Original Manuscript

"Right-Peopling" the State: Nationalism, Historical Legacies, and Ethnic Cleansing in Europe, 1886-2020

Carl Müller-Crepon[1] , Guy Schvitz[2], and Lars-Erik Cederman[3]

Abstract
Many European nation-states were historically homogenized through violent ethnic cleansing. Despite its historical importance, we lack systematic evidence of the conditions under which groups where targeted with cleansing and how it impacted states' ethnic demography. Rising nationalism in the nineteenth century threatened multi-ethnic states with "right-sizing" through secessionism and irredentism. States therefore frequently turned to brutal "right-peopling", in particular where cross-border minorities and those with a history of political independence increased the risk of territorial losses. We test this argument with new spatial, time-variant data on ethnic geography and ethnic cleansing from 1886 to the present. We find that minorities that politically dominated another state and those that have lost political independence were most at risk of ethnic cleansing, especially in times of interstate war. At the macro-level, our results show that ethnic cleansing increased European states' ethnic homogeneity almost as much as border change. Both produced today's nation-states by aligning states and ethnic nations.

[1]Department of Government, London School of Economics and Political Science, UK
[2]Center for Comparative and International Studies, ETH Zürich and European Commission, Joint Research Centre, Italy
[3]Center for Comparative and International Studies, ETH Zürich, Switzerland

Corresponding Author:
Carl Müller-Crepon, Department of Government, London School of Economics and Political Science, Houghton St, London WC2A 2AZ, UK.
Email: c.a.muller-crepon@lse.ac.uk

Data Availability Statement included at the end of the article

Keywords

Contemporary Europe consists of states that are ethnically comparatively homogeneous. Although often taken for granted, Europe's current ethnic geography is the result of a long history of ethnic homogenization that involved extreme levels of violence. Throughout the late nineteenth and early twentieth centuries, European states targeted many ethnic minorities with forced assimilation, resettlements, and mass killings in an effort to homogenize their populations. The practice of "right-peopling" states[1] is not limited to Europe. Recent examples include the genocide of Myanmar's Rohingya, China's forced assimilation of Uyghurs, repeated displacement of Armenians from Nagorno-Karabakh since 2020, and current fears of permanent ethnic cleansing of the Palestinian population in Gaza and parts of the West Bank. Despite the tremendous human costs of such campaigns, it remains unclear under which conditions some minorities become targets of ethnic cleansing while others are spared and to what extent ethnic cleansing has shaped today's societies.

Relying mostly on qualitative case studies, early research explains patterns of ethnic cleansing as the result of a security dilemma (Posen 1993) and internal threats (Harff 2003; Harff and Gurr 1988; Straus 2013; Valentino 2004) or analyzes its macro-historical and ideological roots (Mann 2005; O'Leary 2001). Applying quantitative research methods, more recent studies make important contributions to explaining ethnic cleansing at the group level (e.g., Bulutgil 2015, 2016; McNamee and Zhang 2019; Mylonas 2012) and ethnically targeted one-sided violence (Balcells and Stanton 2021; Fjelde et al. 2021; Fjelde and Hultman 2014). Yet their methodological progress comes at the cost of neglecting macro-historical processes and legacies, as well as spatial dynamics across country-borders.

Adopting a geopolitical perspective, we argue that perceived territorial threats motivated many ethnic cleansing campaigns characterized by mass killings and/or ethnically targeted forced displacement. Following the rise of nationalism in nineteenth century Europe, multi-ethnic polities faced increasing risks of being "right-sized" through secession and irredentism. In response, states increasingly sought to homogenize their populations to pre-empt the loss of territory settled by non-dominant ethnic groups.[2] Violent homogenization efforts concentrated on regions with a high risk of territorial conflict: regions where ethnic groups were divided by state borders and where past border changes invited revisionist nationalism. More specifically, we expect that non-dominant groups with transborder ethnic kin (TEK) and a historical experience of controlling an independent state (past "home rule") were more likely to challenge their host states, in particular where autocratic institutions prevented accommodation. This ultimately made them more likely targets of ethnic cleansing than non-dominant groups without TEK or past home rule.

We test these arguments with our newly collected *Historical Ethnic Geography* dataset that maps Europe's ethnic geography since 1886 based on 73 historical maps of 120 ethnic groups.[3] As a complement, a new list of ethnic cleansing episodes during the same period records 113 cleansing campaigns with a conservatively estimated 56 million victims.

At the level of ethnic groups nested within countries, our analyses show that non-dominant groups with TEK and those with a history of lost home rule were frequent targets of ethnic cleansing by their host states. We find that non-dominant groups with transborder ties to a group dominating another state face a yearly risk of ethnic cleansing that is 180 percent higher than non-TEK groups. Similarly, looking back at 20 years of independent home rule increases groups' risk of ethnic cleansing by 74 percent. These effects are mostly driven by autocratic states and robust to alternative specifications, stringent country-year fixed effects, and a randomization inference test. While the results are not driven solely by ethnic cleansing during the World Wars, we find that TEK links increase the risk of cleansing in particular during times of international warfare between the states a group is part of. This finding further supports our argument that threats of "right-sizing" are a main driver of ethnic cleansing.

Our empirical analysis lastly turns to disentangling the contribution of violent "right-peopling" on the increasing alignment of states and ethnic nations at the European macro-level over the past 160 years. Based on a new, information-theoretic alignment measure, our analysis suggests that more than 40 percent of the overall increase in state-to-nation congruence is due to "right-peopling" that violently changed Europe's ethnic map. This finding has important implications for our understanding of ethnic demography and its socio-political effects.

Literature Review

Following the resurgence of ethnic violence in the 1990s, a broad research community started to explain its occurrence (Korb 2016). Building on international relations theory, Posen's (1993) seminal account explains ethnic cleansing as resulting from a security dilemma which leaves ethnic groups unprotected after the collapse of multi-ethnic states such as the Soviet Union or Yugoslavia. Left to fend for themselves, increased threat perceptions can motivate some groups to strike first to rescue potentially vulnerable co-ethnics in ethnically mixed regions, an escalation that can result in outright ethnic cleansing. A strength of Posen's model is its account for spatial patterns of ethnic violence, often targeted at enclaves. Yet, in focusing entirely on ethnic violence in the wake of state collapse, the inter-group security dilemma says little about the vast majority of modern ethnic cleansing campaigns that was carried out by governments.

Differentiating ethnic cleansing and genocide from more general "ethnic conflict", Mann (2005) takes a macro-historical perspective and argues that the global diffusion of democracy often gave rise to exclusionary ideologies and racial definitions of the *demos*, resulting in forced assimilation, displacement, and outright genocide of out-groups. In fact, this "dark side of democracy" implies that most liberal democracies are

built on a violent history of ethnic cleansing. While singling out nationalist ideology as an important driver of ethnic cleansing, this argument fails to explain why exclusionary nationalism prevailed in some states but not others. Moreover, although macro-historical patterns explain temporal trends, they say little about why some groups became targets of cleansing while others were spared.

Focusing on the latter question, other studies argue that states resort to ethnic cleansing in response to perceived security threats, targeting groups they suspect of collaborating with internal (Harff 2003; Harff and Gurr 1988; Straus 2013; Valentino 2004) or external enemies. Mylonas (2012) importantly shows how states accommodate groups supported by their allies, but tend to exclude, repress, and cleanse groups with ties to rival states. Similarly, Bulutgil (2015, 2016) links ethnic cleansing to external threats (see also Hong and Kim 2019), while highlighting the mitigating effects of cross-cutting class cleavages. Focusing on the border region between China and the USSR, McNamee and Zhang (2019) provide further evidence on ostensibly protective "demographic engineering" (see also Carter 2010; McNamee 2018).

While the literature highlights the strategic logic of ethnic cleansing, it exhibits four shortcomings: First, previous research mostly focuses on explaining ethnic cleansing within existing state borders. However, the restriction to fixed territorial units risks mischaracterizing the link between ethnic cleansing and border-transforming events such as secession and conquest. Second, studying the direct causes of ethnic cleansing can come at the expense of attention to its broader macro-historical context. The occurrence of ethnic cleansing varies over time and is often connected to processes of nation-state formation. Ethnic cleansing should therefore be seen as part of long-term historical developments. Third, many large-N studies focus on country or group-level attributes without much consideration to the effects of spatial configurations. Yet analyses of more fine-grained data from single countries show that local geography plays an important role (e.g., McNamee 2018; McNamee and Zhang 2019). Fourth and finally, the macro-historical transformation of states through violent ethnic cleansing remains understudied. We particularly lack evidence on the impact of ethnic cleansing on the socio-demographic structure of states. To better understand when and where ethnic cleansing occurs and how it impacted European states' demography, it is therefore necessary to implement a historically "deep" large-N research design with meso-level spatial precision similar to single-country studies while covering the whole of Europe.

Theoretical Argument

We seek to explain what triggers state-led ethnic cleansing campaigns.[4] We define ethnic cleansing as the attempt to forcibly and permanently remove members of an ethnic group from a region through violence. Our definition covers two types of ethnic cleansing. *Forced displacement* uproots ethnic groups, typically moving them from their host states' territory to another state. In turn, *ethnic mass killing* refers to efforts to annihilate an ethnic group as a whole or in parts by killing its members.[5] This definition

covers the most violent strategies such as ethnic mass killing and exclusionary politics (Mylonas 2012), yet excludes homogenization policies that operate over a comparatively long time-horizon.

We argue that governments strategically employ ethnic cleansing to establish control over contested territory.[6] Cleansing the territory of a threatening group, states seek to prevent secession or foreign annexation. We focus on two main factors that increase the perceived threat potential of ethnic groups: the presence of trans-border ethnic ties and historical legacies of past home rule.

Nationalism and Territorial Contestation

Although often described as "primitive" or "barbaric", ethnic cleansing is an inherently modern phenomenon (Mann 2005; Ther 2014). Most pre-modern states did not have the capacity to kill or displace entire ethnic groups, nor did they have the motives to do so, as ethnicity was mostly politically irrelevant (O'Leary 2001).

