	Log GDP per capita, 1960–2017 average	7.703	1.489	5.100	10.630	5.705	9.937

A.5 Robustness Checks for the Ethnicity-Level Analyses

Table A.8: Population Diversity and the Number of Conflicts or Conflict-Related Deaths across Ethnic Homelands

	Log spatio-temporal prevalence of UCDP GED conflicts, 1989–2008							
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS
cnumcentroid								
Genetic diversity (observed)	64.871***	55.390***	62.882***	66.407***				
	[14.816]	[18.066]	[20.814]	[17.498]				
Genetic diversity (predicted)					57.211***	49.322***	48.224***	44.033***
					[8.696]	[8.419]	[11.113]	[10.582]
Regional dummies	×	×	×	×	×	×	×	×
Geographical controls		×	×	×		×	×	×
Climatic controls			×	×			×	×
Fragmentation controls				×				×
Development outcomes				×				×
Disease environment controls				×				×
Sample	Observed	Observed	Observed	Observed	Observed	Observed	Observed	Observed
Observations	230	230	230	230	1238	1238	1238	1238
PseudoR^2	.	.	.	.	.	.	.	.
mecoef	93.355***	79.714***	90.497***	95.571***	40.805***	35.178***	34.398***	31.408***
meserr	[27.497]	[28.122]	[31.114]	[27.252]	[8.871]	[7.822]	[9.429]	[8.893]

Notes: This table exploits cross-ethnicity variations to establish a significant positive reduced-form impact of contemporary population diversity on the log spatio-temporal prevalence of UCDP/PRIO conflicts and conflict-related deaths 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. Cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table A.9: Population Diversity and Alternative Conflict Outcomes across Ethnic Homelands

	Log number of conflicts			Log number of deaths		
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Genetic diversity (observed)	5.759**		28.352***		22.594***	
	[2.294]		[10.412]		[8.504]	
Genetic diversity (predicted)		8.242***		38.132***		29.890***
		[1.516]		[6.876]		[5.615]
Linguistic fractionalization	0.129	0.179	1.269	0.785	1.140	0.606
	[0.288]	[0.119]	[1.199]	[0.542]	[0.971]	[0.438]
Linguistic polarization	0.103	-0.165*	-0.250	-0.761*	-0.352	-0.596*
	[0.241]	[0.087]	[1.049]	[0.398]	[0.849]	[0.322]
Regional dummies	×	×	×	×	×	×
Geographical controls	×	×	×	×	×	×
Climatic controls	×	×	×	×	×	×
Sample	Observed	Predicted	Observed	Observed	Observed	Predicted
Observations	230	1238	230	1238	230	1238
Effect of 10th90th %ile move in diversity	1.559**	1.540***	11954.160***	6390.014***	1113.854***	794.434***
meserrest	[0.621]	[0.283]	[4389.922]	[1152.334]	[419.263]	[149.228]
Adjusted R^2	0.332	0.254	0.347	0.260	0.332	0.249

Notes: This table exploits cross-ethnicity variations 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 effect on the outcome variables in levels (i.e., not logged) associated with increasing population diversity from the tenth to the ninetieth percentile of its cross-country distribution is expressed in terms of the prevalence of spatio-temporal of conflict. Cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent.

Table A.10: Observed Population Diversity and the Spatiotemporal Prevalence of Conflict across Ethnic Homelands – Robustness to Accounting for Spatial Dependence

	Log spatio-temporal prevalence of UCDP GED conflicts, 1989–2008						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS
Log Average Area-Time Conflict Share							
Genetic diversity (observed)	36.221***	33.733***	33.996***	33.218***	34.151***	39.657***	40.660***
	[9.565]	[8.629]	[8.551]	[8.531]	[8.552]	[8.859]	[8.837]
Icontcomb2	-2.686	-2.071	-1.097	-1.061	-1.151	-1.652	-1.831
	[2.155]	[1.878]	[1.806]	[1.798]	[1.808]	[1.784]	[1.790]
Icontcomb3	-2.144	-4.330**	-2.693	-2.594	-2.805	-3.188	-3.511*
	[2.322]	[2.135]	[2.111]	[2.102]	[2.121]	[2.100]	[2.118]
Icontcomb4	-5.084**	-2.463	-1.594	-1.597	-1.570	-2.087	-2.119
	[2.267]	[2.043]	[1.965]	[1.956]	[1.965]	[1.932]	[1.931]
Icontcomb5	-5.083**	-2.607	-1.434	-1.558	-1.393	-1.769	-1.700
	[2.498]	[2.199]	[2.140]	[2.132]	[2.140]	[2.133]	[2.130]
wblac	5.469***	1.624	1.425	1.428	1.362	1.315	1.221
	[1.414]	[1.386]	[1.364]	[1.358]	[1.369]	[1.332]	[1.336]
wbssa	-3.624*	-5.590***	-4.622**	-4.728***	-4.632**	-5.826***	-5.862***
	[2.156]	[1.888]	[1.838]	[1.831]	[1.837]	[1.883]	[1.883]
Absolute latitude		-0.135***	-0.101***	-0.103***	-0.102***	-0.084**	-0.083**
		[0.018]	[0.037]	[0.037]	[0.037]	[0.037]	[0.037]
Ruggedness		0.082	0.046	0.046	0.050	0.024	0.026
		[0.155]	[0.161]	[0.160]	[0.161]	[0.161]	[0.161]
Elevation (mean)		-0.390	0.246	0.242	0.235	0.312	0.304
		[0.300]	[0.392]	[0.391]	[0.393]	[0.393]	[0.394]
Elevation (range)		-0.000***	-0.000**	-0.000**	-0.000**	-0.000*	-0.000*
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Land suitability (mean, post-1500)		-0.000***	-0.000***	-0.000***	-0.000***	-0.000**	-0.000**
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Land suitability (range, post-1500)		0.000***	0.000***	0.000***	0.000***	0.000**	0.000**
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
island		-0.299	0.241	0.198	0.276	0.203	0.291
		[0.878]	[0.873]	[0.869]	[0.875]	[0.860]	[0.862]
Nearest waterway		-0.053	-0.038	-0.042	-0.036	-0.111	-0.110
		[0.088]	[0.085]	[0.085]	[0.085]	[0.086]	[0.086]
Average temperature			0.150***	0.144***	0.151***	0.166***	0.172***
			[0.048]	[0.048]	[0.048]	[0.049]	[0.049]
Volatility of temperature			5.296***	5.166***	5.348***	4.930***	5.060***
			[1.508]	[1.504]	[1.511]	[1.526]	[1.528]
Average precipitation			0.005	0.005	0.005	0.003	0.003
			[0.005]	[0.005]	[0.005]	[0.006]	[0.006]
Volatility of precipitation			-0.025	-0.021	-0.024	-0.008	-0.009
			[0.036]	[0.036]	[0.036]	[0.038]	[0.038]
Linguistic fractionalization				0.595		0.302	
				[0.447]		[0.443]	
Linguistic polarization					-0.231		-0.387
					[0.438]		[0.428]
Time since initial settlement						0.281	0.251
						[0.498]	[0.500]
Malaria endemicity						0.386	0.352
						[1.308]	[1.305]
Oil or gas reserve discovery						0.449	0.487
						[0.352]	[0.349]
Log luminosity						-0.274***	-0.288***
						[0.102]	[0.101]
Ipopdenspe2						-0.298	-0.295
						[0.630]	[0.629]
Ipopdenspe3						0.485	0.560
						[0.595]	[0.589]
Ipopdenspe4						-0.596	-0.537
						[0.647]	[0.645]
Ipopdenspe5						0.314	0.401
						[0.661]	[0.660]
Ipopdenspe9999						0.431	0.536
						[0.593]	[0.589]
dobj							
Log Average Area-Time Conflict Share	0.769***	0.884	0.798***	0.804***	0.795***	0.817***	0.812***
	[0.149]	[0.098]	[0.141]	[0.139]	[0.142]	[0.133]	[0.135]
e.Log Average Area-Time Conflict Share	0.909***	0.903***	0.885***	0.887***	0.883***	0.857***	0.850***
	[0.087]	[0.089]	[0.106]	[0.104]	[0.109]	[0.129]	[0.137]

