Received: 2 February 2022 | Revised: 8 May 2023 | Accepted: 8 May 2023

DOI: 10.1111/ajps.12838

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

ARTICLE

Shaping states into nations: The effects of ethnic geography on state borders

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

[1]Department of Government, London School of Economics and Political Science, London, UK

[2]European Commission, Joint Research Centre, Ispra, Varese, Italy

[3]Center for Comparative and International Studies, ETH Zürich, Zurich, Switzerland

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

Abstract
Borders define states, yet little systematic evidence explains where they are drawn. Putting current challenges to state borders into perspective and breaking new methodological ground, this paper analyzes how ethnic geography and nationalism have shaped European borders since the 19th century. We argue that nationalism creates pressures to redraw political borders along ethnic lines, ultimately making states more congruent with ethnic groups. We introduce a Probabilistic Spatial Partition Model to test this argument, modeling state territories as partitions of a planar spatial graph. Using new data on Europe's ethnic geography since 1855, we find that ethnic boundaries increase the conditional probability that two locations they separate are, or will become, divided by a state border. Secession is an important mechanism driving this result. Similar dynamics characterize border change in Asia but not in Africa and the Americas. Our results highlight the endogenous formation of nation-states in Europe and beyond.

Borders are constitutive features of the modern state system that define the size and shape of states and specify the limits of state sovereignty.[1] A growing literature documents borders' attributes (Simmons & Kenwick, 2021) and consequences (Abramson & Carter, 2016; Carter & Goemans, 2011; Michalopoulos & Papaioannou, 2016; Simmons, 2005). Yet, their origins remain understudied with much research treating states and their borders as exogenous. Border formation has however gained renewed relevance as Russia invaded Ukraine, majorities support territorial revisionism in Hungary, Greece, Bulgaria, and Turkey (Fagan & Poushter, 2020), and secessionists challenge states' territorial integrity in Scotland, Northern Ireland, and Catalonia. Ethno-nationalist demands to redraw state borders along ethnic lines are central to all these cases.

Yet, despite their intuitive appeal, explanations that seek borders' origins in ethnicity are contested and not systematically tested. Addressing this gap, we ask whether, how, and to what extent ethnic geography has shaped Europe's partitioning into states since the 19th century. Following macrosociological theories, we argue that the historical rise of nationalism, "a political principle which holds that the political and national unit should be congruent" (Gellner, 1983, p. 1), created demand for nation-states. As most nations are ethnically defined, nationalism prompted popular pressures to redraw borders along ethnic lines through secessionism, unification, and irredentism (Hechter, 2000; O'Leary, 2001; Weiner, 1971). Of these mechanisms, secessionism is most common and systematically studied. While the ethno-political roots of secessionist conflict are well evidenced (e.g., Cederman et al., 2013; Germann & Sambanis, 2021), some studies of secessions discount ethnicity and nationalism in favor of pre-existing political units and power politics (Coggins, 2014; Griffiths, 2016; Roeder, 2012). We contribute to this debate by integrating

[1] See Sack (1986) on human territoriality more generally.

Verification Materials: The materials required to verify the computational reproducibility of the results, procedures, and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: http://doi.org/10.7910/DVN/T1NXVQ.

secessionist, unificationist, and irredentist border change into a common analytical framework[2] and by overcoming previous studies' problematic reliance on geographically fixed units of analysis.[3]

We thus innovate the study of border determinants, which so far lacks a robust quantitative estimator to test theoretical arguments against potentially confounding alternative hypotheses. Realists argue that borders emerge along mountains and rivers, facilitating internal power-projection and effective defense (Morgenthau, 1985, also Kitamura & Lagerlöf, 2020). From an institutionalist perspective, borders are coordination devices based on states' preferences for territory and stability (Simmons, 2005) and often follow local "focal" lines—rivers, watersheds, or historical precedents (Abramson & Carter, 2016; Carter & Goemans, 2011; Goemans, 2006; Goemans & Schultz, 2017). A third perspective highlights borders' origins in ethnic geography. Alesina and Spolaore (1997, 2005) theorize the trade-off between economies of scale and costs of ethnic heterogeneity in large states (see also Friedman, 1977; Desmet et al., 2011). We empirically test the effect of ethnic geography on state borders and provide comprehensive evidence that accounts for alternative explanations.

To do so, we overcome three challenges of assessing the determinants of borders and the spatial partitioning they produce. First, border formation is an intractable problem as infinitely many borders can partition space into an ex ante unknown number of units. Second, borders entail significant and complex spatial dependencies as they form contiguous, nonoverlapping units. Third, unbiased estimation of ethnic geography's effect on borders requires consideration of confounding geographic features that affect both.

We address these challenges with a new *Probabilistic Spatial Partition Model* (PSPM), which allows us to estimate the conditional effect of spatial features (e.g., ethnic settlement patterns) on the partitioning of geographic space into nonoverlapping units (e.g., states). The model discretizes geographic space as a planar network of points that encodes the main dependent and independent variables. It makes partitionings tractable, accounts for spatial dependencies, estimates effects conditional on covariates, and yields valid uncertainty estimates. Beyond our present use, the PSPM can be applied to model other partitionings, for example, administrative or electoral units. We provide an accompanying open-source R package and code for handling spatial network data.[4]

We use the PSPM to estimate the effect of ethnic geography on state borders. Our new, time-varying data on ethnic geography predate (changing) state borders in Europe since 1855. Digitized from 73 historical maps, the data set enables us to analyze borders and border change based on preexisting ethnic settlement areas. We address omitted variable and reverse causality bias by pairing a cross-sectional with a lagged dependent variable (LDV) model that captures the effect of ethnic geography on border change.

We find that an ethnic boundary between two locations increases the probability that they are or will become separated by an international border by 34 and 17 percentage points, respectively. This finding is robust to accounting for potentially endogenous changes in ethnic geography, alternative measures of ethnic differences, additional controls, and changes to the spatiotemporal data structure. Additional analyses highlight ethnic secession as a key mechanism: Since 1946, areas home to peripheral ethnic groups saw secessionist claims, civil wars, and border change 11, 21, and 50 times more often than other areas. Moving beyond Europe, we find that ethnic boundaries explain border change since the 1960s in Asia but not elsewhere.

NATIONALISM AND THE SHAPING OF STATES

We argue that the rise of nationalism created a demand for ethnically homogeneous nation-states, which caused an increasing alignment of Europe's borders with the underlying ethnic map. This development is part of a larger process of the "right-peopling" and "right-sizing" of states (O'Leary, 2001). The former has received much attention in nationalism studies evidencing the formation of nations within states through assimilationist policies and ethnic violence (Bulutgil, 2016; McNamee & Zhang, 2019; Weber, 1976; White, 2004) or local dissimilation processes along state borders (Sahlins, 1989). Yet, an exclusive focus on state-led identity formation that follows Hobsbawm's (1990, p. 10) claim that "[n]ations do not make states and nationalisms but the other way around" neglects parallel changes in state borders and risks underestimating the full impact of nationalism.[5] We therefore focus on the nationalist right-sizing of states along ethnic lines and address reverse processes as an empirical challenge.

How did nationalism transform Europe's borders? We start by defining ethnic groups as "those human groups that entertain a subjective belief in common descent" (Weber, 1978, pp. 385–398), most frequently distinguished by their language and religion.

[2] See also Cederman et al. (2023) who study states' size but not their borders.
[3] See, for example, Griffiths (2016, chapter 2).
[4] Available at https://github.com/carl-mc/pspmand https://github.com/carl-mc/SpatialLattice.

[5] The two processes are linked as ethnic homogenization often focuses on contested territories (Bulutgil, 2015, 2016; McNamee & Zhang, 2019; Mylonas, 2012).