Things changed in the nineteenth century, as nationalism spread across Europe and beyond. In Western Europe, states introduced territorial approaches to citizenship that treated most inhabitants as potential members of the nation. In contrast, most aspiring nations in Central and Eastern Europe adopted "organic" brands of nationalism that viewed nationality as ethnically pre-defined (Mann 2005). In these cases, the "ethnos" rather than the "demos" grounded demands to realize Gellner's (1983, 1) nationalist principle that "ethnic boundaries should not cut across political ones, and […] that ethnic boundaries within a given state should not separate the power-holders from the rest." Where this principle was violated, nationalist mobilization for self-determination, border change, and the creation of ethnically homogeneous nation-states often followed (Cederman, Girardin, and Müller-Crepon 2023; Müller-Crepon, Schvitz, and Cederman 2023).

The ideological shift towards nationalism therefore represented a fundamental challenge to the existing political order. Most affected were the large and ethnically diverse Ottoman, Habsburg and Russian empires, but many newly established nation-states, such as Greece, Serbia, and Poland, also faced mismatches between political and ethnic borders. Ethnic diversity made effective rule increasingly difficult and posed a security threat. Regions inhabited by non-dominant groups threatened to secede, while neighboring states claimed or attempted to annex territories inhabited by "their" ethnic kin (Weiner 1971). Such tensions were fueled by major powers in an effort to destabilize their rivals (Mylonas 2012).

In this environment, governments became increasingly pre-occupied with homogenizing their populations. In principle, they could reduce geopolitical risks by "right-sizing" their territory, abandoning claims to regions populated by minorities. Given the high value of scarce territory in Europe, however, they were unlikely to do so voluntarily (O'Leary 2001). Instead, many governments opted for "right-peopling" strategies that allowed them to retain their territory.

In contrast to non-violent homogenization efforts (Darden and Mylonas 2016; Weber 1976), some states, we argue, choose forced resettlement and mass killing of ethnic groups as a last-resort, in particular if groups are viewed as an urgent threat to state survival (Cattaruzza 2010; Ther 2014). Ethnic cleansing can remove the nationalist incompatibility altogether or, if not all-encompassing, reduce a groups' capacity for mobilization by fragmenting it (see Schubiger 2022). An ethnic groups' threat potential largely depends on its motives and opportunities for secession, and on whether the it's presence in a region increases the risk of foreign annexation. Both are affected by the presence of trans-border ethnic kin (TEK) and a history of political independence through "home rule".

Trans-border Ethnic Kin

As a violation of nationalist principles, the division of ethnic groups by state borders can motivate resistance against the status quo. Leaders of divided groups commonly portray the group's fragmentation as an injustice, setting the stage for tensions between the group and its host state government. Viewing current borders as illegitimate, divided groups are likely to demand political concessions that may range from regional autonomy to independence or the unification with a neighboring state (Cederman, Rüegger, and Schvitz 2021). In turn, host states are more likely to target such groups with aggressive nation-building policies.

To consider the effect of TEK linkages in greater detail, we distinguish between three configurations of state borders and ethnic settlement areas, as shown in Figure 1. Throughout the discussion, we reserve the term *ethnic groups* for those ethnic communities that exist independently of country borders and refer to *group segments* as those parts of an ethnic group that belong to a given state. For example, the collapse of Austria-Hungary led to Hungarian group segments in Austria, Hungary, Czechoslovakia, Romania and Yugoslavia.
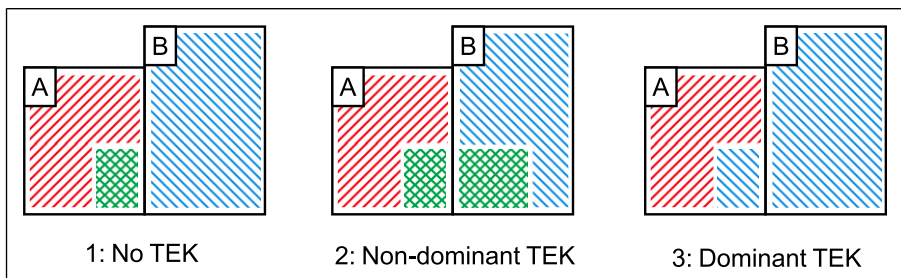

Figure 1. Three ethno-political configurations with with and without transborder ethnic kin (TEK).
Note. Boxes indicate state borders, hatched areas represent ethnic groups.

The first configuration shows a non-dominant group segment in state A without TEK, a situation the Scottish in the United Kingdom find themselves in. The second configuration features a non-dominant group segment in state A with ethnic ties to a non-dominant segment in state B. An example are the Armenians in the Ottoman Empire, which had stateless ethnic kin in both Russia and Iran. Given the contradiction between nationalist principles and a group's current territorial division, trans-border ethnic groups are more susceptible to separatist conflict than groups without TEK linkages (Cederman, Rüegger, and Schvitz 2021). We posit that states are more likely to view such groups as a security threat, given their opportunity to stage cross-border insurgencies (Salehyan 2007). TEK groups also represent an opportunity for rival states to destabilize their neighbors by stoking ethnic tensions (Mylonas 2012). This situation was feared by the Ottoman government, who aimed to salvage their rule over Anatolia and counter the threat of Russian invasion and Armenian independence through genocide in 1915 (Akçam 2012). We thus expect that

Hypothesis 1: *Non-dominant group segments with non-dominant TEK are more likely to become targets of ethnic cleansing than non-dominant group segments without TEK links.*

The third configuration in Figure 1 shows a non-dominant segment in state A with ethnic ties to the dominant group in state B. Adding to risks of secession and foreign interference, this third configuration also increases the risk of annexation. The existence of a kin state and the unrealized potential of national unity can inspire irredentist claims on both sides of the border. Leaders in target state A may view the non-dominant group as a "fifth column" that poses a security threat (Mylonas and Radnitz 2022; Weiner 1971). Even in the absence of open conflict or territorial disputes, the risk of instability may prompt states to pre-emptively resettle "stranded" groups to their homeland state across the border. The existence of a homeland state also creates an opportunity to negotiate formal population exchange agreements, which were long seen as acceptable on the international stage (Ther 2014). The 1923 population exchange of more than 1.5 million people between the late Ottoman empire and Greece exemplifies this logic. In particular, the Ottoman government feared Greek irredentism, while Greek nationalists eyed material gain and a "modern" homogeneous Greek nation-state (Shields 2013). This motivates the following: expectation:

Hypothesis 2: *Non-dominant group segments with dominant TEK in neighboring states are more likely to become targets of ethnic cleansing than other non-dominant group segments.*

The dynamic underlying this effect might be weakened if (particularly large) states can deter their neighbors from violently targeting its ethnic kin (Van Houten 1998). While such deterrence may have a pacifying effect in normal times, it is unlikely to work once a TEK state has raised territorial claims or is even engaged in active war with the segments' host state.

Past Home Rule

In addition to a group's current territorial division, historical legacies play a decisive role in shaping the risk of territorial conflict. Most territorial disputes are rooted in claims of historical ownership (Carter 2017). Such demands are widely seen as more legitimate than other types of claims, and hence are more likely to attract domestic and international support (Murphy 1990). Even where just used as a pretext, historical precedents can still create opportunities for revisionism, for example through their continued existence as subnational administrative units which facilitate secessionist mobilization (Griffiths 2016) and through their lasting effects on the local social fabric (e.g. Abramson, Carter, and Ying 2022), which makes reinstating old borders more feasible than drawing new ones (Abramson and Carter 2016).

Historical border change motivates ethnic secession and irredentism, especially if such changes entailed a loss of political power for an ethnic segment. Three types of border change constitute such a loss of home rule, each also impacting groups' transborder ethnic ties (Figure 2). First, border change from secession or conquest can separate a segment from its surviving home state. This was the case of Muslim and German populations stranded outside the remains of the collapsed Ottoman empire and Nazi Germany, respectively. Large parts of both groups were forcibly displaced with the goal of "repatriatiation" and prevention of future conflict (İçduygu and Sert 2015; Snyder 2011, ch. 10). Second, ethnic groups can lose their home rule and through foreign annexation of their entire home state. This was the fate of Estonians, Lithuanians, and Latvians who lost their independence to the USSR in 1940, followed by wide-spread deportations. Third, some instances of conquest and annexation split a group across several states. Poland's partition at the end of the eighteenth century split its territory and population between three empires.

Groups with a history of independent statehood likely have stronger national identities, which can be mobilized through backward-looking myths of the group's glorious past and the trauma of status decline. The loss of autonomy thus creates powerful motives for group segments to push for revisionist border changes (Germann and Sambanis 2021; Hechter 2000; Siroky and Cuffe 2015) and threaten their host states' territorial integrity. In response, states are likely to target such groups with increasingly violent nation-building efforts:

> **Hypothesis 3:** *Non-dominant group segments with a history of past home rule are more likely to become targets of ethnic cleansing than other non-dominant group segments.*