/

Table A.11: Predicted Population Diversity and the Spatiotemporal Prevalence of Conflict across Ethnic Homelands – Robustness to Accounting for Spatial Dependence

	Log spatio-temporal prevalence of UCDP GED conflicts, 1989–2008								
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS	(8) 2SLS	(9) 2SLS
Log Average Area-Time Conflict Share									
Genetic diversity (observed)							40.015***	41.457***	41.923***
							[12.076]	[10.616]	[10.505]
Genetic diversity (predicted)	57.403***	52.186***	52.974***	53.445***	52.071***	50.047***			
	[5.333]	[5.285]	[5.307]	[5.332]	[9.462]	[9.151]			
Icontcomb2	0.753*	0.872**	0.648	0.481	3.380***	2.909**	-1.172	-1.741	-1.890
	[0.430]	[0.414]	[0.406]	[0.403]	[1.197]	[1.147]	[1.775]	[1.741]	[1.733]
Icontcomb3	0.838	0.830	0.894*	0.689	3.632***	2.980**	-3.672*	-4.334**	-4.552**
	[0.551]	[0.539]	[0.529]	[0.526]	[1.252]	[1.210]	[2.156]	[2.107]	[2.060]
Icontcomb4	-0.632	0.700	0.728	0.699	3.098**	2.752**	-1.651	-2.124	-2.125
	[0.460]	[0.461]	[0.453]	[0.454]	[1.384]	[1.321]	[1.922]	[1.881]	[1.865]
Icontcomb5	2.275***	3.178***	3.591***	3.725***	5.199***	4.784***	-1.359	-1.792	-1.728
	[0.553]	[0.547]	[0.552]	[0.553]	[1.382]	[1.330]	[2.124]	[2.093]	[2.076]
wblac	2.896***	0.910**	0.666*	0.559	1.556**	1.513**	1.602	1.217	1.039
	[0.378]	[0.388]	[0.382]	[0.381]	[0.698]	[0.686]	[1.434]	[1.341]	[1.331]
wbssa	0.859**	-0.043	-0.649*	-0.581	0.834	0.235	-4.636**	-5.592***	-5.577***
	[0.385]	[0.382]	[0.394]	[0.394]	[1.248]	[1.207]	[1.803]	[1.841]	[1.829]
Absolute latitude		-0.041***	-0.047***	-0.045***	-0.015	0.008	-0.117***	-0.098**	-0.096**
		[0.013]	[0.013]	[0.013]	[0.027]	[0.026]	[0.039]	[0.040]	[0.041]
Ruggedness		-0.001	0.025	0.031	0.148	0.070	0.045	-0.032	-0.050
		[0.057]	[0.057]	[0.057]	[0.117]	[0.117]	[0.164]	[0.161]	[0.160]
Elevation (mean)		0.396***	0.440***	0.470***	0.271	0.185	0.216	0.320	0.357
		[0.129]	[0.128]	[0.128]	[0.227]	[0.220]	[0.391]	[0.390]	[0.386]
Elevation (range)		-0.000	-0.000	-0.000	-0.000	-0.000	-0.000**	-0.000	-0.000
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Land suitability (mean, post-1500)		-0.000	-0.000	-0.000*	-0.000**	-0.000***	-0.000**	-0.000**	-0.000**
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Land suitability (range, post-1500)		0.000*	0.000*	0.000**	0.000	0.000	0.000**	0.000**	0.000**
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
island		-0.810***	-0.723***	-0.784***	-1.139***	-1.062***	0.514	0.328	0.268
		[0.247]	[0.250]	[0.251]	[0.373]	[0.380]	[0.962]	[0.946]	[0.931]
Nearest waterway		0.017	-0.015	-0.010	-0.291	-0.357**	-0.056	-0.129	-0.128
		[0.034]	[0.035]	[0.035]	[0.179]	[0.177]	[0.083]	[0.084]	[0.083]
Average temperature		0.061***	0.060***	0.067***	0.089**	0.111***	0.123**	0.141***	0.143***
		[0.019]	[0.019]	[0.019]	[0.043]	[0.042]	[0.050]	[0.051]	[0.051]
Volatility of temperature		-0.548	-0.256	-0.290	0.678	0.792	5.412***	4.834***	4.868***
		[0.643]	[0.632]	[0.634]	[1.333]	[1.299]	[1.535]	[1.538]	[1.537]
Average precipitation		-0.001	-0.003	-0.002	0.002	0.001	0.005	0.004	0.005
		[0.002]	[0.002]	[0.002]	[0.003]	[0.003]	[0.006]	[0.006]	[0.006]
Volatility of precipitation		0.001	0.008	0.006	0.004	0.014	-0.025	-0.010	-0.007
		[0.011]	[0.011]	[0.011]	[0.015]	[0.014]	[0.039]	[0.039]	[0.038]
Time since initial settlement			0.014	-0.023		0.727***		-0.035	-0.027
			[0.154]	[0.154]		[0.280]		[0.485]	[0.471]
Malaria endemicity			2.268***	3.458***		1.923		0.271	0.385
			[0.544]	[0.446]		[1.533]		[1.504]	[1.536]
Oil or gas reserve discovery			0.287**	0.303**		0.167		0.467	0.519
			[0.123]	[0.124]		[0.287]		[0.343]	[0.337]
Log luminosity			-0.066**	-0.056*		-0.121*		-0.261***	-0.269***
			[0.032]	[0.032]		[0.064]		[0.101]	[0.100]
Ipopdenspe2			0.172	0.129		0.300		-0.098	-0.033
			[0.169]	[0.170]		[0.332]		[0.615]	[0.610]
Ipopdenspe3			0.314*	0.305*		1.000**		0.435	0.478
			[0.179]	[0.180]		[0.405]		[0.586]	[0.576]
Ipopdenspe4			0.230	0.237		0.867**		-0.454	-0.408
			[0.181]	[0.182]		[0.405]		[0.634]	[0.627]
Ipopdenspe5			0.417**	0.427**		0.824*		0.418	0.412
			[0.198]	[0.199]		[0.427]		[0.642]	[0.633]
Ipopdenspe9999			0.253	0.228		-0.214		0.507	0.572
			[0.156]	[0.156]		[0.342]		[0.577]	[0.568]
Linguistic fractionalization			0.460**					0.293	
			[0.181]					[0.428]	
Linguistic polarization				0.081					-0.336
				[0.152]					[0.413]
dobj									
Log Average Area-Time Conflict Share	0.549***	0.661***	0.871***	1.039***	0.630***	0.632***	0.704***	0.704***	0.681***
	[0.049]	[0.046]	[0.064]	[0.034]	[0.143]	[0.140]	[0.160]	[0.145]	[0.128]
e.Log Average Area-Time Conflict Share	3.664***	3.922***	3.694***	3.694***	0.944***	0.932***	1.847***	2.176***	2.442***
	[0.058]	[0.035]	[0.043]	[0.043]	[0.050]	[0.060]	[0.398]	[0.556]	[0.562]
/									
var(e.lcsnm)	2.249***	1.889***	1.835***	1.849***	2.357***	2.117***			
	[0.096]	[0.081]	[0.079]	[0.079]	[0.211]	[0.189]			
Sample	Predicted	Predicted	Predicted	Predicted	Homogenous	Homogenous	Observed	Observed	Observed
Direct Impact of Genetic Diversity	57.498***	52.330***	53.416***	52.333***	52.683***	50.641***	40.648***	42.112***	42.517***
	(5.337)	(5.295)	(5.352)	(5.337)	(9.534)	(9.233)	(12.315)	(10.839)	(10.698)
Total Impact of Genetic Diversity	101.917***	114.794***	255.533**	-677.393	87.584***	84.475***	61.099***	63.303***	62.486***
	[10.954]	[15.415]	[111.747]	[646.422]	[22.695]	[22.117]	[22.105]	[19.970]	[18.427]
Observations	1097	1097	1097	1097	254	254	202	202	202
Effect of 10th90th %ile move in diversity									
meserrest									
First-stage F statistic									
Adjusted R^2									