Once groups' members desire to control a state, they become ethnic nations, "a community of sentiment which would adequately manifest itself in a state of its own" (Weber, 1978, p. 176). In consequence, ethnonationalist ideology requires "that ethnic boundaries should not cut across political ones, and, in particular, that ethnic boundaries within a given state [...] should not separate the power-holders from the rest" (Gellner, 1983, p. 1). Three constellations violate Gellner's congruence principle, each motivating a specific type of border change.

First and most common are ethnic minorities in a state dominated by a different group. Such "alien rule" (Hechter, 2013) deprives groups of self-determination and state services that often favor ruling groups (De Luca et al., 2018). In response, stateless nations may try to attain statehood by secession. The break-up of empires and multiethnic states exemplifies this process (Beissinger, 2002; Germann & Sambanis, 2021). With many more potential ethnic nations than states,[6] secessionism is the most common type of border change (Gellner, 1983; Griffiths, 2016; Hechter, 2000).

Second, ethnonationalist grievances can also emerge if an ethnic group is divided by state borders, prompting nationalist calls for unification (Cederman et al., 2022). The promise of benefits from governance over a larger and ethnically homogeneous territory and population can help their cause (Alesina & Spolaore, 2005). Such efforts sometimes yield the merger of coethnic units, as illustrated by 19th-century Germany and Italy and the more recent reunifications of Vietnam, Yemen, and Germany. Concomitant to the decline of state death since 1945 (Fazal, 2004, 2007), ethnic unification is exceedingly rare.

Third, a configuration in which an ethnic group dominates one state but forms a minority in another can pressure the homeland government to "liberate" their kin, thus resulting in irredentist nationalism (Weiner, 1971; Siroky & Hale, 2017). Named after Italian Veneto and Trento that remained "unredeemed" after the first wave of Italian unification, the stronger territorial integrity norm has reduced irredentist border change after World War II (Zacher, 2001).

Nationalist ideology equips revisionist activists of all three situations with powerful arguments that legitimize their claims over ostensibly "indivisible" territory and mobilize elites and citizens for their projects (Goddard, 2006; Hroch, 1985; Murphy, 2002). While collective action problems and resistance by the incumbent state can inhibit actual border change (Hardin, 1995), nationalist grievances can lower the bar by making activists less risk averse (Petersen, 2002; Nugent, 2020; Germann & Sambanis, 2021). Still, revisionist nationalism is unlikely to succeed without considerable material and organizational resources (Tilly, 1978). Alternatively, geopolitical and economic crises create opportunities for change by weakening existing states, as illustrated by imperial collapse after the World Wars (Abramson & Carter, 2021; Skocpol, 1979). In addition, nationalist "successes" can inspire nationalists elsewhere, further reinforcing the spatiotemporal clustering of border change. Nationalist ideas spread through 19th-century Europe and globally thanks to the "Wilsonian moment" after World War I (Manela, 2007).

Yet, the diffusion of nationalism beyond Europe did not necessarily produce ethnonationalist congruence. The disintegration of the massively multiethnic European colonial empires led to new borders that cut through ethnic groups and created ethnically diverse independent states (Englebert et al., 2002). While some activists supported pan-nationalism, the prevailing elites in the Global South generally subscribed to the legal norm of *uti possidetis*. This implied that new borders would follow colonial administrative borders regardless of their ethnic fit (Ratner, 1996). Where ethnic groups were much smaller than states, as in sub-Saharan Africa, *uti possidetis* was particularly influential (Carter & Goemans, 2011), a tendency that was further reinforced by a lack of interstate competition over sparsely populated areas (Herbst, 2000) and international norms (Zacher, 2001). Even under these conditions, sub-Saharan Africa was far from immune to ethnonationalist revisionism, as evidenced by Somali irredentism and Biafran separatism in Nigeria. In contrast and thanks to the presence of demographically dominant groups, ethnonationalism had a larger influence on border drawing in postcolonial Asia.

Regardless of the specific historical context, those mobilizing for border change will base their territorial claims on their—often self-serving—understandings of ethnic geography. However, even where mobilization successfully achieves border change, "ethnically pure" borders tend to be elusive because of overlapping and noncontiguous ethnic settlement patterns (Sambanis & Schulhofer-Wohl, 2009). As a result, ethnic geography determines the *approximate* location of new borders. In turn, sharp focal lines such as previous administrative borders, historical precedents, rivers, or watersheds inform their local settlement (Carter & Goemans, 2011; Goemans, 2006).

Analyzing the primacy of secession and the global generalizability of the argument in separate analyses, our main empirical focus is on the overall impact of ethnic settlement patterns on European state borders:

Hypothesis 1. Ethnic settlement patterns shape state territories such that ethnic boundaries and state borders become increasingly congruent.

[6] Particularly after the German and Italian unifications outside our empirical scope.

UNIT OF ANALYSIS AND DATA

We test our claims about the effect of ethnic boundaries on state borders using time-variant data on state borders and ethnic geography in Europe since 1886. This section explains how we go beyond previous studies of border determinants by modeling the European landmass as a spatial network of points. We use the network to encode our data and estimate our new PSPM presented subsequently.

Geographic space as a network of points

We model geographic space as a network of points, a move that addresses limitations of previous analyses of border locations. These have followed three approaches. First, Goemans (2006) and Carter and Goemans (2011) show that new borders are frequently drawn along focal lines such as natural frontiers, administrative borders, or historical precedent. This valuable description of border characteristics provides the ground for analyzing border precedents as influential causes of border stability (Abramson & Carter, 2016; Carter & Goemans, 2011). Yet, a focus on observed borders produces limited insights into their causes, since it neglects all potential but unrealized borders. In addition, a focus on locally aligned features risks missing factors such as ethnic geography that only determine borders' *approximate* location at a higher geographic level.

A second approach by Kitamura and Lagerlöf (2020) uses grid cells as seemingly independent units to examine the frequency with which they have been crossed by state borders. Doing so disregards non-monotonic spatial dependencies inherent in the outcome of interest. Because borders partition space into contiguous territorial units, they interdependently emerge in grid cells. For example, a border will cross a string of pairs of neighboring grid cells, violating the assumption of unit-independence in standard regression approaches as the outcome for any unit depends on its relation to the *ensemble* of neighboring cells (not) crossed by a border. Classic spatial error clustering (e.g., Conley, 1999) and spatio(temporal) diffusion models (Wucherpfennig et al., 2021) rely on an exogeneously imposed spatial connectivity matrix and are thus unable to recover such endogenous spatial dependency structures.

A third approach compares observed partitionings with simulated ones. Prominent in the literature on gerrymandering (e.g., Fifield et al., 2020), such comparisons are based on aggregate statistics, as in our case the ethnic homogeneity of observed and simulated states. This approach yields information on the likelihood that an observed partitioning could have originated from the simulated process. Yet, such analyses do not produce inferences about the effects of a given spatial feature on the observed partitioning, in particular in the presence of confounders.

In response to these limitations, we introduce a simplified understanding of space as a planar network G of N points. Discretizing space makes tractable the problem of analyzing the partitioning of a continuous surface, which otherwise has infinitely many possible outcomes. Coupled with the partition model introduced below, the network structure of the data allows us to capture the spatial dependencies that characterize borders. Taking a network of points guarantees that G's vertices have unambiguous partition memberships. G covers Europe[7] as a hexagonal lattice with 1096 nodes and 2905 edges. Its nodes j are connected to their up to six first-degree neighbors k at a distance of ~100 km (Figure 1a).[8]

Data on state borders

Our main outcome is the map of states: the partitioning P_t of the lattice G_t into states in year t. We measure P_t by retrieving the state each vertex belongs to between 1886 and 2019 from the CShapes 2.0 data set (Schvitz et al., 2022). We analyze borders in every 25th year, that is, in 1886, 1911,..., 2011.[9] The quarter-century intervals are long enough for cumulative border change to produce meaningful variation yet short enough to capture varying patterns of border change since 1886.