Figure 2 clearly shows an inherent connection between the type of border change that led to the loss of segments' past home rule and the presence and type of their transborder ethnic ties. This raises the question whether transborder ties to a dominant group and past home rule have cumulative or substitutive effects on the risk of ethnic cleansing: on the one hand, past home rule could increase the risk of territorial change
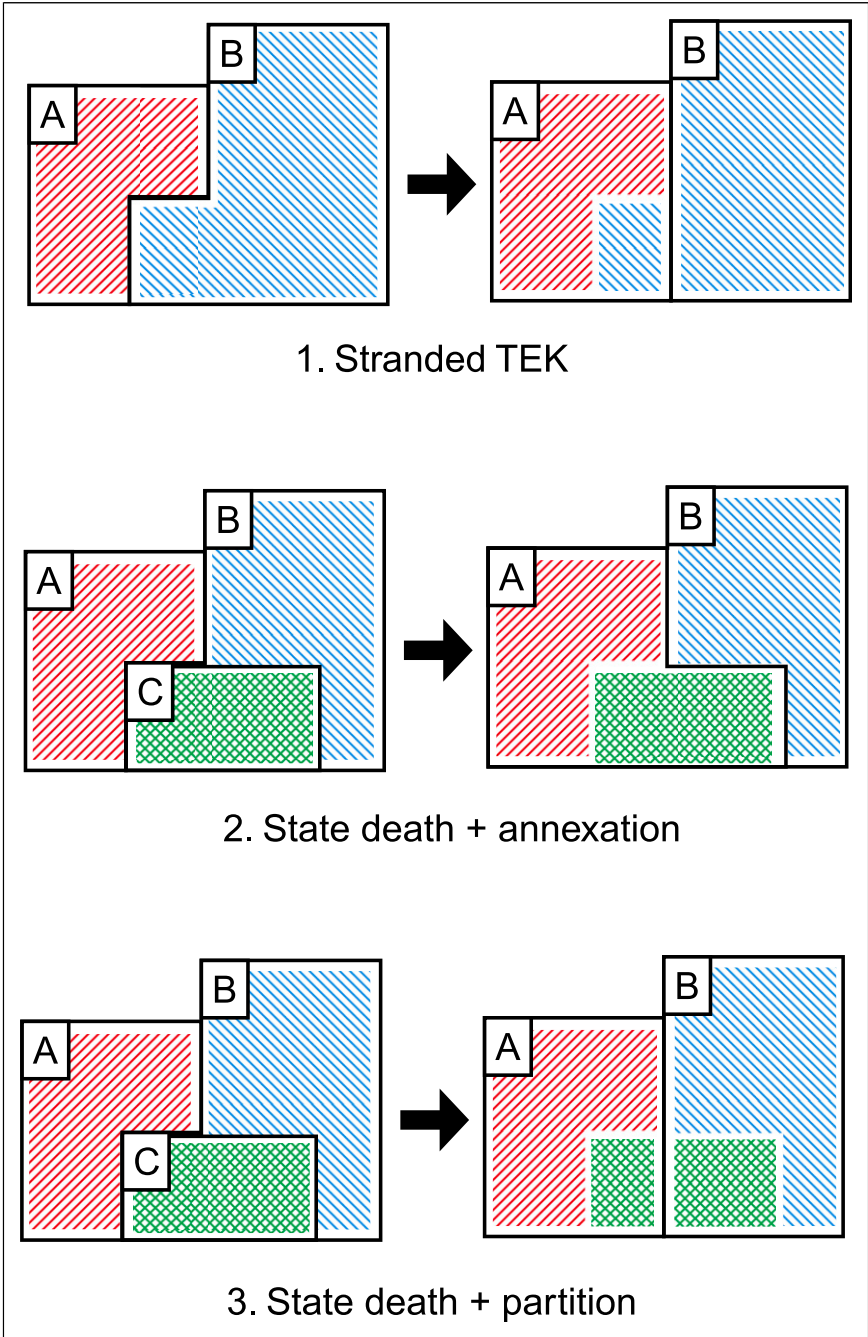
Figure 2. Three scenarios where border changes entail a loss of "home rule".
Note. Boxes indicate state borders (before and after), hatched areas represent ethnic groups.

and reactive ethnic cleansing among segments with dominant TEK, indicating a cumulative effect. Yet, this effect may be substituted for by the effect of the dominant TEK group and small as compared to segments with no or non-dominant TEK. For the latter, the example of home rule in the past may in turn substitute for the absence of an example of home rule in the present. We empirically investigate this issue below.

Ethnic Cleansing and the Making of Homogeneous Nation-States

The effects of ethnic cleansing on the macro-level follow directly from its proposed geopolitical origins. If states' are successful in addressing the risk of secessionism and irredentism through violent right-peopling of their populations, the targeted ethnic segment will shrink in size, potentially to the point of complete annihilation from a given state territory and population.

While ethnic cleansing campaigns are comparatively rare events, they have shaped Europe's ethnic demography through their sheer historical magnitude. Few demographic traces remind us of past ethnic diversity in Eastern Europe, the "Bloodlands"[7] where the mass murders committed by Hitler's and Stalin's regimes killed approximately 14 million civilians only between 1933 and 1945 and displaced many more (Snyder 2011). Similarly, ethnic demography in the Balkans, in Greece and Turkey has been violently influenced by, for example, the Armenian Genocide (1915-197) or population exchanges between Turkey and Greece after 1923, which followed genocide, targeted killings, and mass-displacement in the late Ottoman Empire. In sum, we argue that, at the macro-level:

> **Hypothesis 4:** *Ethnic cleansing significantly contributed to the ethnic homogenization of states' populations.*

While we claim that our argument about the roots of ethnic cleansing in nationalist territorial competition captures important historical dynamics, its applicability is likely restricted by a number of influential scope conditions. Most prominently, these consist in the absence of a bundle of liberal and democratic norms that have led to ever-stronger norms of territorial integrity (Zacher 2001), reduced the likelihood of ethnic conflict by enabling power-sharing and accommodation (Cederman, Gleditsch, and Wucherpfennig 2017; Gurr 2000), and prevented interstate war (e.g. Imai and Lo 2021).

The New Historical Ethnic Geography Dataset

Testing our argument about the effects of transborder ethnic kin and historical precedents on ethnic cleansing and its impact on states' demography requires spatially disaggregated and complete data on Europe's ethnic geography since the nineteenth century.[8] Unfortunately, existing data on ethnic geography, such as the Atlas Narodov

Mira (Bruk and Apenchenko 1964) and GeoEPR Wucherpfennig et al. (2011), are time-invariant and date from after World War II and most ethnic cleansing campaigns.

We fill this gap by collecting, digitizing, and standardizing 73 historical ethnic maps of Europe. Coupled with hand-coded data on violent and peaceful periods of ethno-demographic change, the resulting *Historical Ethnic Geography* (HEG) dataset constitutes the foundation of our analysis. The data map ethnic geography in Europe—defined expansively to include the Caucasus, the Levant, and Northern Africa —from 1886 to 2020 using time-variant rasters that provide estimates of the ethnic composition of local populations. Compared to prior polygon-based data (e.g., Weidmann, Rød and Cederman 2010; WLMS 2006; Wucherpfennig et al. 2011), HEG efficiently combines information across multiple maps, captures local ethnic diversity, and avoids imposing arbitrary population thresholds. In addition and important for the present purposes, the data are time-variant, based on historical information, and independent of changing state borders. Because the data is based on historical maps, it only captures spatially broad patterns of ethnic geography rather than local ethnic diversity resulting from individual-level migration.

Historical Ethnic Map Collection. Our data collection started with identifying and scanning all potentially relevant ethnic maps from the digital catalogues of major libraries such as the Library of Congress, the British Library, and the Bibliothèque nationale de France.[9] From this collection, we selected a total 73 ethnic maps that had (1) a high resolution, (2) broad spatial coverage, (3) authors of varying nationality, and (4) no obvious political biases. For each map, we digitized all groups they depict including their respective group labels. Figure 3 depicts the distribution of maps across time.[10]

Standardizing Ethnic Maps. To combine our maps, we next standardize all ethnic groups they depict. While almost all of them are linguistically defined, the data are enlisted at
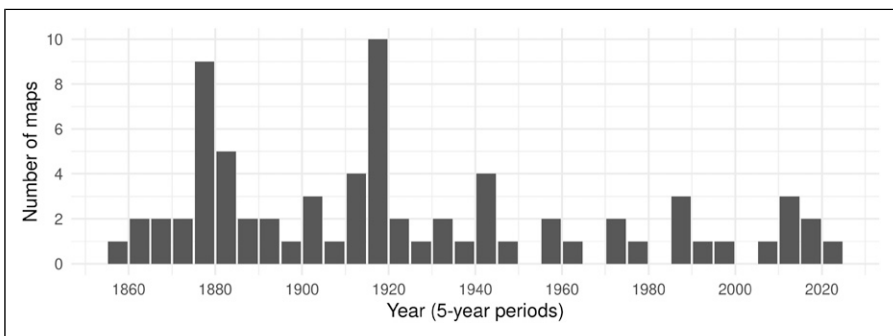

Figure 3. Digitized maps by year of creation.
Note. Binned into 5 year periods.

differing levels of granularity. We follow Müller-Crepon, Pengl, and Bormann (2020) and match all ethnic labels to the tree of known languages compiled by Ethnologue (Lewis 2009). We then applying a majority "voting rule," using those language tree nodes that appear on the majority of maps that depict a given tree branch.[11]

Coding Episodes of Rapid Ethnic Change. To fill the temporal gaps between maps, we define episodes within which we assume certain maps to be valid. The starting points of episodes are determined by periods of large-scale ethnic change which likely changed a group's settlement area in a state. Specifically, we consult the secondary literature for each group to identify instances of forced resettlement, genocide, and less-violent cases of mass migration. We record the state(s) in which each change occurred, the actor(s) responsible, and the approximate size of the affected population. Because these data are less complete for smaller events, we drop those that have likely affected less than 1'000 individuals.

Figure 4(b) shows the episodes of rapid ethnic change for the case of Polish-speaking populations. The episodes determine the spatio-temporal validity of our maps, assuming that data on a group is valid until the group is subject to change in the respective state territory. While this approach avoids potentially fraught interpolations across periods of rapid change, it demands many and temporally granular maps in cases of repeated ethnic change. Where such data is not available, we use maps from preceding periods where available.

Constructing Gridded Ethnic Geography Data. Finally, we convert the standardized and periodized ethnic maps into spatio-temporal raster data with a resolution of 0.0833 decimal degrees. For each raster cell, group, and period, we aggregate the information across maps by calculating the share of maps that show the group to be present in a cell. Overlapping settlement areas are discounted accordingly.[12] Figure 4(c) visualizes the resulting data on the Polish in 1918 and 1951. All group shares in a point add to unity and proxy cells' local ethnic composition. Figure 5 maps the full HEG data for the year 1890 and 2020, showing much ethnic change in Central and Eastern Europe.