Notes: This table exploits cross-ethnicity variations to establish a significant positive reduced-form impact of predicted population diversity on the log spatio-temporal prevalence of UCDP/PRIO conflicts during the 1989–2008 period, conditional

Table A.12: Predicted Population Diversity and the Spatiotemporal Prevalence of Conflict across Ethnic Homelands – Robustness to Accounting for Predicted Population Diversity as a Generated Regressor

	Log spatio-temporal prevalence of UCDP GED conflicts, 1989–2008					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Genetic diversity (predicted)	80.672***	77.583***	73.810***	75.385***	79.380***	74.089***
	[5.738]	[7.142]	[7.513]	[7.488]	[10.765]	[11.029]
Genetic diversity (observed)						
Linguistic fractionalization			0.671**			
			[0.272]			
Linguistic polarization				0.255		
				[0.211]		
Regional dummies	×	×	×	×	×	×
Geographical controls		×	×	×	×	×
Climatic controls		×	×	×	×	×
Development outcomes			×	×		×
Disease environment controls			×	×		×
Sample	Predicted	Predicted	Predicted	Predicted	Homogenous	Homogenous
Observations	1238	1238	1238	1238	276	276
Effect of 10th90th %ile move in diversity	1.328***	1.277***	1.215***	1.241***	1.116***	1.042***
meserrest	[0.094]	[0.118]	[0.124]	[0.123]	[0.151]	[0.155]
Adjusted R^2	0.372	0.409	0.438	0.435	0.441	0.491
Bootstrapped standard error	[7.643]***	[6.558]***	[6.896]***	[6.910]***	[8.665]***	[9.082]***

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. 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 6. The reader is therefore referred to Table 6 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 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. Cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent.

A.6 Descriptive Statistics at the Ethnicity Level

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. (2010): "Cross-National Time-Series Data Archive," Databanks International, Jerusalem, Israel. http://www.databanksinternational.com/ (accessed March 23, 2012).

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.

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

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.

Maddison, A. (2010): "Statistics on World Population, GDP and Per Capita GDP, 1–2008 AD," Groningen Growth and Development Centre, Groningen, The Netherlands. http://www.ggdc.net/maddison/oriindex.htm (accessed January 2, 2011).