Figure 1b plots the outcome data in 1886. While we can distinguish "Spain" from "France," these labels are, for our purposes, completely interchangeable. Because we do not ex ante know the number or names of states, we are not interested in whether some vertices became part of "France." Instead, we study whether certain vertices together form a contiguous state—a partition. The set of all partitions defines the partitioning of Europe into states.

Data on historical ethnic settlement patterns

We collect new data on ethnic settlement areas in Europe since 1855. Our main independent variable measures whether an edge crosses and ethnic boundary or not. We construct this measure from 73 historical maps that together capture changes in ethnic

[7] We avoid state-based definitions and define Europe's eastern border from the Bosporus, via the Black Sea, the Carpathian mountain ridge, the Caspian Sea, and the Ural.

[8] This minimizes geographic distortion. Supporting Information (SI) Section D.6 (p. 22) shows robustness to varying the graph's location, resolution, and structure.

[9] SI Section D.5 (p. 21) analyzes alternative temporal structures.

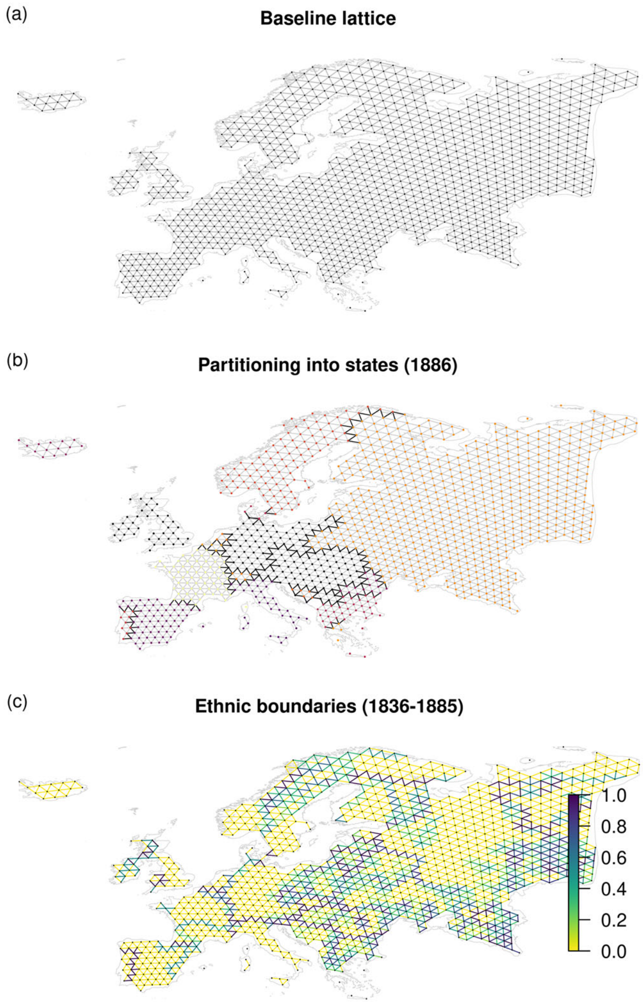

FIGURE 1 Europe as a hexagonal spatial lattice. *Note*: In (b), nodes are colored by state and border-crossing edges are drawn in black. In (c), edges' color denotes the fraction of maps in which an edge crosses an ethnic boundary.

settlement patterns over the past 165 years. Changes from genocides and population exchanges are well documented, while assimilation has more gradually altered ethnic geography. Our historical and time-varying data avoid reverse causality that may arise from backwards-projecting contemporary ethnic data.

Ethnic maps first emerged in the mid-19th century and proliferated due to at least two factors. First, innovations in statistics and cartography enabled the linguistic and religious categorization of local populations. Second, the rise of state-driven and peripheral nationalisms created a demand for maps of ethnic groups (Hansen, 2015; Kertzer & Arel, 2002). Initial efforts by German and Austrian geographers in the 1840s were followed by authors from Russia, the Balkans, and other parts of Europe, resulting in a scientific community dedicated to ethnic cartography.

For the most part, maps were drawn from census data on the town or district level, and defined ethnicity based on native languages (Cadiot, 2005; Hansen, 2015). The production of ethnic maps was generally viewed as a scientific endeavor, motivated by enlightenment-era ideals of measuring and classifying the "natural" world (Livingstone & Withers, 1999). Cartographers therefore sought to establish common standards and provided detailed justifications (Hansen, 2015).

However, ethnic maps and census data were also used politically, employed by states and nationalist movements to shape perceptions of national homelands and support territorial claims (Anderson, 1991; Herb, 2002).[10] This was most evident at the Paris Peace conference of 1919, where all parties relied on their own maps to support their demands (Palsky, 2002). Yet, the scope for manipulation was limited. Because cartographers largely relied on similar data and methods, they could not arbitrarily "invent" ethnic boundaries without jeopardizing their reputation (Hansen, 2015; Herb, 2002). Instead, most attempts to manipulate maps and census data involved the subtle use of politically convenient criteria such as the choice of sources, population thresholds (Hansen, 2015), and the underlying list of ethnic groups (Cadiot, 2005; Hirsch, 1997).[11] While ethnic categorizations may have additionally affected ethnic consciousness (Anderson, 1991; Kertzer & Arel, 2002), such ethnic malleability was restricted too: while unifying German dialects into one self-conscious group was possible, more salient and sticky linguistic divides between mutually unintelligible languages were very difficult, if not impossible, to alter, invent, or make disappear.

As with all data on ethnic demographics, the political importance and potential manipulation of ethnic maps could bias our analysis. Our empirical strategy to test for and mitigate such biases consists of five components.

First, we carefully screened our map material. Starting with over 350 maps, we selected 73 maps based on high quality and spatial precision, and the absence of obvious political bias (SI Section C.1, p. 9). Drawn by 64 authors from 18 nationalities, the maps cover various parts of Europe at different points in time using sometimes diverging categorizations of ethnic groups. Second, our spatial graph G is coarse with a resolution of 100 km and up to 200 km in a robustness check. Most differences between and likely manipulations of ethnic maps affect much smaller areas (see Figure 2).

Third, we average ethnic settlement patterns across all maps from a given period, reducing the impact of potential biases on any one map. Additionally, we find no "outlier maps" when re-estimating our main models for each map separately. Fourth, we show that our

[10] See Branch (2013) for parallel consequences of mapping states.

[11] For example, Kertzer and Arel (2002) note that Greek, Serbian, and Bulgarian nationalists used alternative linguistic criteria to justify claims on parts of Macedonia.