A variant of the rasterization procedure produces yearly rasters that interpolate the ethno-demographic information derived from the raw data across time[13] while still respecting the sharp periodization. This improves data quality where ethnic change through assimilation occurs over slowly over decades. For example, many groups such as the French in Western Europe have not experienced ethnic cleansing but have changed nevertheless. In contrast to the baseline approach, the yearly raster give more weight to maps closer to the year of observation, thus capturing slowly changing ethnic geography.

Validation. Our data validation relies on three comparisons. First, we analyze the face validity of our maps by computing the extent to which their depictions of ethnic groups overlap. We find that pairs of maps that are drawn with a time-difference of less than
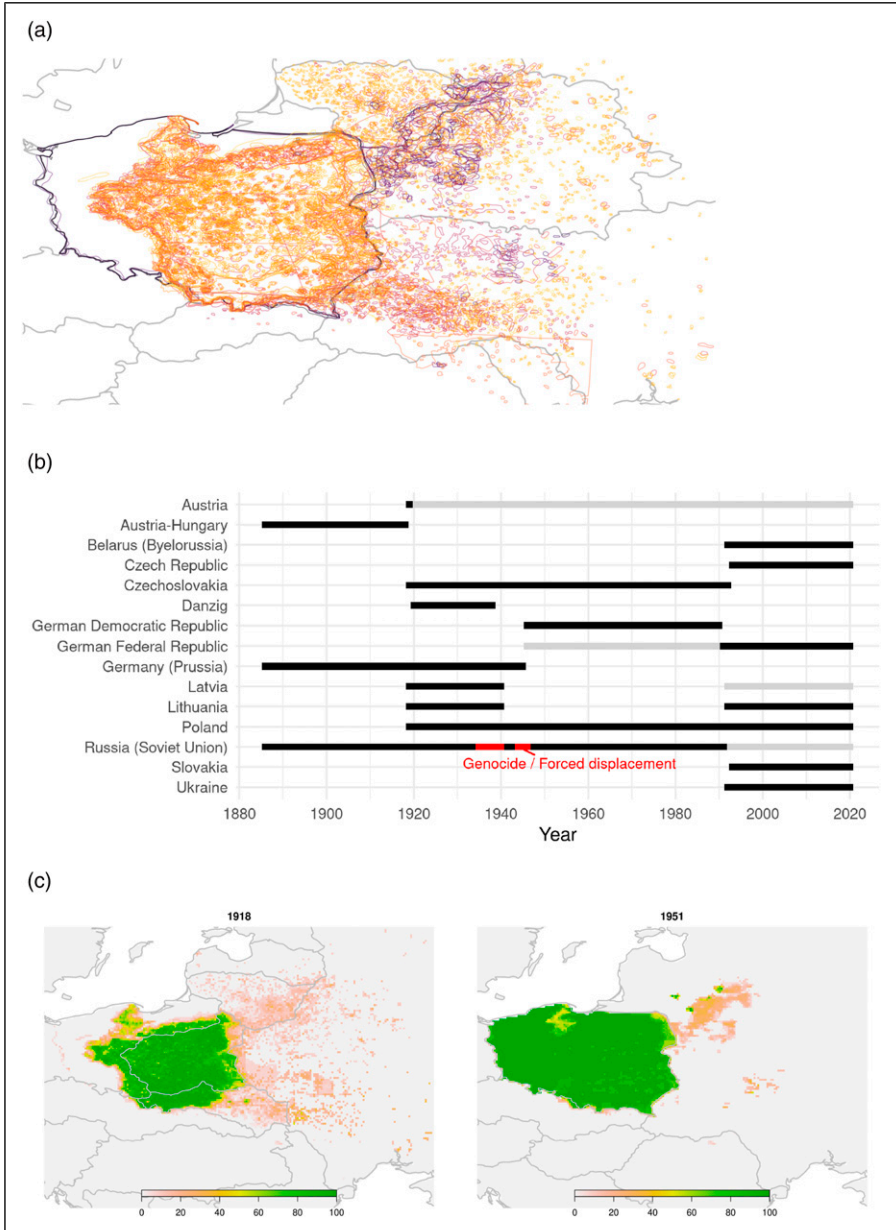
Figure 4. HEG data construction for the Polish people. (a) Digitized maps of ethnic Polish. (b) Overview of Polish people across countries. (c) Map of Polish People in 1918 and 1951 as raster data.
Notes. (a) Lighter colors denote earlier years. State borders in light grey.
(b) Black bars denote existence of Poles in a state. Grey bars denote that state without the group. Episodes of ethnic cleansing are marked in red.
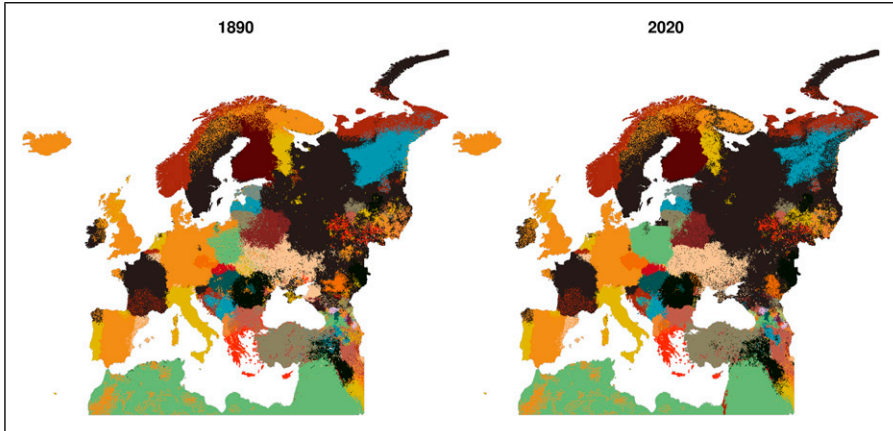(c) Colored areas indicate the estimated population share in each cell.

Figure 5. Full HEG data in 1890 and 2020.
Note. Each ethnic group is colored differently from its neighboring groups. For each pixel, one group is sampled from the ethnic composition in the HEG data.

25 year (including across periods of rapid ethnic change) have, on average, an 85 percent overlap in their depiction of the same ethnic group.[14] As an expected result of ethno-demographic change, overlaps decrease with growing time-differences between maps.

Second, we compare groups' country-level population shares derived from the 1990 HEG data with Fearon's (2003) data on ethnic groups. The aggregated HEG data explain 93 percent of the variation in Fearon's data. Third, we gauge the subnational validity of our data compared to census data of the Austrian-Hungarian empire in 1910. The HEG data explain 87 percent of the variation in the shares of the nine largest ethnic groups across 450 districts. Together, these results indicate that our approach yields valid data on ethnicity at the local and national levels.

Empirical Strategy

Our analysis proceeds with a test of Hypotheses 1-3 conducted at the level of ethnic group segments. After presenting the research design and results, we return to the macro-level and measure the impact of ethnic cleansing on the homogenization of European states.

Main Data

Unit of Analysis. Our main unit of analysis is the segment s of ethnic group e present in country c at time t between 1886 and 2020. Segments are derived by intersecting the HEG raster data for year t with the respective set of state borders retrieved from the

CShapes 2.0 dataset (Schvitz et al. 2022).[15] The resulting dataset contains 39'003 group segment-years across 6'125 country-years and 120 ethnic groups.

We systematically assign dominant group status to group segments that have the largest population share in a state's capital, resolving conflicting cases by recurring to secondary sources. Our analysis focuses only on non-dominant ethnic group segments, since dominant groups are theoretically unlikely and have not been empirically observed to be cleansed by states that are governed by their respective co-ethnics.

Ethnic Cleansing. Similar to Bulutgil (2015, 2016), we take as our main outcome of interest the onset of an episode of ethnic cleansing through mass killings and/or forced displacement (e.g., Garrity 2022) executed by the government of the host state of an ethnic segment since 1886. We differ from Bulutgil's (2015; 2016) operationalization mainly by using an absolute threshold of 1'000 victims which reduces data requirements as compared to her relative criterion of 20 percent of groups' population. We use our data on ethnic change presented above to retrieve this information and code all post-onset years during an episode of ethnic cleansing as missing.

Our final dataset includes 113 onsets of ethnic cleansing with more than 1'000 victims carried out by host state governments, equivalent to an onset in 0.34 percent segment-years.[16] The overall number of victims of ethnic cleansing campaigns is extremely difficult to gauge, as definitions of victimhood are contested, historical sources at times unreliable, and secondary studies not always conclusive. Drawing on estimates from the secondary literature on the number of killed or displaced civilian individuals during each campaign, our (imprecise) estimate of the victims of state-led ethnic cleansing since 1886 amounts to a staggering 56 million individuals[17] or 28 percent of the population of the affected ethnic group segments (198 million).[18] A back-of-the-envelope calculation indicates that individual Europeans' risk of becoming a victim of ethnic cleansing at any point in their life was non-trivial since 1886, amounting to roughly 3 percent.[19]

Figure 6 shows that ethnic cleansing campaigns are mostly concentrated in the first half of the twentieth century, in particular during the violent reign of Hitler's Germany and Stalin's Soviet Union over large parts of the continent, as well as during the aftermath of World War II. But minorities are still at risk today, as in the Balkans in the 1990s, in Azerbaijan in 2020, and in the Russian-occupied territories of Ukraine since 2014.

Main Independent Variables. We construct two independent variables to test our main arguments. First, the TEK status of each ethnic segment captures whether, in a given year, it has (1) no transborder ethnic kin (TEK, ca. 23 percent), (2) only TEK without dominant status (ca. 41 percent), or (3) at least one dominant TEK group (ca. 36 percent). These categories are mutually exclusive. We assign a TEK status to all groups located in more than one state at time *t*. Group segments are assigned the dominant status in a country if they make up a majority of the population in the capital.
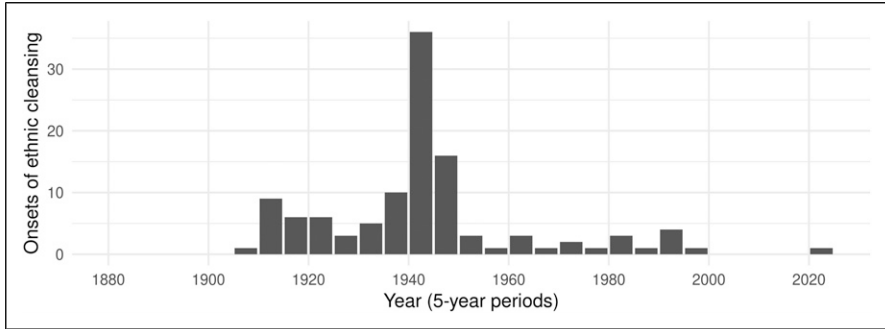
Figure 6. Onsets of ethnic cleansing by year.
Note. Binned into 5 year periods.