Marshall, M. G. (2010): "Major Episodes of Political Violence (MEPV) and Conflict Regions, 1946–2008," Center for Systemic Peace, Vienna, VA. http://www.systemicpeace.org/inscrdata.html (accessed February 10, 2011).

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. (2017): "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2016.1227711.

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.

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.

Ross, M. L. (2013): "Oil and Gas Data, 1932-2011," https://hdl.handle.net/1902.1/20369.

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.

Themnér, L. and P. Wallensteen (2012): "Armed Conflicts, 1946–2011," *Journal of Peace Research*, 49, 565–575.

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

Wang, S., C. M. L. Jr., M. Jakobsson, S. Ramachandran, N. Ray, G. Bedoya, W. Rojas, M. V. Parra, J. A. Molina, C. Gallo, et al. (2007): "Genetic Variation and Population Structure in Native Americans," *PLoS Genetics*, 3, 2049–2067.

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

World Values Survey, T. (2006): "European and World Values Surveys, Four-Wave Integrated Data File, 1981–2004, version 20060423," The World Values Survey Association, Stockholm, Sweden. http://www.worldvaluessurvey.org (accessed August 24, 2010).

——— (2009): "World Values Survey, 1981–2008 Official Aggregate, version 20090914," The World Values Survey Association, Stockholm, Sweden. http://www.worldvaluessurvey.org (accessed August 24, 2010).

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.1, however, we establish that population diversity remains a significant predictor of civil conflicts when we augment our baseline set of covariates in Table 1 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.2, 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 1 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.3, we check the robustness of our findings from Table 1 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.4, we examine the sensitivity of our baseline findings from Table 1 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.5, 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 from Ethnologue on the spatial distribution of language homelands in combination with gridded population data, and they enter our regressions in Table SA.5 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.4–SA.5 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.6, we exploit the initial or year-1960 values of the time-dependent baseline controls employed by our analysis in Table 1 (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 1 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.7 therefore reports, for our key specifications from Table 1, 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 1. 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.8 checks the qualitative robustness of the results associated with our fully specified empirical models in Columns 8 and 12 of Table 1, 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.1: 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 1. 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 1. The reader is therefore referred to Table 1 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.2: 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	0.046	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 1. 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 1. The reader is therefore referred to Table 1 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.3: 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	0.048	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 1. 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 1. The reader is therefore referred to Table 1 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.4: 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	0.040	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 1. 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 1. The reader is therefore referred to Table 1 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.5: 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***	0.324***	0.286**	0.556***	0.598***	0.543***	0.562***
	(0.066)	(0.103)	(0.115)	(0.115)	(0.121)	(0.191)	(0.207)	(0.176)	(0.177)
Linguistic fractionalization (georeferenced)				0.005	0.001		0.008		0.003
				(0.009)	(0.009)		(0.012)		(0.008)
Linguistic polarization (georeferenced)				0.004	0.002		0.005		0.003
				(0.011)	(0.009)		(0.010)		(0.008)
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	0.040	0.047	0.070	0.088		
Adjusted R^2	0.030	0.188	0.214	0.206	0.356	0.226	0.391		
Effect of 10th–90th %ile move in diversity	0.014***	0.029***	0.020***	0.021***	0.019**	0.027***	0.026***	0.035***	0.038***
	(0.004)	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)	(0.009)	(0.011)	(0.012)
First-stage F statistic								157.089	99.924

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 1. Specifically, it establishes robustness to accounting for the potentially confounding influence of linguistic fractionalization and polarization, constructed using georeferenced information from Ethnologue on the spatial distribution of language homelands 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 1. The reader is therefore referred to Table 1 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.6: 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	0.049	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 1. 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 1. The reader is therefore referred to Table 1 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.7: 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 1. 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 1, 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 1. The reader is therefore referred to Table 1 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.8: 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 institutions	×	×	×	×	×	×	×	×	×	×
Controls for ethnic diversity	×	×	×	×	×	×	×	×	×	×
Controls for oil, population, and income	×	×	×	×	×	×	×	×	×	×
Observations	147	147	105	105	131	131	132	132	126	126
Partial R^2 of population diversity	0.087		0.093		0.099		0.062		0.056	
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***	0.018**	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 1. 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 1. The reader is therefore referred to Table 1 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.9: 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)
Controls for geography		×	×		×	×		×	×
Continent dummies			×			×			×
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.10, 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 2 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.11 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 2 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.12, we check the robustness of our findings from Table 2 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 but statistically insignificant coefficient. 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.13 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 2, 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 a positive and significant coefficient. 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.14 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 2, 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.14 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.15, 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 2 and in odd-numbered columns of Table SA.14, 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 or, in the specifications for conflict incidence, noticeably smaller in magnitude than their classically estimated counterparts in Tables 2 and SA.14 that do not admit spatiotemporal dependence across country-time observations.

Robustness to Accounting for Alternative Correlates of Conflict Onset In Table SA.16, 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 6 and 8 of Table 2, 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, the additional control variables considered by the analysis in Table SA.16 generally tend to enter our specifications with positive but statistically insignificant coefficients. 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.17 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 2, 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.17, 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.10: 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***	6.179*	6.061*	7.057**	9.506**
	(4.264)	(4.488)	(4.387)	(5.167)	(3.301)	(3.531)	(3.421)	(4.277)
Ecological fractionalization	−0.368	−0.080	−0.503	−0.394	0.022	−0.399	−0.023	−0.430
	(0.456)	(0.524)	(0.432)	(0.494)	(0.274)	(0.370)	(0.275)	(0.376)
Ecological polarization	0.865**	0.327	1.086***	0.927**	0.234	0.329	0.402	0.527
	(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	0.097	0.001	0.111	0.088
	(0.178)	(0.166)	(0.167)	(0.176)	(0.124)	(0.138)	(0.122)	(0.140)
Lagged volatility of temperature	−0.576*	−0.416	−0.356	−0.274	0.305	0.246	0.218	0.238
	(0.342)	(0.382)	(0.307)	(0.332)	(0.287)	(0.281)	(0.272)	(0.263)
Lagged volatility of precipitation	−1.326	−1.363	−0.504	−0.439	−0.284	−0.153	−0.568	−0.222
	(0.814)	(1.096)	(0.603)	(0.742)	(0.594)	(0.709)	(0.596)	(0.648)
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	0.162		
Marginal effect of diversity	2.675***	1.873**	3.364***	2.981***	0.323*	0.316*	0.332*	0.456*
	(0.796)	(0.833)	(0.908)	(1.046)	(0.177)	(0.186)	(0.173)	(0.233)