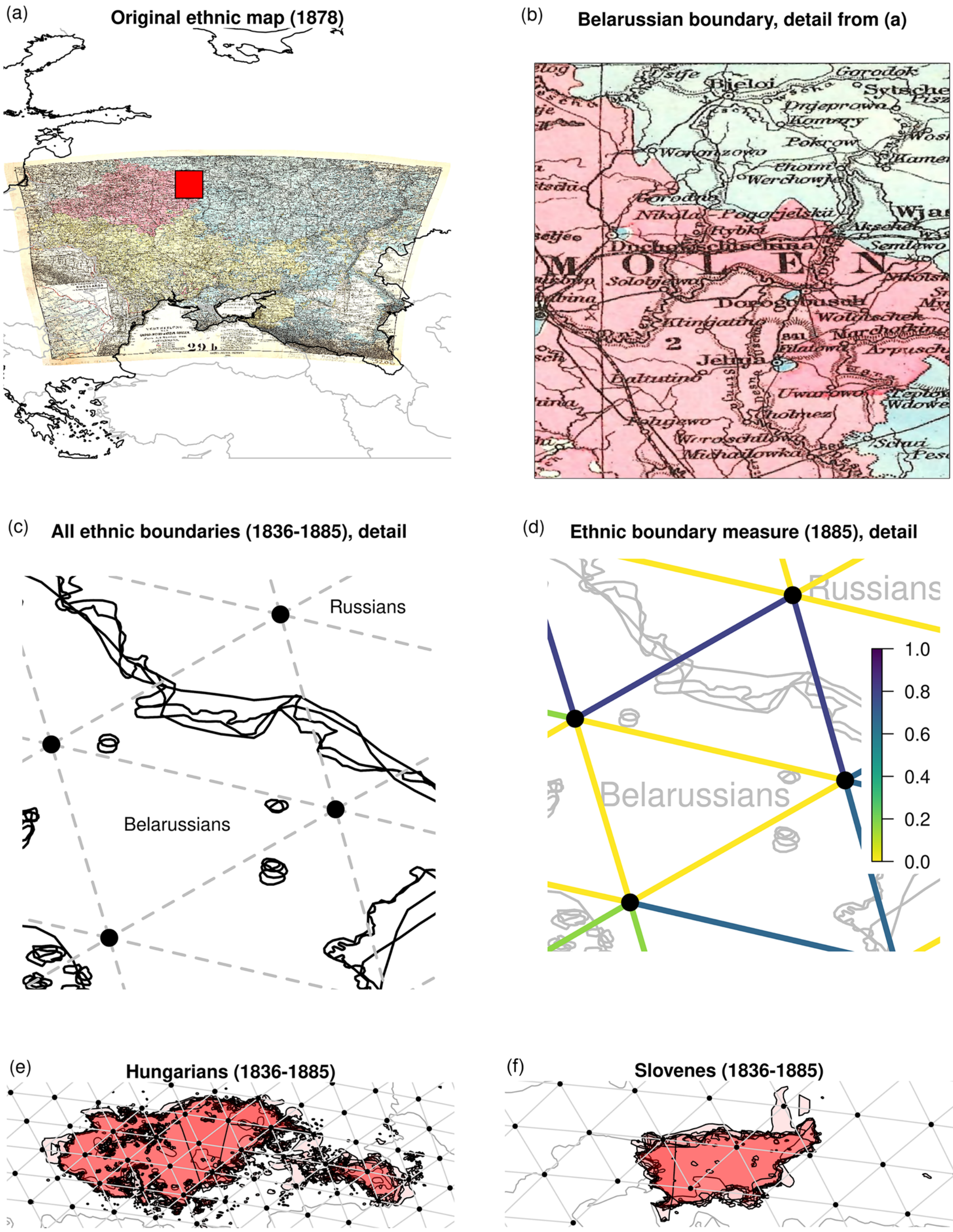

FIGURE 2 Constructing ethnic boundary from historical ethnic maps. *Note*: (a)–(d) show the transfer of ethnic settlement data onto graph *G*, with (b) to (d) depicting the area marked in red in (a). (e) and (f) show Hungarian and Slovenian settlement areas from multiple maps.

results are robust to exclusively using pre-1886 ethnic boundaries to explain changes of state borders between 1886 and 2011. This severely limits potential reverse causality, as well as strategic map manipulations during the World Wars. Fifth, we employ linguistic distances and ethnodemographic Austro-Hungarian census data as two alternative, continuous measures of "ethnic distance" to address remaining concerns of political biases. Discussed below, our results are robust across all tests.

We construct our main independent variable ethnic boundary as the proportion of maps from a given period in which an edge crosses an ethnic boundary:

$$\text{ethnic boundary}_{j,k,t} = \frac{1}{M_{j,k,t}} \sum_{m=1}^{M_{j,k,t}} \mathbb{1}_{g_{m,j} \neq g_{m,k}}, \quad (1)$$

where j and k are an edge's constitutive nodes observed in year t. $M_{j,k,t}$ denotes the set of maps that cover the geographic location of j and k in one of the 50 years prior to t. The variable ethnic boundary$_{j,k,t}$ is the simple arithmetic mean of the map-level indicators that are 1 if a map m shows nodes j and k in different ethnic settlement areas and 0 otherwise.[12]

MODELING AND ESTIMATION

We start from the intuition that the partitioning of space into states results from "attractive" and "repulsive" forces active between different locations. These forces correspond to factors that affect border formation, such as a river or an ethnic boundary separating two locations. If two points attract each other, they are likely part of the same state. If pushed apart by repulsive forces, they may become divided by a border. Each point is attracted to or repulsed by multiple neighboring points, but can only be part of one state. Directly capturing spatial dependence by only allowing for contiguous and nonoverlapping state territories, a point's ultimate state "membership" is therefore the probabilistic result of the interplay of the attraction and repulsion exerted by and among all its neighbors and their state memberships.

Our PSPM captures this logic by modeling the partitioning of a planar graph. The model allows us to estimate the attractive or repulsive forces resulting from attributes of the graph's edges. When estimating the effect of ethnic differences on state borders, we can thus account for covariates that influence ethnic settlement patterns and state borders. We next present and validate the PSPM and then introduce our empirical strategy to test our theoretical argument.

Probabilistic spatial partition model

We model state territories as contiguous and mutually exclusive clusters of nodes (partitions) of graph G introduced above. Our modeling objective is to estimate the magnitude and uncertainty of the effects of edge-level attributes while accounting for spatial dependencies in the graph. We here present the models' fundamentals, explain our approach to estimation and uncertainty, and validate the results with Monte Carlo experiments. SI Section A (pp. 1–4) contains all further details.

The model: We model the distribution over all possible partitionings P of lattice G as a Boltzmann distribution:

$$Pr(P = p_i) = \frac{e^{-\epsilon_i}}{\sum_{i=1}^{|\mathbb{P}|} e^{-\epsilon_i}}, \quad (2)$$

where the realization probability of partitioning p_i decreases with its *energy* ϵ_i. The term energy reflects the origin of the Boltzmann distribution in modeling the condition of a system in statistical mechanics (e.g., Park & Newman, 2004).[13] Because systems typically move toward a low energy, low-energy partitionings have higher probabilities. Applied to the partitioning of space into states, we can interpret the energy ϵ_i as the sum of inter- and intrastate tensions that result from a given partitioning.

Figure 3 illustrates this intuition for a simple graph of four vertices. The plot maps five (out of 12 possible) partitionings, with "countries" shown as nodes' color and number. Solid edges run within country borders and dashed ones across them. The top and bottom edges span across the red boundary between two ethnic groups, while the top and left edges cross the blue river. For illustrative purposes, we assume that political tensions ϵ result when states are too small (b, d), multiethnic (a, c), or divided by the river (a, e). Intuitively, Equation (2) holds that partitionings with ubiquitous tensions on the left have a lower probability than those with less tension to the right. Note also the spatial consistency of the graph. We cannot, for example, switch the left edge in (a) from dashed to solid since this would make the partitioning intransitive.

We assume that a partitioning's total energy ϵ_i is determined by the sum of realized energies of the edges that connect all first-degree neighbor node pairs L on the lattice:[14]

$$\epsilon_i = \sum_{j,k \in L} \epsilon_{j,k} * s_{j,k}, \quad (3)$$

whereby the potential energy $\epsilon_{j,k}$ of the edge between nodes j and k is realized if j and k are part of the same partition ($s_{j,k} = 1$, solid lines in Figure 3) and is not realized if they are part of different partition ($s_{j,k} = 0$, dotted lines in Figure 3). Our empirical interest focuses on the determinants of each edges' potential energy:

$$\epsilon_{j,k} = \beta_0 + \beta\, \mathbf{x}_{j,k}, \quad (4)$$

which defines the potential energy ϵ of the edge between nodes j and k as the sum of a constant β_0 that captures the baseline repulsion between nodes and edge-level characteristics $\mathbf{x}_{j,k}$ weighted by the parameter vector β. In our case and as discussed in the next section, $\mathbf{x}_{j,k}$ includes the indicator ethnic boundary$_{j,k}$ and additional edge-level covariates. While we have

[12] Where settlement areas overlap, we compute the share of nonmutual groups in j and k.