Second, the geocoded historical state borders after 1816[20] enable us to trace each group segment's recent history of past home rule. In particular, we compute the number of years since 1816 in which the average inhabitant of a group segment's settlement area at time t belonged to a state in which the segment's ethnic group had dominant status.[21] The larger fraction of a group's settlement area has been under rule of a co-ethnic state for longer time, the higher our indicator of past home rule. On average, 12 percent of the non-dominant segments in our data have a history of any past home rule since 1816. Of those with a history, the median number of home rule years is 18 years, and the mean 44 years. We log-transform the variable in our analysis to account for this right skew.

Control Variables. We use the HEG raster data on ethnic segments in combination with various other geographic datasets to measure a series of factors that may affect our main independent variables and the likelihood of ethnic cleansing. Unless otherwise noted, these control variables are population weighted averages across each groups' settlement area.

For each segment, we first measure the log-transformed population size as larger segments may be more likely to become targets of ethnic cleansing and more often have TEK as well as past home rule. In addition, we control for the population size of the country and the entire ethnic group a segment belongs to.

Second, we account for segments' average distance to their host state's capital, since peripheral segments are more likely to have TEK and may be at a higher risk of ethnic cleansing. In a similar vein, we measure segments' geography as their average altitude, ruggedness, temperature, precipitation, evaporation, and the ratio of the latter two.[22]

Estimation Strategy

We use these data to estimate the effect of TEK status and past home rule on the onset of ethnic cleansing in an OLS fixed effects setup:[23]

$$\text{onset}_{g,s,c,t} = \gamma_c + \tau_t + \beta_1 \text{TEK status}_{s,t} + \beta_2 \text{past home rule}_{s,t} + \\ \beta_3 \text{TEK status}_{s,t} \times \text{past home rule}_{s,t} + \boldsymbol{X}_{s,g,c,t} + \epsilon_{s,g,c,t}, \tag{1}$$

where the risk of an onset of ethnic cleansing in a segment-year is modelled as the sum of a country-specific (γ_c) and year-specific (τ_t) risk, the effects associated with our control variables X, and segments' TEK status, past home rule and their interaction. We introduce these main treatment variables one-by-one.

As foreshadowed in the theoretical argument, we model their interaction in the last step to account for the close connection between TEK status and past home rule and test for the effect of all theoretically possible configurations. We note that TEK status is often causally *posterior* to past home rule as states dominated by large ethnic groups (e.g., the Ottoman and Habsburg empires or the Soviet Union) often shrank but survived as rump states with "stranded" segments abroad. These non-dominant segments (e.g., ethnic Turks on the Balkan) have a history of past home rule and links to a dominant TEK group. TEK status therefore captures part of the effect of past home rule.

We cluster standard errors on the level of ethnic groups s to account for dependence over time and between segments. In order to account for the small number of groups with a history of home rule but no or non-dominant TEK, we also compute bootstrapped standard errors for the full interaction model (see Cameron, Gelbach, and Miller 2008).

Results

Our analysis supports Hypotheses 2 and 3 but not Hypothesis 1. Non-dominant ethnic segments with transborder ethnic kin (TEK) are at higher risk of being targeted by campaigns of ethnic cleansing, yet only if their kin has dominant status in another state. In addition, the risk of ethnic cleansing is higher in segments with a history of home rule. This effect partially works through the aforementioned dominant TEK mechanism, but is also present for groups without dominant TEK.

Table 1 presents the results from our main empirical specification. We start by assessing the effect of ethnic segments with non-dominant TEK and dominant TEK ties in Model 1. Ethnic segments that have exclusively non-dominant TEK (Hypothesis 1) exhibit a yearly risk of being ethnically cleansed which is an imprecisely estimated 0.1 percentage point higher than the risk of segments without any TEK links. Confidence intervals are overlapping with the null hypothesis of no effect, which we therefore cannot reject. The effect associated with non-dominant TEK is also significantly smaller (p < .01) than related to dominant TEK ties (Hypothesis 2), which have a 0.62 percentage points higher risk which amounts to 1.8 times the average risk of ethnic cleansing (0.34 percent).

Model 2 presents similar support for Hypothesis 3 in that past home rule has a consistent association with the risk of ethnic cleansing. A doubling of the number of years of ethnic home rule experienced by an ethnic segment since 1816 is associated with an increase in the risk of ethnic cleansing by 0.06 percentage points. Moving from

Table 1. Ethnic Cleansing 1886–2020 (OLS): TEK Links and Past Home Rule.

	Ethnic Cleansing (0/100)			
	(1)	(2)	(3)	(4)
Non-dominant TEK	0.110 (0.140)		0.136 (0.136)	0.182 (0.136)
Dominant TEK	0.621 (0.228)**		0.613 (0.232)**	0.728 (0.233)**
Past home rule (yrs, log)		0.084 (0.041)*	0.063 (0.043)	0.858 (0.355)*
Non-dominant TEK × past home rule				0.098 (0.414)
Dominant TEK × past home rule				−0.848 (0.357)*
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Mean DV	0.34	0.34	0.34	0.34
Observations	32,578	32,578	32,578	32,578
Adjusted R^2	0.021	0.020	0.021	0.022

Notes. OLS linear models. Sample excludes dominant groups. Control variables described in main text. Standard errors clustered on the ethnic group level. Significance codes: $^{\dagger}p < .1$; $*p < .05$; $**p < .01$.

zero years of past home rule—the predominant case in our sample—to 20 years, which is close to the median number of years for segments with past home rule, thus raises the likelihood of ethnic cleansing by 0.25 percentage points, or three quarters of the average risk of 0.34 percent.

Models 3 and 4 then assess the joint impact of segments' TEK ties and history of home rule. Combining all three main variables of interest into the same model, Model 3 shows a diminished and imprecisely estimated, yet positive, effect associated with past home rule. We take this as sign that part of the effect of past home rule works through TEK links: Many segments with extensive past home rule are minorities "stranded" outside their home states after the break up of empires. These segments, such as German populations across the former territories of Germany and Austria-Hungary, were often cleansed after the empires they commanded fell apart. Motivating the expulsion of Germans from post-Second-World-War Poland as preventing future ethno-territorial revisionism, Winston Churchill (Churchill, 1944) declared in 1944 that

> [e]xpulsion is the method which, in so far as we have been able to see, will be the most satisfactory and lasting. There will be no mixture of populations to cause endless trouble, as has been the case in Alsace-Lorraine. A clean sweep will be made.

Lastly, a full interaction of TEK links and past home rule in Model 4 sheds light on the comparative risks of all possible configurations. Plotted in Figure 7, we find that TEK and dominant TEK without previous home rule to have similar effects as
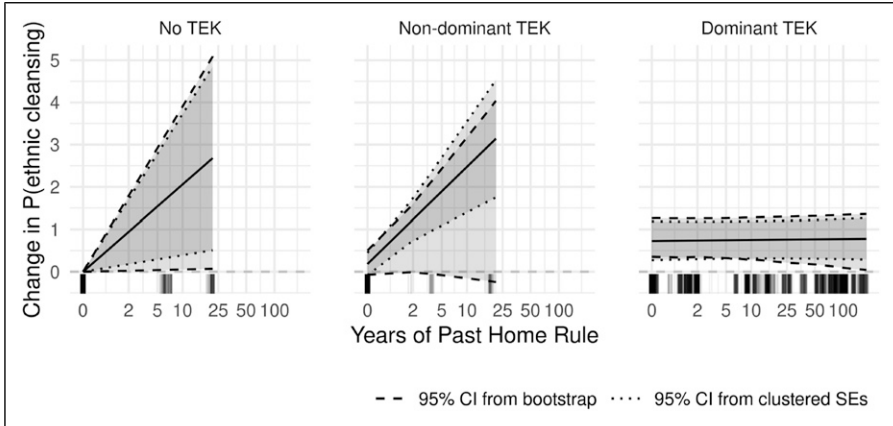
Figure 7. Change in the probability of ethnic cleansing by TEK status and past home rule. *Note.* Comparison group is "No TEK" and no past home rule. Dotted lines demarkate 95 percent CIs from clustered standard errors, dashed lines demarkate 95 percent CIs from an ethnic group-level bootstrap with 1'000 iterations.

in Model 1, the latter significantly increasing the risk of ethnic cleansing. Past home rule, in contrast, only increases the risk for segments without dominant TEK. A doubling of the years of past home rule increases the risk of ethnic cleansing for these segments by approximately 0.86 percentage points, or more than twice the baserate. Due to the small number of affected groups – more than 90 percent of groups without dominant TEK have no history of home rule – these estimates remain statistically significant (p < .05 and <.1, respectively) but exhibit larger uncertainty when computing bootstrapped confidence intervals (dashed lines in Figure 7). However, past home rule does not further increase the risk of ethnic cleansing for segments with dominant TEK, at least partially because its effect is already captured by the dominant TEK dummy itself.

A set of additional analyses investigates whether democratic institutions moderate the effects associated with TEK connections and ethnic segments' history of past home rule. Using electoral democracy ('polyarchy') scores from VDEM (Coppedge et al. 2021), we find that our results are almost exclusively driven by segments in states with autocratic institutions (Online Appendix Table A8). This finding aligns with previous research suggesting important impacts of liberal norms on territorial integrity and peace within and across state borders (e.g. Cederman, Gleditsch, and Wucherpfennig 2017; Imai and Lo 2021; Zacher 2001).