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 2. 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 2. The reader is therefore referred to Table 2 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.11: 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*	4.803*	8.543**	11.651***
	(4.670)	(4.781)	(5.404)	(6.386)	(2.937)	(2.779)	(3.652)	(4.247)
Log years since Neolithic Revolution	0.085	0.187	−0.290	−0.243	0.333**	0.323*	0.030	−0.159
	(0.270)	(0.296)	(0.285)	(0.334)	(0.147)	(0.173)	(0.194)	(0.232)
Log index of state antiquity	0.244***	0.076	0.286***	0.143	0.093**	0.035	0.125**	0.095
	(0.088)	(0.103)	(0.101)	(0.116)	(0.041)	(0.056)	(0.051)	(0.070)
Log duration of human settlement	0.000	0.070	−0.024	−0.009	0.038	0.043	0.003	0.018
	(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	0.048	0.050	−0.004	−0.018
	(0.052)	(0.051)	(0.040)	(0.047)	(0.031)	(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			0.142	0.151		
Marginal effect of diversity	2.992***	1.901**	3.885***	3.105**	0.293*	0.264*	0.436**	0.603**
	(0.896)	(0.936)	(1.140)	(1.333)	(0.165)	(0.154)	(0.203)	(0.256)

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 2. 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 2. The reader is therefore referred to Table 2 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.12: 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***	6.680**	6.836**	7.880***	9.104***
	(3.867)	(3.801)	(3.782)	(4.046)	(2.857)	(2.944)	(2.967)	(3.364)
Ethnic inequality in luminosity	0.593	0.675	0.331	0.277	0.329	0.329	0.263	0.142
	(0.372)	(0.451)	(0.376)	(0.445)	(0.260)	(0.262)	(0.257)	(0.255)
Spatial inequality in luminosity	−0.035	0.150	0.294	0.519	−0.052	−0.014	0.069	0.087
	(0.409)	(0.425)	(0.392)	(0.410)	(0.256)	(0.259)	(0.246)	(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	0.171		
Marginal effect of diversity	2.838***	2.626***	3.272***	3.094***	0.348**	0.348**	0.369**	0.431**
	(0.717)	(0.702)	(0.787)	(0.843)	(0.153)	(0.153)	(0.152)	(0.181)

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 2. 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 2. The reader is therefore referred to Table 2 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.13: 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)

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 2. 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 2, 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 2. The reader is therefore referred to Table 2 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.14: 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) Ordinary Logit	(2) Rare-Event Logit	(3) Ordinary Logit	(4) Rare-Event Logit	(5) Ordinary Logit	(6) Rare-Event Logit	(7) Ordinary Logit	(8) Rare-Event 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***	13.855**	13.406**	13.275**	12.537*
	(6.653)	(6.529)	(6.703)	(6.479)	(6.262)	(6.173)	(6.562)	(6.495)
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		0.132		0.164	
Marginal effect of diversity	3.733***	3.949***	2.992***	3.178***	0.192**	0.201**	0.158**	0.162
	(1.009)	(1.159)	(0.937)	(1.123)	(0.087)	(0.102)	(0.081)	(0.104)

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 2. 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 2. The reader is therefore referred to Table 2 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.15: 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***	6.167**	13.855**	6.380*	13.275*
	(2.616)	(4.261)	(3.381)	(6.025)	(2.904)	(6.520)	(3.475)	(7.361)
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	0.132	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 2 and in odd-numbered columns of Table SA.14. 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 2 and in odd-numbered columns of Table SA.14. The reader is therefore referred to Table 2 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.16: 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) IV Probit	(4) IV Probit
	Annual PRIO25 civil conflict onset, 1960–1999			
Population diversity (ancestry adjusted)	7.364*	6.902*	9.983**	9.196**
	(3.926)	(4.014)	(4.236)	(4.302)
Ethnic dominance		0.086		0.127
		(0.130)		(0.119)
Political instability, lagged		0.172		0.115
		(0.127)		(0.122)
New state dummy, lagged		0.204		−0.156
		(0.536)		(0.530)
Controls for all baseline covariates	×	×	×	×
Observations	2,539	2,539	3,345	3,345
Countries	95	95	116	116
Pseudo R^2	0.146	0.148		
Marginal effect of diversity	0.442*	0.413*	0.565**	0.513*
	(0.243)	(0.247)	(0.273)	(0.269)