[13] The PSPM can be reformulated as an Exponential Random Graph Model, where $P(Y = y_i)$ is the probability of the realization of subgraph y_i of lattice G where y_i exclusively connects members of the same partition.

[14] More complex total energy functions could account for higher level predictors working, for example, at the level of emerging partitions (e.g., their size) or the partitioning as a whole (e.g., number of partitions or their size distribution).

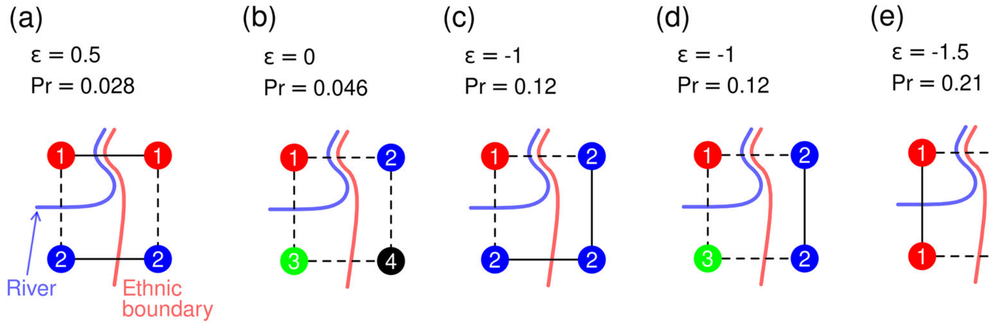

FIGURE 3 Illustration of the PSPM. *Note*: See main text for discussion. For illustrative purposes, we set parameters as $\beta_0 = -1$; $\beta_{\text{ethnic boundary}} = 1$, $\beta_{\text{river}} = 0.5$. The potential energy of each edge (from top, clockwise) is therefore .5, −1, 0, and −.5 (Equation 5).

manually set the β parameters in Figure 3 for illustrative purposes, our empirical goal is to estimate them from the observed partitioning of Europe.

Because the realization probability of a partitioning decreases with its total energy (Equation 2), coefficient estimates can be interpreted as follows: Variables associated with a positive estimate exert a *repulsive* force on nodes and increase the probability of them ending up in different partitions. Those with a negative estimate exert an *attractive* force, decreasing the chance that a border separates two points.

Applied to Figure 3 where we have manually set $\beta_{\text{ethnic boundary}} > \beta_{\text{river}}$, this means that ethnically aligned state territories have the highest probability (Panels d and e). Borders along the river in Panel (c) have a reduced probability. Finally, because of a baseline attraction between nodes ($\beta_0 < 0$), partitionings with many small countries have a low likelihood (Panels b and d).

Because edges' values of $s_{j,k}$ are interdependent, it is difficult to interpret coefficients directly. This holds except for *bridge edges* that connect two otherwise disjoint network parts (i.e., a peninsula with the continent) and can therefore independently switch $s_{j,k}$ without violating transitivity. For these edges, we can interpret coefficient estimates as in a logistic regression model, computing odds ratios, predicted probabilities, and marginal effects (see also Cranmer & Desmarais, 2011, p. 73).

Estimation and uncertainty: We estimate the β-parameters in Equation (4) using a maximum composite likelihood approach (Lindsay, 1988). Here, the likelihood function is the product over the conditional probabilities of vertices' observed partition memberships, defined based on their neighbors' memberships. We implement a Gibbs sampler that follows this logic to sample from the set of possible partitionings $|\mathbb{P}_G|$ of graph G, given edge-level predictors $\mathbf{x}_{i,j}$ and known parameters β. The sampler allows us to derive standard errors from a parametric bootstrap.[15]

Validation: We test the validity of inferences drawn from our model in an extensive series of Monte Carlo experiments presented in detail in SI Section B (pp. 4–8). Our estimator is asymptotically unbiased in the size and number of independent networks across varying β parameter combinations, and parametric bootstrapping produces consistent frequentist uncertainty estimates. SI Section D.8 (p. 25) compares the PSPM with a benchmark that disregards spatial dependence, showing that the latter produces upwards-biased and overconfident estimates.

Empirical strategy

To test our main hypothesis, we estimate the effect of ethnic geography on the partitioning of our spatial lattice G_t into states specifying the edge-level energy function as:

$$\epsilon_{j,k,t} = \beta_0 + \beta_1 \text{ ethnic boundary}_{j,k,t} + \gamma \mathbf{X}_{j,k}, \tag{5}$$

where β_0 is the baseline repulsion between nodes and ethnic boundary$_{j,k,t}$ captures whether the nodes of an edge are located in different ethnic settlement areas (Equation 1 above). To avoid bias from omitted spatial features, $\mathbf{X}_{j,k}$ must capture factors that cause ethnic as well as state borders. We therefore include time-invariant indicators for the length of each edge in kilometers, the size of the largest river[16] and watershed[17] crossed by an edge, and the mean

[15] See SI Section A.2 (p. 2).
[16] Based on a river scale in the Natural Earth data: https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-rivers-lake-centerlines/. SI Section D.3 (p. 18) shows robustness to nonlinear river effects.
[17] We derive an ordinal variable from Pfaffstetter watershed codes (Lehner et al., 2008).

elevation[18] along it. Taken together, these covariates capture important geographic causes of ethnic geography and state borders (e.g., Kitamura & Lagerlöf, 2020). We scale all variables to range between 0 and 1 to ensure coefficients' comparability.

Our second analysis uses a lagged dependent variable (LDV) model to test whether ethnic boundaries affect border *change* such that both become increasingly congruent and address reverse causality as the main inferential threat affecting the baseline model. If ethnic settlement patterns results from identity formation within state borders (e.g., Hobsbawm, 1990) the estimate of β_1 in Equation (5) could be systematically biased. We therefore account for past borders leaving ethnic boundary to affect only border change:

$$\epsilon_{j,k,t} = \beta_0 + \beta_1 \text{ ethnic boundary}_{j,k,t-1} + \beta_2 \text{ state border}_{j,k,t-1} + \beta_3 \text{ deep lag}_{j,k} + \gamma \mathbf{X}_{j,k}, \quad (6)$$

where we model edges' potential energy in period t as depending on ethnic and state borders 25 years earlier in $t-1$. In other words, to explain state borders in 1936, we control for state borders in 1911 and construct ethnic boundary$_{j,k,t-1}$ from ethnic maps drawn between 1860 and 1910. Because ethnic boundaries are measured in data from the 50 years preceding the lagged dependent variable (Equation 1), border change between $t-1$ and t cannot impact ethnic boundary$_{j,k,t-1}$. This avoids bias from reverse causality. SI Section D.1 (p. 14) shows robustness to interacting controls with state border$_{j,k,t-1}$ to differentiate between border emergence and persistence and to modeling duration dynamics.

Furthermore, borders in the deep historical past may have caused ethnic boundaries and may form precedents for "new" borders (Abramson & Carter, 2016; Simmons, 2005). To avoid such omitted variable bias, we add a "deep lag" of state borders, the share of years in which an edge crosses a border in AD 1100, 1200, …, 1600, and 1790.[19] Because we lack early-19th century ethnic maps, we cannot estimate the LDV model for the 1886 outcome data.

We first estimate our baseline and LDV models on the pooled sample of all periods. In a second step, we estimate separate models for each period to gauge temporal variation ethnic geography's effects. Throughout, we use a parametric bootstrap to derive confidence intervals.[20]

[18] From Hastings et al. (1999).
[19] Data are from Abramson (2017) and stop in 1790.
[20] SI Section D.7 (p. 25) shows robustness to varying burn-in rates of the underlying Gibbs sampler.