Robustness Checks

We assess the robustness, reporting all results in the Online Appendix. We first document that our results are robust to estimating logistic regressions. Second, we analyze robustness regarding the choice of control variables.

Several of our baseline confounders are arguably "post-treatment": Because past and current state borders constitute (part of) our treatment in that they determine past home rule and TEK status, some attributes of ethnic segments such as their (relative size) and geography are co-determined by these very same borders. As a remedy, we drop all controls and obtain very similar results than in the main specification. In a similar vein, we show robustness to dropping all fixed effects.

On the other hand, there are a host of characteristics of ethnic segments and states left out of the baseline specification that may constitute omitted variables. We therefore add a series of covariates that capture ethnic segments' dispersion and share of the state's population, the overlap of their settlement area with that of their state's dominant group, a segments' distance to the border, as well as the ethnic fractionalization of their host state and fractionalization of their larger kin group across state borders. While these could correlate with the main variables of interest and cause ethnic cleansing, adding them does not substantively change the results. In order to control for potentially biasing omitted characteristics of states, we additionally add state-year fixed effects to our models which control for any time-variant characteristic of the countries our segments find themselves in. Thus only comparing segments within the same year and state, the respective specification shows stable results.

Because the onset of ethnic cleansing is a comparatively "rare" (yet still too common) event that affects 113 observations in our data, our results may be due to pure chance or driven entirely by particular historical (sub-)episodes such as the world wars. We find neither to be likely. We first conduct a randomization inference test (Figure A1) in which we randomly re-allocate the onsets of ethnic cleansing across observations in our data 1'000 times. Our main estimates are located at the very margins of the resulting distributions of estimates. Second, we test whether our results are exclusively driven by the two World Wars. While they constitute "most-likely" historical episodes for our argument and contain half the ethnic cleansing episodes we analyze, their complexity increases the risk of unobserved confounding. Dropping the respective years (1914-1918 and 1939-1945) decreases the effect of dominant TEK and past home rule by 50 percent and increases uncertainty ($p = .10$ and $.12$, in respectively). These findings suggest that our findings are weaker outside these two episodes of large-scale violence in Europe and further motivate the subsequent analysis of the effect of territorial claim and war on ethnic cleansing.

Mechanisms: Ethno-territorial Competition and Warfare

Our main results suggest that in particular group segments with dominant TEK abroad are likely targets of ethnic cleansing. Two likely triggers are territorial claims by a TEK state and violent irredentism in particular. Both raise the immediacy of the territorial threat and sharply limit any potential deterring effect by the TEK state. To shed light on this mechanism, we join our data on ethnic segments with the Correlates of War datasets on territorial claims (Hensel 2001) and interstate warfare (Sarkees and Wayman 2010). For each claim targeted at, and war involving, the host state A of an ethnic segment s, we code whether the claimant or

enemy-state *B* has an ethnic tie to the ethnic segment *s* in question. Our expectation is that such a "fifth-column" status of the group segment lagged by 1 year increases the risk of ethnic cleansing. In the respective model, we also control for whether a state is targeted by claims or involved in wars at all and whether an ethnic segment has any TEK links at all. These terms are included to prevent the estimates of interest to be driven purely by all the claims/wars a state is engaged in or the TEK links a group has.

Table 2 presents the results. While territorial claims from states with a TEK link do not increase the risk of cleansing of the respective ethnic segments (Model 2), interstate warfare does so substantively. Warfare alone increases a group segment's risk of ethnic cleansing by 0.9 percentage points, but that risk increases by another 1.8 percentage points, that is 4.5 times the baseline rate, in cases where the segment has a TEK link to the opposing state.

This result supports our argument that ethnic cleansing is oftentimes driven by territorial competition along ethnic lines, in particular once it materializes as territorial claims and warfare. Once nationalism holds sway and territory can only be legitimately ruled by a state on the basis of a common nationality, states have perverse incentives to ethnically cleanse its territory from non-dominant groups to remove the nationalist incompatibilities and uphold its rule.

Ethnic Cleansing and the Making of European Nation-states

Our empirical analysis has so far focused on the roots of ethnic cleansing in states' incentives to mitigate risks of territorial rightsizing rooted in groups' trans-border ties to dominant groups and historical home rule. These findings open the way to conduct a back-of-the-envelope assessment of the extent to which ethnic cleansing contributed to the alignment between ethnically defined nations and European states, an effect that has been hitherto unquantified in the literature.

There are currently two common yet unsatisfying approaches to measuring the alignment of states and ethnic nations. Most measure states' *ethnic homogeneity* as one minus Herfindahl's Fractionalization Index (e.g., Alesina, Baqir, and Easterly 1999). In turn, the degree to which ethnic nations enjoy *territorial unity* inside the same state can be captured by one minus the degree of political fractionalization of ethnic groups by state borders (e.g. Cederman, Rüegger, and Schvitz 2021). Each of these indeces captures one of the two core dimensions of the state-to-nation alignment affected by ethnic cleansing, yet neither measure is in itself sufficient. We therefore turn towards a third, information-theoretic measure of the *Mutual Information* the geography of states and ethnic nations provide on each other.

Our *Mutual Information* index assesses the amount of information the partitioning of Europe's population into states *S* carries about its partitioning into ethnic nations *N* (Vinh, Epps, and Bailey 2010).[24] We start from a grid of 8739 points that are the centroids of a hexagonal grid that covers the European landmass. Our normalized mutual information (MI) metric is defined as

$$MI(S,N) \quad = H(S) - H(S|N) \tag{2}$$

Table 2. Ethnic cleansing 1886–2020 (OLS): TEK, Territorial Claims, and Interstate Wars.

	Ethnic cleansing (0/100)				
	(1)	(2)	(3)	(4)	(5)
Has TEK		0.010		−0.025	−0.044
		(0.167)		(0.155)	(0.165)
Terr. claims$_{t-1}$	−0.115	−0.220			−0.295
	(0.131)	(0.128)[†]			(0.128)*
Claims from state w/TEK		0.301			0.144
		(0.155)[†]			(0.157)
Interstate war$_{t-1}$			1.314	0.884	0.937
			(0.303)**	(0.269)**	(0.275)**
War with state w/TEK				1.832	1.793
				(0.855)*	(0.872)*
Country FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Mean DV	0.41	0.41	0.39	0.39	0.41
Observations	27,150	26,991	28,736	28,736	26,991
Adjusted R^2	0.020	0.020	0.023	0.024	0.024

Notes. OLS linear models. Sample excludes dominant groups. Control variables described in main text. Standard errors clustered on the ethnic group level. Significance codes: [†]$p < .1$; *$p < .05$; **$p < .01$.

$$MI_{norm}(S,N) \quad = \frac{MI(S,N)}{(H(S) \; H(N))^{.5}} \tag{3}$$

where $H(S)$ and $H(S|N)$ are the (conditional) entropies of points' membership in states S. MI returns the quantity of information S carries on their membership in nations N in bits, normalized in MI_{norm} with the entropies of partitionings S and N. This yields a measure that strictly varies between 0 (no mutual information) and 1 (full mutual information).

Figure 8(a) shows ethnic homogeneity of states, territorial unity of groups, and mutual information between states and ethnic nations in our data on state territories and ethnic geography for each year between 1886 and 2019. Showing a rising state-to-nation alignment, states' ethnic homogeneity increased from 0.55 (approx. the US today) to 0.8 (approx. Sweden today). Ethnic nations' high levels of territorial unity have remained comparatively constant. Combining both dimensions, our measure of mutual information increases from 0.74 in 1886 to 0.86 in 2019.

As a final step in this analysis, we can disaggregate each year-on-year *change* in the mutual information measure into change that resulted from border change and from shifts in ethnic geography. Figure 8(b) shows that, in 2019, changes in ethnic geography have cumulatively contributed 44 percent to the increased alignment between European states and ethnic groups. The remaining 56 percent are due to border change (e.g., Cederman, Girardin, and Müller-Crepon 2023). Clear temporal patterns are
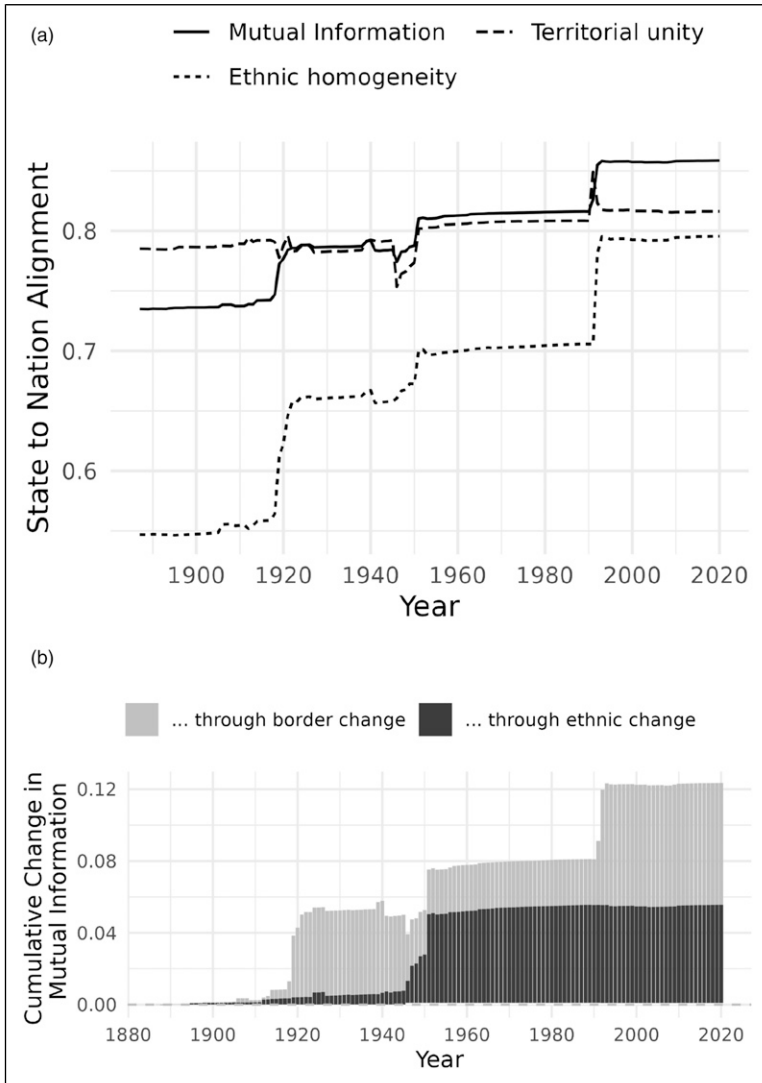
Figure 8. State-to-nation alignment in Europe, 1886–2020. (a) Increasing state-to-nation alignment, 1886-2020 (b) Cumulative change: Contributions of rightsizing and rightpeopling.

visible with the end of World War I and the break-up of the Soviet Union coming with increasing alignments due to border changes, while ethnic cleansing dominated the end of World War II.