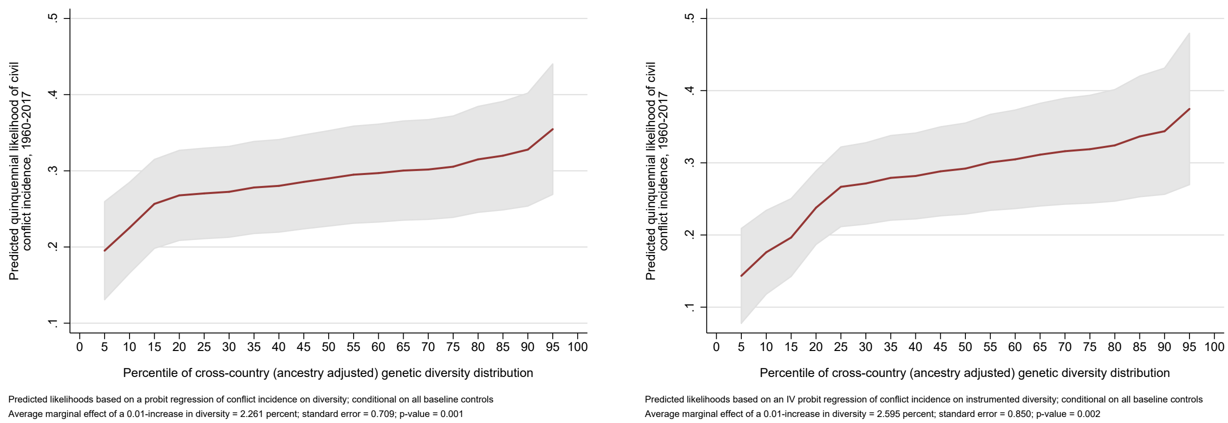
Notes: This table conducts a robustness check on the results from the baseline analysis of the reduced-form impact of contemporary population diversity on the annual onset of civil conflict in repeated cross-country data, as shown in Columns 6 and 8 of Table 2. Specifically, it establishes robustness 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). In light of constraints imposed by the availability of data on these additional control variables, 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 fully specified baseline models of conflict onset, as reported in Columns 6 and 8 of Table 2. The reader is therefore referred to Table 2 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.17: 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)	7.952**	8.959**	7.960**	8.739**	9.044***	10.652**	9.108***	10.584**
	(3.622)	(3.887)	(3.557)	(3.892)	(3.426)	(4.531)	(3.412)	(4.564)
Aggregate price shock	−0.152***	−0.159***			−0.180***	−0.190***		
	(0.056)	(0.059)			(0.053)	(0.056)		
Aggregate price shock, lagged	0.018	0.021			0.014	0.016		
	(0.062)	(0.069)			(0.057)	(0.062)		
Aggregate price shock, twice lagged	−0.165***	−0.179***			−0.117*	−0.121*		
	(0.062)	(0.066)			(0.060)	(0.064)		
Annual crop price shock			−0.181**	−0.191**			−0.208***	−0.223***
			(0.080)	(0.083)			(0.071)	(0.075)
Annual crop price shock, lagged			−0.052	−0.048			−0.048	−0.045
			(0.086)	(0.093)			(0.081)	(0.087)
Annual crop price shock, twice lagged			−0.160*	−0.179*			−0.105	−0.113
			(0.090)	(0.094)			(0.088)	(0.095)
Perennial crop price shock			−0.150**	−0.144**			−0.157***	−0.154***
			(0.069)	(0.070)			(0.058)	(0.059)
Perennial crop price shock, lagged			0.120**	0.119**			0.094*	0.089*
			(0.052)	(0.054)			(0.048)	(0.050)
Perennial crop price shock, twice lagged			−0.134***	−0.145***			−0.083*	−0.083*
			(0.051)	(0.053)			(0.047)	(0.049)
Extractive crop price shock			−0.238***	−0.247***			−0.261***	−0.275***
			(0.088)	(0.092)			(0.082)	(0.086)
Extractive crop price shock, lagged			0.047	0.055			0.035	0.041
			(0.092)	(0.098)			(0.088)	(0.093)
Extractive crop price shock, twice lagged			−0.324***	−0.332***			−0.264***	−0.264**
			(0.108)	(0.111)			(0.100)	(0.104)
Continent dummies	×	×	×	×	×	×	×	×
Time dummies	×	×	×	×	×	×	×	×
Controls for temporal spillovers	×	×	×	×	×	×	×	×
Controls for geography	×	×	×	×	×	×	×	×
Controls for ethnic diversity		×		×		×		×
Controls for institutions		×		×		×		×
Observations	2,647	2,626	2,647	2,626	3,621	3,599	3,621	3,599
Countries	81	81	81	81	103	103	103	103
Pseudo R^2	0.116	0.150	0.129	0.161				
Marginal effect of diversity	0.490**	0.536**	0.484**	0.517**	0.510**	0.577**	0.508**	0.568**
	(0.229)	(0.242)	(0.223)	(0.240)	(0.222)	(0.280)	(0.220)	(0.280)

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 2. 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 2, 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 2 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

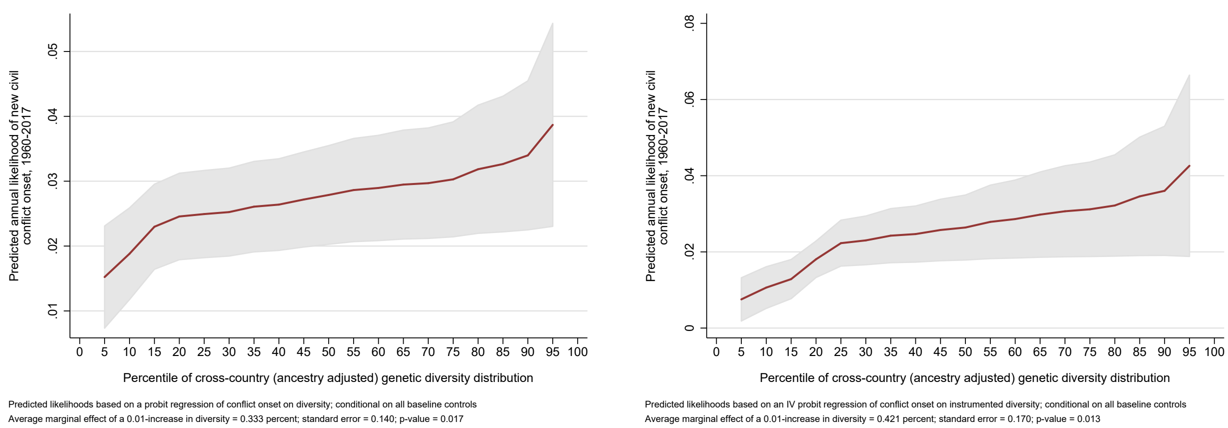


(a) Old-World sample

(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 2. 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.

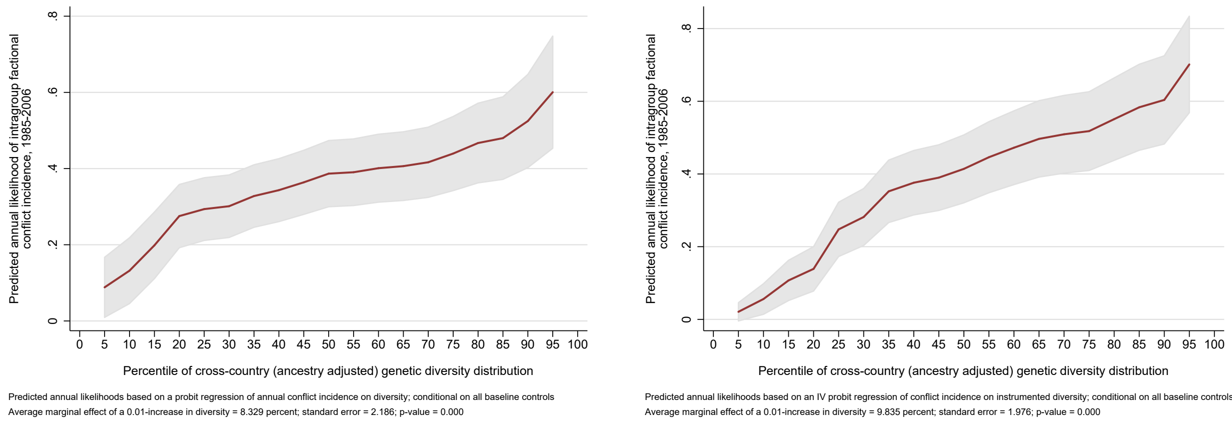


(a) Old-World sample

(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 2. 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 3. 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.