TABLE 1 Determinants of state borders in Europe, 1886–2011.

	Baseline	Lagged dependent variable
Constant	−2.50*	−3.01*
	[−3.04, −1.91]	[−3.98, −2.47]
Ethnic boundary$_t$	1.22*	
	[1.06, 1.40]	
Ethnic boundary$_{t-1}$		1.02*
		[0.79, 1.24]
State border$_{t-1}$		1.65*
		[1.46, 1.96]
Deep lag		0.74*
		[0.36, 1.15]
Number of periods	6	5
Number of vertices	6769	5412
Number of edges	17923	14243
Number of states	189	177
Controls	Yes	Yes

Note: Each period t has a length of 25 years; 95% confidence intervals from parametric bootstrap in parenthesis.
*Statistically significant at the 95% level.

RESULTS

Overall, we find consistent support for our theoretical argument with a strong correlation of ethnic boundaries with state borders in the baseline model. Moreover, we find similarly sized effects in our LDV model: Even when accounting for current and past political borders, ethnic boundaries are strongly and positively related to the formation of new borders over the next 25 years.

Main results: Table 1 presents the main results obtained from estimating the baseline and LDV models on the pooled data. The findings support our theoretical argument and corroborate further predictions from the broader literature. The negative constant shows that the nodes in our lattice are generally *attracted* to each other when we set all covariates to zero. This attraction is mitigated by our independent variables.

First, the coefficient of (lagged) ethnic boundaries is positive: Nodes separated by an ethnic boundary repulse each other and likely become separated by state borders. The respective effect is only slightly larger in the baseline than in the LDV model, which accounts for past borders and their determinants. The baseline estimates are thus not simply driven by reverse effects of state borders on ethnic geographies and omitted variables that affect both. Importantly, the effects of ethnic boundaries are sizeable. They are associated with almost two thirds of the energy attributed to a lagged state border. Consistent with the

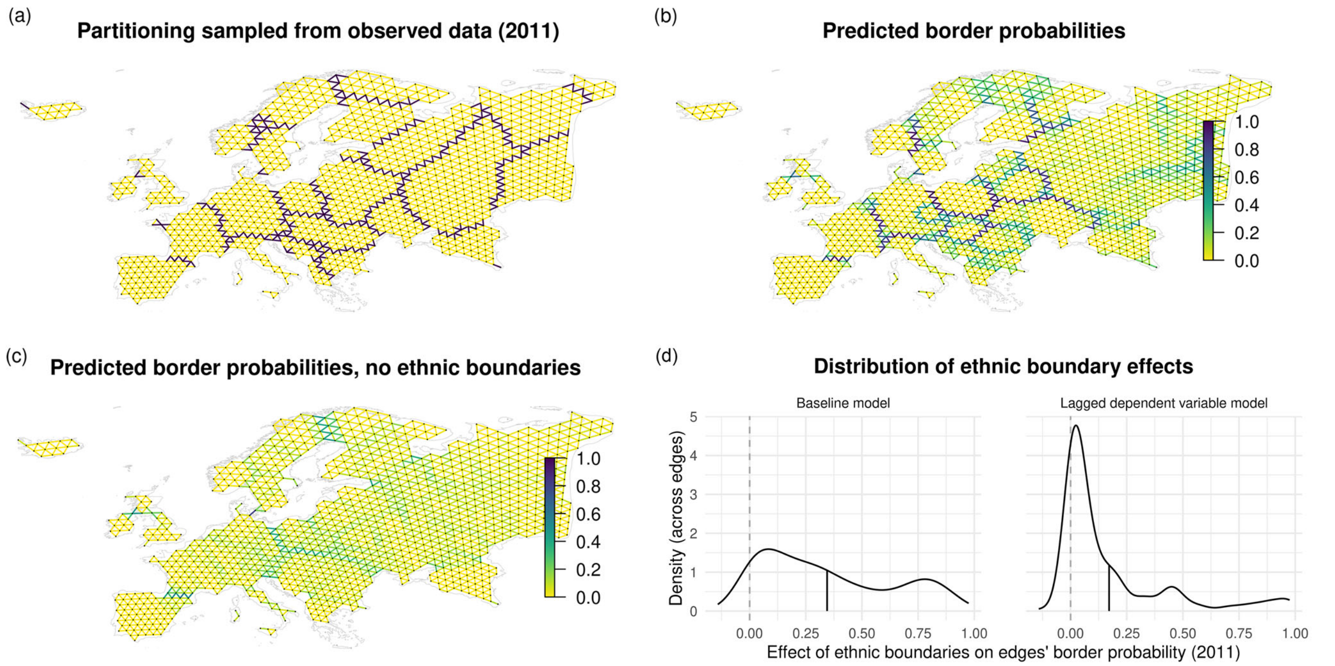

FIGURE 4 Effect of ethnic boundaries on edges' predicted border probability. *Note*: Panel (a) plots one partitioning sampled from the observed data. (b) shows edge's average border probability across 120 partitionings sampled from the observed data (2011). (c) shows border probabilities derived from data without ethnic boundaries. (d) shows the distribution of the effect of all ethnic boundaries across edges. Computations in (a)–(c) are based on the baseline model.

prevalence of secessionist border change since 1886, we find that ethnic boundaries affect the emergence of new borders more than the stability of old ones (SI Section D.1, p. 14).

Consistent with the findings by Abramson and Carter (2016), the LDV model shows that state borders from between the 10th- and 18th-century continue to separate nodes after 1886 conditional on ethnic geography. Shown in SI Section D.4 (p. 21), estimated effects of natural border determinants support previous arguments. Large watersheds and rivers, but not high altitudes are likely to divide locations into different states, in particular at a high spatial resolution and without conditioning on posttreatment ethnic boundaries and historical state borders.

Interpretation of effect sizes: Table 1 says little about the estimated absolute effect of ethnic boundaries on state borders. As discussed above, we can interpret the coefficients in parallel to those of a logistic regression for edges that bridge otherwise disjoint parts of the lattice and are therefore independent. For these bridge edges, the coefficient of ethnic boundary implies an odds ratio of 3.4 [2.9, 4.0][21] for the baseline model. Holding all covariates at their median values, an ethnic boundary thus leads to an increase in the probability of crossing a state border from 11.2% [9.7, 12.4] to 29.9% [27.6, 33.0]. The LDV model yields an odds ratio of 2.8 [2.2, 3.4] and a change in the border probability from 6.1% [4.6, 7.9] to 15.3% [11.0, 19.4].[22] These substantial effects constitute a lower bound to the effects of ethnic boundaries, which increase as they cross multiple interdependent edges.

For the more common case of *interdependent* edges, we use our estimates to sample 120 partitionings of the type plotted in Figure 4a and compute predicted border probabilities as the fraction of partitionings in which an edge crosses a border. The joint effect of all ethnic boundaries can be assessed by sampling two types of partitionings. The first type is sampled from the observed data in 2011 (Figure 4b). The second, counterfactual type is sampled assuming that all of Europe belongs to the same ethnic group[23] but holding all other covariates at their observed values (4c).

Predicted probabilities based on observed data in 4b overall closely resemble Europe's political map. Portugal is a prominent false negative, likely due its small size, narrowness, and rivers and watersheds that cross it. In the Balkans, diffuse border probabilities reflect overlapping ethnic settlement areas. Lastly, false positives cross Switzerland, a state that defies ethnically aligned borders. Yet, a comparison to Panel 4c shows that incorporating ethnic boundaries greatly improves our prediction, increasing the area under the ROC curve from 63% to 88%.

[21] Ninety-five percent CI in parentheses.

[22] This change is conditional on no border in $t-1$, hence the lower probability.

[23] That is, setting all ethnic boundaries to zero.