The effects of ethnic change on European state-to-nation congruence in Figure 8(b) are overwhelmingly due to violent ethnic cleansing. This can be seen by comparing the

above results, which take non-violent ethnic change into account with results based on the baseline HEG data where temporal variation originates only from periods of rapid, most often violent[25] ethnic change. While we do not currently knowing whether the observed changes indeed *entirely* resulted from violence, the historical literature and sheer scale of ethnic cleansing campaigns indicates violence to be their main driver. Using the baseline HEG data leads to very similar results. The respective cumulative contribution of ethnic change amounts to 39 percent of the increasing alignment of states and ethnic nations.

Our limited data[26] makes it difficult to precisely and causally distinguish the effects of ethnic cleansing on increases in state to nation alignments. Yet, the above exercise in macro-level accounting provides nevertheless a first measurement of the contribution of ethnic displacement and mass-killings on the development of the comparatively homogeneous nation states in today's Europe. Given the scale of victimization brought about by the transformation of European states over the past 140 years, we believe that taking this step is important and encourage future improvements.

Conclusion

Many contemporary nation-states in Europe were ethnically homogenized by violent means. Since the nineteenth century, ethnic cleansing was, and still is, among the most important sources of human suffering. It is at the same time a root cause of the current ethnic homogeneity of European states, achieved in large part by violently "right-peopling" their populations.

Building on historical and political science literatures, we argue that threats to the territorial integrity of states constituted an important driver of ethnic cleansing since the nineteenth century. In this age of nationalism, the boundaries of ethnic nations became the paramount legitimizing principle of states territorial rule. Multi-ethnic states were at risk of being "right-sized" through secessionism and irredentism, in particular where groups could draw on transnational ethnic ties or follow a historical precedent of home rule. By ethnically cleansing these territories, states sought to reduce the disjunction between political and ethnic borders that nationalists despise. Ethnic cleansing is thus one of the perverse, if logical, consequences of the ethnic nationalism that has reshaped the European state system since the nineteenth century.

We test the effect of transnational ethnic ties and past home rule of non-dominant ethnic groups on their risk of ethnic cleansing with new data of the changing settlement areas of European ethnic groups since 1886. Combined with a new enumeration of episodes of ethnic cleansing, we find general support for our arguments. Non-dominant ethnic groups with transnational kin that dominate another state, are exposed to a severely increased risk of ethnic cleansing while ties to groups that do not dominate a state have no sizeable or statistically significant effect. Relatedly, ethnic segments that can draw on a history of home rule are at increased risk of becoming targeted by ethnic cleansing campaigns. These effects are weaker under democratic institutions, which can offer pathways to accommodation and power-sharing. Importantly, the risk of

ethnic cleansing is closely associated with times of interstate warfare, especially with states in which an ethnic group has transnational ethnic kin. Ethnic cleansing is thus often rooted in territorial competition structured along ethno-nationalist lines. Moving back to the macro-level, we find that changes in ethnic geography associated with ethnic cleansing explain approximately 40 percent of the increasing congruence between states and ethnic nations. Thus playing a key role in the making of today's European nation-states, these findings highlight the need to take the historical origins of ethnic demography seriously when studying its effects.

Our argument and findings resonate with many past cases of ethnic cleansing, such as the Armenians in the Ottoman Empire and Nagorno-Karabakh, the cleansing of ethnic Turks from the Balkans, or the persecution of Poles under Hitler and Stalin. It does not, however, apply to all cases of ethnic cleansing. Nor does it exhaustively explain those cases where the logic is present. Some groups became targets for reasons unrelated to territorial threats, most importantly the Jewish and Roma populations during the Holocaust. In addition, even where ethnic cleansing can be linked to territorial threats, other factors such as cross-cutting cleavages (Bulutgil 2015, 2016) or war-fighting strategies (Lichtenheld 2020) have determined the conduct, scope, and timing of governments' campaigns. Constrained by our macro-historical abstraction and the scope of our empirical data, we have studied nationalist state-to-nation discrepancies as structural drivers of the ethnic cleansing that violently right-peopled European states.

ORCID iDs

Carl Müller-Crepon https://orcid.org/0000-0001-8536-7861
Lars-Erik Cederman https://orcid.org/0000-0002-8511-0617

Data Availability Statement

The data that supports the findings of this study are available in the supporting information of this article.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. In analogy to "right-sizing," we use thisterm to refer to state efforts to homogenize their populations, without of course implying any legitimacy of such efforts (see O'Leary et al. 2001).
2. Non-dominant groups that do not demographically and politically dominate a state.
3. The data and an associated R-Package are available at github.com/carl-mc/hegdata.
4. Although non-state actors can also engage in ethnic cleansing, the vast majority of Europe's ethnic cleansing campaigns were carried out by states.
5. It is difficult to draw a sharp line between the two types of ethnic cleansing, as forced displacement is often accompanied by mass killings and the latter also frequently involves mass deportations. Therefore, our analysis focuses on ethnic cleansing as an overarching category.
6. Alternative "right-peopling" strategies (O'Leary 2001) are assimilation (forced or voluntary) and state-sponsored resettlement of dominant group members into contested regions (McNamee and Zhang 2019).
7. Approximately Poland, Belarus, Ukraine, Estonia, Latvia, Lithuania, parts of eastern Romania, and western Russia.
8. Given the driving force of nationalist ideology, we do not expect our argument to hold before that time.
9. Search terms included in particular "ethnic*", "ethnograph*", "linguistic", and "language" in all major European languages.
10. Online Appendix Table A9 enlists all maps.
11. If two maps depict "Bavarian," a German dialect, and while twenty show "Germans," "German" enters the master list, subsuming all German dialects.
12. When a map depicts only a group's parent-language, we drop the respective settlement area as we know that the group is not present outside that area but know little about the group inside its parent's settlement area.
13. This is done through a kernel regression smoother at the grid-cell level.
14. See Figure A4.
15. We drop segments that are smaller than 10'000 inhabitants and less than 1 percent of group e to remove tiny artificial "spill-over" segments along international borders. Population estimates rely on the 1800 HYDE 3.1 data (Goldewijk, Beusen, and Janssen 2010).
16. In comparison and because of her higher victimization threshold, Bulutgil (2015, 2016) identifies 41 cases of ethnic cleansing in Europe during the same period. Butcher et al. (2020)

use a yearly threshold of 25 deaths, enlisting 201 target mass killing episodes globally since 1946.

17. The estimate is likely conservative as we take the lower value where the literature indicates a range.

18. This is the sum of segments' population in the years of onset of ethnic cleansing.

19. This computation is challenging without individual level data. The number is calculated as $1 - (1 - (\sum_{y=1886}^{2020} Victims_y / \sum_{y=1886}^{2020} Population_y))^{50} = .0306$, i.e. the Europe-wide, individual-level probability of becoming a victim of ethnic cleansing in any given year accumulated over an assumed (and rather low) life expectancy of 50 years.

20. Constructed from the CShapes 2.0 data (Schvitz et al. 2022), Centennia (Reed 2008), and secondary sources. See also Cederman, Girardin, and Müller-Crepon (2023).

21. As above, we derive dominant groups for historical states by taking the largest ethnic group in their capitals. We use the earliest set of ethnic maps for that purpose but are not aware of any case in which the largest group in capitals changed dramatically.

22. All from FAO's GAEZ database.

23. Online Appendix Table A2 shows robustness to estimating logistic regressions.

24. For each point, we sample ethnic groups from the compositional HEG data.

25. Episodes of forced displacement and genocide make up 83 percent of these periods

26. Importantly, we here neglect increasing ethnic diversity due to spatially diffuse immigration uncaptured by our HEG data.

References

Abramson, Scott F., and David B. Carter. 2016. "The Historical Origins of Territorial Disputes." *The American Political Science Review* 110 (4): 675.

Abramson, Scott F., David B. Carter, and Luwei Ying. 2022. "Historical Border Changes, State Building, and Contemporary Trust in Europe." *American Political Science Review* 116 (3): 875–895.

Akçam, Taner. 2012. *The Young Turks' Crime against Humanity.* Princeton: Princeton University Press.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114 (4): 1243-1284.

Balcells, Laia, and Jessica A. Stanton. 2021. "Violence against Civilians during Armed Conflict: Moving beyond the Macro-And Micro-level Divide." *Annual Review of Political Science* 24: 45-69.

Bruk, Solomon I., and V. S. Apenchenko. 1964. *Atlas Narodov Mira.* Moscow: Glavnoe upravlenie geodezii i kartografii gosudarstvennogo geologicheskogo komiteta SSSR and Institut etnografii im. H. H. Miklukho-Maklaia, Akademiia nauk SSSR.

Bulutgil, H. Zeynep. 2015. "Social Cleavages, Wartime Experience, and Ethnic Cleansing in Europe." *Journal of Peace Research* 52 (5): 577-590.

Bulutgil, H. Zeynep. 2016. *The Roots of Ethnic Cleansing in Europe.* Cambridge: Cambridge University Press.