Supplement B Supplement to the Ethnicity-Level Analyses

B.1 Definitions of Main Variables

This section describes the construction of the main variables.

Migratory Distance from East Africa
In estimating the migratory distance from Addis Ababa (East Africa) for each of the ethnic groups in the data, 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). Furthermore, for ethnicities that reside in a distance that exceed 100km from the traversable area connected to Addis Ababa, the distance was computed in the following way. A point set was created by clipping the extended traversable area to world boundaries and aggregating it to a resolution of 2,096,707 pixels which was then converted into points. For each ethnicity centroid, the nearest four distance points were identified. These distances was then added to the migratory distance from Addis Ababa at the distance point to obtain the total migratory distance from the ethnicity centroid from Addis Ababa to each of these four points. The point with the shortest total migratory distance from Addis Ababa was selected to represent the total migratory distance for the ethnicity.

Control Variables
Linguistic fractionalization and polarization Using the Ethnologue map of the spatial distribution of language homelands in combination with the Gridded Population of the World dataset, we estimate the number of individuals living in each intersection between ethnic homelands and language homelands, assuming that population counts in overlapping language homeland areas are split equally between each language homeland-ethnic homeland intersection. Based on these population counts, we calculate the degree of fractionalization using the formula $1 - \sum_i s_i^2$ and the degree of polarization using the formula $4 \sum_i s_i^2(1 - s_i)$, where s_i is the population share of language group i in the homeland.[1]

Other control variables The control variables are based on a range of sources. The data includes a range of geographic variables derived from a number of sources. These geographic variables include elevation, ruggedness, length and density of rivers in the area. Furthermore, the agricultural suitability variables are calculated as the average and standard deviations of the pre-1500 caloric suitability index constructed by Galor and Özak (2016). In addition, the average temperature, and average diurnal temperature range over the period 1901–2012 as constructed by the Climate Research Unit (see Harris et al., 2014).

B.2 Construction of the Georeferenced Ethnicity-Level Dataset

The novel geo-referenced data set of population diversity across ethnic groups is based on several sources. It links the measurements of observed genetic diversity of the 232 ethnic group (as provided by Pemberton et al. (2013)), as well as the measurement of predicted diversity for the

[1] Mapped language areas show the first-language distribution of speakers of languages. Where the data are available, Ethnologue typically use the convention of showing a language in an area if at least 25% of the population in that area speak the language fluently as a first language. In other cases, the language locations have been plotted after research by language surveyors or by other linguists. Ethnologue does not map the locations of immigrant language populations until the language groups are recognised as established in the country and thus receive a full entry in the Ethnologue.

entire set of ethnic groups in the *Ethnographic Atlas* (as constructed in the current paper) to: (i) the geographical area of the historical homelands of these ethnic groups, (ii) the ethnographic characteristics of these ethnic groups, (as reported by the *Ethnographic Atlas* and the *Standard Cross-Cultural Sample*), and (iii) the geographical characteristics of the homelands of these ethnic groups.

The link between population diversity of each ethnic group and the geographical area of the historical homeland of these ethnic groups exploits several sources. Polygons for observations in the *Ethnographic Atlas* is based on Fenske (2013), who linked observations in the *Ethnographic Atlas* to the: (i) polygons found in Murdock (1959), (ii) the Handbook of North American Indians (Heizer, 1978), (iii) Global Mapping International's (GMI) World Language Mapping System, (iv) the Geo-Referencing Ethnic Groups (GREG) map of Weidmann et al. (2010), and (v) data for modern administrative boundaries. We used the link between observations in the *Ethnographic Atlas* and James Fenske's collection of polygons that was implied by the reported centroid coordinates in the data by Fenske (2013).

The matching process of observed population diversity for the 232 ethnic groups in Pemberton et al. (2013) was based on four phases. First, 65 observations from the Pemberton data was merged with name-based matches with the *Ethnographic Atlas* and via that to James Fenske's polygons.[2] Second, the geo-coded points of the ethnic groups reported in Pemberton et al. (2013) was overlaid with the map of James Fenske's polygons and proximate pairs of polygons and points were classified as as either separate, similar, or disparate groups, yielding 84 matches between polygons and points. Third, an additional 97 merges were achieved using a similar method with polygons from the GMI data set and their associated Ethnologue information. Fourth, for some remaining ethnic groups, a plausible polygon could be constructed based on secondary information about the ethnic group.[3]

[2]This matching process required the use of the various names given to each group in different sources.

[3]For instance, Tuscans were merged to the modern region of Toscana, Orcadians were merged to the South Orkney Islands, the Zenú were merged to the Zenú reserve, and the Sengwer were merged to the Embobut Forest area.