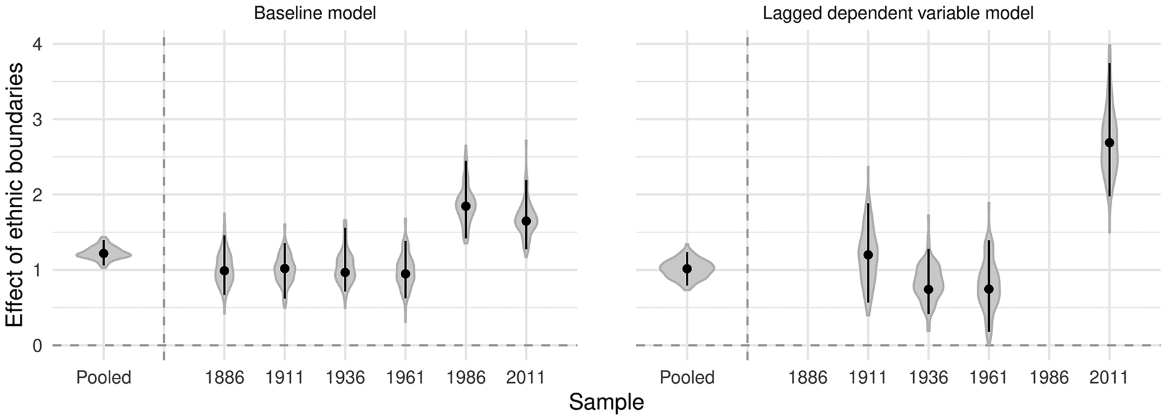

FIGURE 5 Effect of ethnic boundaries on the partitioning of Europe into states. *Note*: Ninety-five percent CIs and gray areas show the distribution of bootstrapped estimates.

The difference between border probabilities in Panels 4b and 4c constitutes the joint effect of all observed ethnic boundaries, shown in Panel 4d. Being larger than the bridge-edge effects discussed above, ethnic boundaries increase border probabilities by 34 percentage points in the baseline model. In the LDV model, border probabilities increase by 17 percentage points over a relatively small baseline probability of border change. In sum, these results confirm a substantial effect of ethnic boundaries on the location of (newly drawn) state borders.

Variation over time: Figure 5 sheds light on temporal dynamics by showing separate estimates for each 25th year since 1886. Consistent with our argument, the baseline association between state borders and ethnic boundaries increases over time. The temporally disaggregated LDV models show that ethnic geography affected *changes* in state borders particularly around the turn of the 19th century, World War I, and between 1986 and 2011 when the Soviet Union and Yugoslavia collapsed.[24] World War II brought slightly lesser ethnic alignment of state borders, and borders were stable from 1961 to 1986. In sum, systemic instability comes with nationalist border change (cf., Skocpol, 1979; Abramson & Carter, 2021).

ROBUSTNESS CHECKS

Our robustness checks assess whether the main findings are driven by potentially endogenous changes in ethnic geography, the choice of data on ethnicity and control variables, or the spatiotemporal data structure. SI Section D (pp. 14–26) presents all details.

[24] Post-Soviet and post-Yugoslav borders mostly followed administrative borders often drawn based on ethnic geography (Hirsch, 2000).

Pre-1886 ethnic boundaries: Political biases may affect in particular ethnic data produced during the World Wars. In addition, our results could be biased by omitted factors that first changed ethnic settlement patterns and, temporarily lagged, correlated border change. As a remedy, we use ethnic boundaries observed in the 50 years prior to 1886 as time-invariant predictor. The estimates in Figure 6 show effects of historical ethnic boundaries that are only marginally smaller than our baseline estimates. We also find an increasing alignment of state borders to ethnic boundaries in the LDV models. Reaffirming the absence of reverse causality and providing evidence against political bias, the LDV results show that pre-1886 ethnic boundaries affected border change even a century later. These results hold when we account for subnational regional borders in 1800 and 1900 (SI Section D.3, p. 18).

Alternative measures of ethnic difference: We further test robustness regarding three alternative measures of edge-level ethnic differences (SI Section D.2, p. 15). First, estimating our main specification for each ethnic map yields a smooth estimate distribution without "outlier maps" and evidences no undue influence of any one map. Second, we inquire whether effects of ethnic maps on ethnic identities may have caused our results. Such effects would most likely arise between linguistically close groups, yet our estimates increase with the linguistic distance between groups. Third, politically biased ethnic maps may result from manipulated population thresholds. Using continuous census data on compositional ethnic differences between districts in pre-WWI Austria–Hungary yields stronger and more precise results, likely due to more precise measurement. In sum, we find no evidence that political or other biases from our historical ethnic maps affect our results.

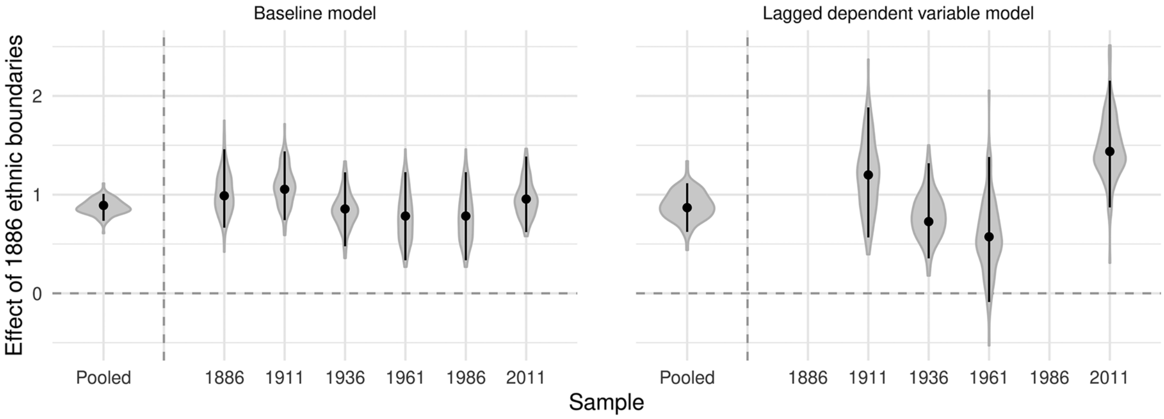

FIGURE 6 Effect of pre-1886 ethnic boundaries on the partitioning of Europe into states. *Note*: Ninety-five percent CIs and gray areas show the distribution ofbootstrapped estimates.

Control variables: Our main results are insensitive to re-estimating models without controls or extending them to account for ruggedness, population density, the edges' geographic orientation, and nonlinear river effects, as well as administrative borders in 1800 and 1900.

Variation of the data structure: We find that our results are robust to varying the length of periods t between 5 and 65 years.[25] We also vary the spatial data structure regarding (1) the graph's exact location, (2) its spatial resolution, and (3) its connectivity structure. For each variation, estimates remain statistically and substantially significant and similar to the baseline results. As additional evidence against potential bias from ethnic maps that are erroneous or manipulated, effects *increase* with coarser networks in which spatial error becomes less relevant and manipulation less likely.

In sum, our robustness checks show that the main results are not due to either endogenous changes in ethnic boundaries over time or potentially arbitrary modeling decisions of ours. The consistency of the results with early and alternative ethnic data as well as coarse spatial networks suggests the absence of substantive bias from political manipulation of ethnic data. In the next section, we provide evidence on secessionist claims and conflicts as an important mechanism through which ethnic geography shapes state borders in the age of nationalism.

MECHANISM: SECESSIONIST CLAIMS AND CONFLICT

Because there are more potential ethnic nations than realized states, secessionism likely drives much of the border-changing effects of nationalism. In an auxiliary analysis in SI Section E (pp. 28–30), we find that ethnically distinct peripheral areas were more likely experience ethnic secessionism since 1946. For this analysis, we recur to the vertices of our spatial network as units of analysis. For each year, we code whether a point became part of a secessonist claim (Germann & Schvitz, 2023), was settled by a politically relevant ethnic group associated with an onset of secessionist civil war (Vogt et al., 2015), and became part of a newly independent state (Schvitz et al., 2022). We model the effect of coethnicity of the point with its state's capital on these outcomes using a Cox proportional hazard model, which mitigates the problem of successful secession leading to selection out of the treatment group.