Butcher, Charles, Benjamin E. Goldsmith, Sascha Nanlohy, Arcot Sowmya, and David Muchlinski. 2020. "Introducing the Targeted Mass Killing Data Set for the Study and Forecasting of Mass Atrocities." *Journal of Conflict Resolution* 64 (7-8): 1524-1547.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics* 90 (3): 414-427.

Carter, David B. 2010. "The Strategy of Territorial Conflict." *American Journal of Political Science* 54 (4): 969-987.

Carter, David B. 2017. "History as a Double-Edged Sword: Historical Boundaries and Territorial Claims." *Politics, Philosophy & Economics* 16 (4): 400-421.

Cattaruzza, Marina. 2010. "'Last Stop Expulsion'–The Minority Question and Forced Migration in East-Central Europe: 1918–49 to Hans Lemberg, in Memory." *Nations and Nationalism* 16 (1): 108-126.

Cederman, Lars-Erik, Kristian Skrede Gleditsch, and Julian Wucherpfennig. 2017. "Predicting the Decline of Ethnic Civil War: Was Gurr Right and for the Right Reasons?" *Journal of Peace Research* 54 (2): 262-274.

Cederman, Lars-Erik, Luc Girardin, and Carl Müller-Crepon. 2023. "Nationalism and the Puzzle of Reversing State Size." *World Politics* 75 (4): 692-734.

Cederman, Lars-Erik, Seraina Rüegger, and Guy Schvitz. 2021. "Redemption through Rebellion: Border Change, Lost Unity and Nationalist Conflict." *American Journal of Political Science* 66 (1): 24–42.

Churchill, Winston. 1944. "Speech in Commons on Soviet-Polish Frontier."

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, Nazifa Alizada, David Altman, Michael Bernhard, Agnes Cornell, M. Steven Fish, Lisa Gastaldi, Haakon Gjerlow, Adam Glynn, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Anja Neundorf, Pamela Paxton, Daniel Pemstein, Oskar Ryden, Johannes von Römer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundström, Eitan Tzelgov, Luca Uberti, Yi-Ting Wang, Tore Wig, and Daniel Ziblatt. 2021. "V-Dem Dataset." *Dataset*.

Darden, Keith, and Harris Mylonas. 2016. "Threats to Territorial Integrity, National Mass Schooling, and Linguistic Commonality." *Comparative Political Studies* 49 (11): 1446-1479.

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth* 8 (2): 195-222.

Fjelde, Hanne, and Lisa Hultman. 2014. "Weakening the Enemy: A Disaggregated Study of Violence against Civilians in Africa." *Journal of Conflict Resolution* 58 (7): 1230-1257.

Fjelde, Hanne, Lisa Hultman, Livia Schubiger, Lars-Erik Cederman, Simon Hug, and Margareta Sollenberg. 2021. "Introducing the Ethnic One-Sided Violence Dataset." *Conflict Management and Peace Science* 38 (1): 109-126.

Garrity, Meghan. 2022. "Introducing the Government-Sponsored Mass Expulsion Dataset." *Journal of Peace Research*. FirstView.

Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca: Cornell University Press.

Germann, Micha, and Nicholas Sambanis. 2021. "Political Exclusion, Lost Autonomy, and Escalating Conflict over Self-Determination." *International Organization* 75 (1): 178-203.

Goldewijk, Kees Klein, Arthur Beusen, and Peter Janssen. 2010. "Long-term Dynamic Modeling of Global Population and Built-Up Area in a Spatially Explicit Way: HYDE 3.1." *The Holocene* 2010 (1): 1-9.

Griffiths, Ryan D. 2016. *The Age of Secession: The International and Domestic Determinants of State Birth*. Cambridge: Cambridge University Press.

Gurr, Ted Robert. 2000. "Ethnic Warfare on the Wane." *Foreign Affairs* May: 52–64.

Harff, Barbara. 2003. "No Lessons Learned from the Holocaust? Assessing Risks of Genocide and Political Mass Murder since 1955." *American Political Science Review* 97 (1): 57–73.

Harff, Barbara, and Ted Robert Gurr. 1988. "Toward Empirical Theory of Genocides and Politicides: Identification and Measurement of Cases since 1945." *International Studies Quarterly* 32 (3): 359-371.

Hechter, Michael. 2000. *Containing Nationalism*. Oxford: Oxford University Press.

Hensel, Paul R. 2001. "Contentious Issues and World Politics: The Management of Territorial Claims in the Americas, 1816–1992." *International Studies Quarterly* 45 (1): 81-109.

Hong, Mi Hwa, and Nam Kyu Kim. 2019. "How Do External Territorial Threats Affect Mass Killing?" *Journal of Peace Research* 56 (4): 529-544.

İçduygu, Ahmet, and Deniz Sert. 2015. "The Changing Waves of Migration from the Balkans to Turkey: A Historical Account." In *Migration in the Southern Balkans: From Ottoman Territory to Globalized Nation States*, edited by Hans Vermeulen, Martin Baldwin-Edwards, and Riki van Boeschoten, 85-104. Cham: Springer International Publishing.

Imai, Kosuke, and James Lo. 2021. "Robustness of Empirical Evidence for the Democratic Peace: A Nonparametric Sensitivity Analysis." *International Organization* 75 (3): 901-919.

Korb, Alexander. 2016. "Homogenizing Southeastern Europe, 1912–99: Ethnic Cleansing in the Balkans Revisited."

Lewis, M. Paul. 2009. *Ethnologue: Languages of the World*, Vol. 16. Dallas, TX: SIL International.

Lichtenheld, Adam G. 2020. "Explaining Population Displacement Strategies in Civil Wars: a Cross-National Analysis." *International Organization* 74 (2): 253-294.

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

McNamee, Lachlan. 2018. "Mass Resettlement and Political Violence: Evidence from Rwanda." *World Politics* 70 (4): 595-644.

McNamee, Lachlan, and Anna Zhang. 2019. "Demographic Engineering and International Conflict: Evidence from China and the Former USSR." *International Organization* 73 (2): 291–327.

Müller-Crepon, Carl, Guy Schvitz, and Lars-Erik Cederman. 2023. "Shaping States into Nations: The Effects of Ethnic Geography on State Borders." *American Journal of Political Science*. forthcoming.

Müller-Crepon, Carl, Yannick Pengl, and Nils-Christian Bormann. 2020. "Linking Ethnic Data from Africa." *Journal of Peace Research* 55 (5): 687-698.

Murphy, Alexander B. 1990. "Historical Justifications for Territorial Claims." *Annals of the Association of American Geographers* 80 (4): 531-548.

Mylonas, Harris. 2012. *The Politics of Nation-Building: Making Co-nationals, Refugees, and Minorities*. Cambridge: Cambridge University Press.

Mylonas, Harry, and Scott Radnitz, eds. 2022. *Enemies Within: The Global Politics of Fifth Columns*. Oxford: Oxford University Press.

O'Leary, Brendan. 2001. "The Elements of Right-Sizing and Right-Peopling the State." *Right-sizing the State: The Politics of Moving Borders* 15–73.

O'Leary, Brendan, Ian Lustick, and Thomas M. Callaghy. 2001. *Right-sizing the State: The Politics of Moving Borders*. Oxford: Oxford University Press.

Posen, Barry R. 1993. "The Security Dilemma and Ethnic Conflict." *Survival* 35 (1): 27-47.

Reed, Frank. 2008. "Centennia Historical Atlas." URL: https://historicalatlas.com/

Salehyan, Idean. 2007. "Transnational Rebels: Neighboring States as Sanctuary for Rebel Groups." *World Politics* 59 (2): 217-242.

Sarkees, Meredith Reid, and Frank Wayman. 2010. *Resort to War: 1816-2007*. Washington, DC: Cq Press.

Schubiger, Livia. 2022. "One for All? State Violence and Insurgent Cohesion." *International Organization*. forthcoming.

Schvitz, Guy, Luc Girardin, Seraina Rüegger, Nils B. Weidmann, Lars-Erik Cederman, and Kristian Skrede Gleditsch. 2022. "Mapping the International System, 1886-2019: the CShapes 2.0 Dataset." *Journal of Conflict Resolution* 66 (1): 144-161.

Shields, Sarah. 2013. "The Greek-Turkish Population Exchange: Internationally Administered Ethnic Cleansing." *Middle East Report* 267: 2-6.

Siroky, David S., and John Cuffe. 2015. "Lost Autonomy, Nationalism and Separatism." *Comparative Political Studies* 48 (1): 3-34.

Snyder, Timothy. 2011. *Bloodlands: Europe between Hitler and Stalin*. London: Random House.

Straus, Scott. 2013. *The Order of Genocide*. Ithaca: Cornell University Press.

Ther, Philipp. 2014. *The Dark Side of Nation-States: Ethnic Cleansing in Modern Europe*. New York: Berghahn Books.

Valentino, Benjamin A. 2004. *Final Solutions: Mass Killing and Genocide in the 20th Century*. Ithaca: Cornell University Press.

Van Houten, Pieter. 1998. "The Role of a Minority's Reference State in Ethnic Relations." *European Journal of Sociology/Archives Européennes de Sociologie* 39 (1): 110-146.

Vinh, Nguyen Xuan, Julien Epps, and James Bailey. 2010. "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance." *Journal of Machine Learning Research* 11 (Oct): 2837-2854.

Weber, Eugen. 1976. *Peasants into Frenchmen: The Modernization of Rural France 1870-1914*. Stanford: Stanford University Press.

Weidmann, Nils B., Jan Ketil Rød, and Lars-Erik Cederman. 2010. "Representing Ethnic Groups in Space: A New Dataset." *Journal of Peace Research* 47 (4): 491-499.

Weiner, Myron. 1971. "The Macedonian Syndrome an Historical Model of International Relations and Political Development." *World Politics* 23 (4): 665-683.

WLMS. 2006. "World Language Mapping System." URL: https://www.gmi.org/wlms/

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

Zacher, Mark W. 2001. "The Territorial Integrity Norm: International Boundaries and the Use of Force." *International Organization* 55 (2): 215-250.