B.3 Robustness Checks

Table SB.1: Population Diversity and the Spatiotemporal Prevalence of Conflict across Ethnic Homelands – Robustness to Accounting for Alternative Distances

	Log spatio-temporal prevalence of UCDP GED conflicts, 1989–2008					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS
Genetic diversity (observed)	23.506**	24.277**	23.091**			
	[9.820]	[10.009]	[10.018]			
Genetic diversity (predicted)				75.605***	70.438***	73.260***
				[7.760]	[7.517]	[7.596]
Distance to Technological Frontier in Year 1 (in 1000 kms)	0.237			-0.003		
	[0.156]			[0.057]		
Distance to Technological Frontier in Year 1000 (in 1000 kms)		-0.042			-0.205***	
		[0.167]			[0.064]	
Distance to Technological Frontier in Year 1500 (in 1000 kms)			0.092			-0.103*
			[0.153]			[0.060]
Regional dummies	×	×	×	×	×	×
Geographical controls	×	×	×	×	×	×
Climatic controls	×	×	×	×	×	×
Fragmentation controls	×	×	×	×	×	×
Sample	Observed	Observed	Observed	Predicted	Predicted	Predicted
Observations	230	230	230	1238	1238	1238
Effect of 10th90th %ile move in diversity	0.348**	0.359**	0.342**	1.244***	1.159***	1.206***
meserrest	[0.145]	[0.148]	[0.148]	[0.128]	[0.124]	[0.125]
First-stage F statistic						
Adjusted R^2	0.448	0.442	0.443	0.410	0.416	0.412
β^*	19.488	17.907	19.010	76.644	70.392	75.119

Notes: This table exploits cross-ethnicity variations to establish a significant positive impact of observed and predicted population on the log spatio-temporal prevalence of UCDP/PRIO conflicts 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. Cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

TABLE SB.2: Observed Population Diversity and the Spatiotemporal Prevalence of Conflict across Ethnic Homelands – Robustness to Accounting for Measures of Ecological Diversity

	Log spatio-temporal prevalence of UCDP GED conflicts, 1989–2008						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Genetic diversity (observed)	31.263***	26.042***	22.067**	21.445**	22.253**	24.726**	25.558**
	[10.312]	[10.010]	[10.070]	[10.090]	[10.056]	[10.918]	[10.940]
Ecological diversity	-1.710	-0.593	0.438	0.469	0.424	0.568	0.539
	[1.327]	[1.320]	[1.271]	[1.280]	[1.275]	[1.308]	[1.305]
Ecological polarization	1.290	1.128	0.733	0.676	0.741	0.946	0.980
	[1.060]	[1.103]	[1.061]	[1.065]	[1.061]	[1.101]	[1.095]
Linguistic fractionalization				0.469		0.213	
				[0.539]		[0.542]	
Linguistic polarization					-0.294		-0.360
					[0.496]		[0.519]
Regional dummies	×	×	×	×	×	×	×
Geographical controls		×	×	×	×	×	×
Climatic controls			×	×	×	×	×
Development outcomes						×	×
Disease environment controls						×	×
Sample	Observed	Observed	Observed	Observed	Observed	Observed	Observed
Observations	228	228	228	228	228	228	228
Effect of 10th90th %ile move in diversity	0.470***	0.392***	0.332**	0.323**	0.335**	0.372**	0.384**
meserrest	[0.155]	[0.151]	[0.151]	[0.152]	[0.151]	[0.164]	[0.165]
Adjusted R^2	0.165	0.330	0.441	0.440	0.439	0.462	0.463
β^*		22.776	17.454	16.534	17.738	21.614	22.845

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. Cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent.

Table SB.3: Predicted Population Diversity and the Spatiotemporal Prevalence of Conflict across Ethnic Homelands – Robustness to Accounting for Measures of Ecological Diversity

	Log spatio-temporal prevalence of UCDP GED conflicts, 1989–2008								
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS	(8) 2SLS	(9) 2SLS
Genetic diversity (predicted)	85.669***	85.856***	82.303***	83.984***	82.198***	74.758***			
	[5.194]	[6.404]	[6.815]	[6.717]	[11.719]	[12.482]			
Genetic diversity (observed)							102.549***	95.985***	99.456***
							[31.335]	[26.885]	[27.283]
Ecological diversity	-0.768	-0.186	0.199	0.153	1.111	0.704	0.005	0.587	0.564
	[0.654]	[0.653]	[0.645]	[0.644]	[1.425]	[1.389]	[1.481]	[1.303]	[1.305]
Ecological polarization	0.610	0.308	0.109	0.175	-1.256	-0.945	-0.086	0.158	0.134
	[0.530]	[0.520]	[0.502]	[0.501]	[1.090]	[1.041]	[1.270]	[1.165]	[1.163]
Linguistic fractionalization			0.626**					-0.088	
			[0.278]					[0.535]	
Linguistic polarization				0.207					-0.612
				[0.214]					[0.528]
Regional dummies	×	×	×	×	×	×	×	×	×
Geographical controls		×	×	×	×	×	×	×	×
Climatic controls		×	×	×	×	×	×	×	×
Development outcomes			×	×		×		×	×
Disease environment controls			×	×		×		×	×
Sample	Predicted	Predicted	Predicted	Predicted	Homogenous	Homogenous	Observed	Observed	Observed
Observations	1206	1206	1206	1206	251	251	228	228	228
Effect of 10th90th %ile move in diversity	1.450***	1.453***	1.393***	1.422***	1.294***	1.176***	1.542***	1.444***	1.496***
meserrest	[0.088]	[0.108]	[0.115]	[0.114]	[0.184]	[0.196]	[0.471]	[0.404]	[0.410]
Adjusted R^2	0.369	0.411	0.444	0.441	0.434	0.489			
β^*		86.357	76.793	81.157	76.462	61.207			
First stage coefficient							0.712	0.712	0.712
							(0.195)	(0.195)	(0.195)
First-stage F-statistic							13.325	21.221	20.736

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. Cluster-robust standard errors are reported in square brackets. *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent.

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.

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

BOCKSTETTE, V., A. CHANDA, AND L. PUTTERMAN (2002): "States and Markets: The Advantage of an Early Start," *Journal of Economic Growth*, 7, 347–369.

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.

CERVELLATI, M., U. SUNDE, AND S. VALMORI (2017): "Pathogens, Weather Shocks and Civil Conflicts," *Economic Journal*, 127, 2581–2616.

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.

DESMET, K., I. ORTUÑO-ORTÍN, AND R. WACZIARG (2012): "The Political Economy of Linguistic Cleavages," *Journal of Development Economics*, 97, 322–338.

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. (2013): "Does Land Abundance Explain African Institutions?" *Economic Journal*, 123, 1363–1390.

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

GALOR, O. AND Ö. ÖZAK (2016): "The Agricultural Origins of Time Preference," *American Economic Review*, 106, 3064–3103.

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.

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.

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

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.

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.