We find large and statistically significant effects of being ruled from a non-coethnic capital on demands for and realizations of secession. Over 50 years and holding covariates at their median value, Figure 7 shows that ethnically distinct regions have a probability of 35% to be part of a claimed, violently pursued (14%), or realized border change (41%). The respective probabilities for coethnic areas are close to zero. While the break-up of the USSR and Yugoslavia dominate the temporal pattern of secessions, our results hold when we stratify by country-year. In sum, they show that ethnic secessions drive the alignment of state borders with the ethnic map.

GLOBAL COMPARISON

We finally analyze the generalizability of our findings beyond 19th- and 20th-century Europe by comparing the state-shaping effects of ethnic geography across continents. To do so, we create spatial lattices for each continent and use our main PSPM specifications to estimate the effect of ethnic boundaries on state

[25] Sixty-five years is the maximum period length that produces at least two periods.

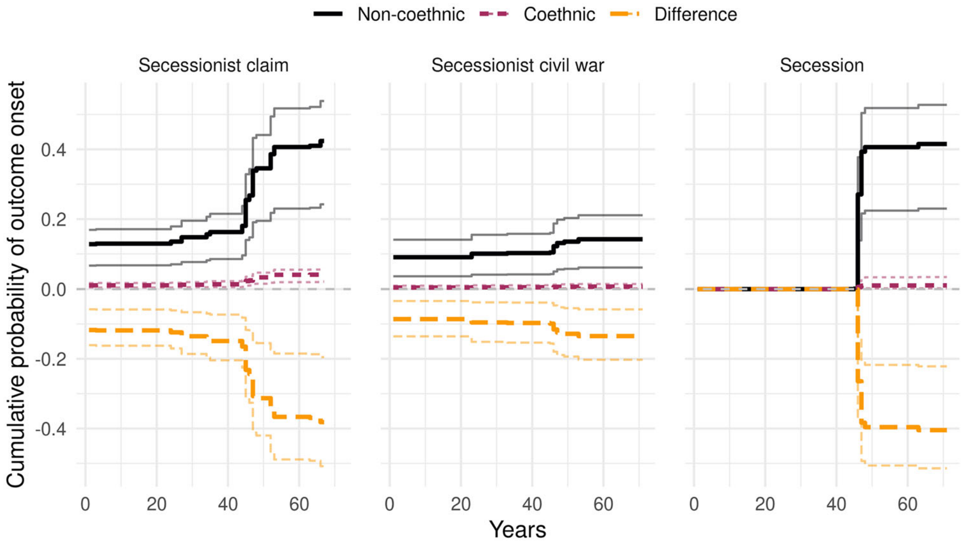

FIGURE 7 Effect of ethnic boundaries on secessionism. *Note*: Predictions with ninety-five percent CIs based on Models 1, 3, and 5 in SI Table A6, p. 29, setting covariates to median values.

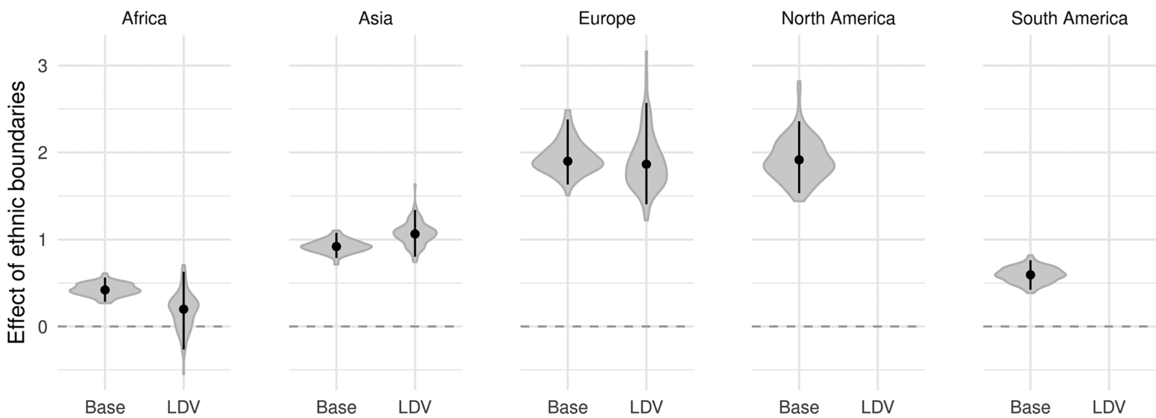

FIGURE 8 Effect of ethnic boundaries in 1964 on state borders across continents. *Note*: Ninety-five percent CIs and gray areas show the distribution of bootstrapped estimates.

borders in 2017. We use the earliest global data on ethnic geography from the 1963 Soviet *Atlas Narodov Mira* (Weidmann et al., 2010) and control for 1964 state borders in the LDV model.[26]

Starting with Africa, the results in Figure 8 support the conventional wisdom that decolonization and the *uti possedetis* norm preserved colonial borders drawn with little reference to ethnic geography (Griffiths, 2015; Michalopoulos & Papaioannou, 2016). The baseline coefficient is relatively small (yet statistically significant, see also Paine et al., 2021) and the LDV result shows no significant effect on border changes since 1964. Ethnic boundaries have had a more substantive effect on Asian borders. Though “only” half the size compared to Europe, ethnic boundaries significantly correlate with borders in 2017 and with post-1964 border change, mostly driven by Soviet Republics' independence. Lastly, we observe a stronger cross-sectional correlation between ethnic and state boundaries in North than in South America. The absence of recent border change prohibits estimating LDV models. In an auxiliary test in SI Section D.9 (p. 26), we find that ethnic boundaries have a larger effects on border change in densely populated regions in Europe and globally, suggesting that the nationalist reshaping of states occurs mostly where territory is of high value and competed over (cf. Herbst, 2000).

In sum, these results yield two insights. First, state borders are cross-sectionally aligned with ethnic boundaries at a global scale, with states in Africa showing the least alignment. Second, ethnic boundaries seem to affect border change in Asia and Europe but not elsewhere. Ongoing ethnonationalist

[26] Lacking global data, we omit the “deep lag.”

conflicts from secessionist Kurdistan to border disputes between India and Pakistan suggest an ongoing risk of ethnic reshaping of Asian states. In contrast, outright secessionist conflict is rare in Africa where the territorial integrity norm is generally upheld (Englebert & Hummel, 2005; Zacher, 2001), low population densities decrease territorial competition (Herbst, 2000), but ethnic conflict fragments some states internally.

CONCLUSION

Assessing nationalism's impact empirically, this study has analyzed whether, by how much, and how the nationalist principle reshaped European states along ethnic boundaries since 1886. Bringing systematic evidence to bear, we contribute to the literature on state and border formation, which has so far been relatively fragmented as regards the ethnic origins of the partitioning of geographic space into states.

Theoretically, we have drawn on a rich yet mostly qualitative literature that highlights the impact of nationalism on international borders through secession and, less frequently, unification and irredentism. Over time, these processes gradually aligned state borders with the ethnic map. We have tested this proposition with new spatial data on ethnic settlement patterns since 1855 and our new PSPM, a statistical method that allows us to estimate the effect of ethnic geography on the partitioning of Europe into states.

While developed for this study, the PSPM can be adapted to study other partitionings such as administrative units or electoral districts. To improve its flexibility, future developments could focus on supra-edge predictors, different samplers, compositional membership outcomes, computational efficiency, and statistical properties. Lastly, innovative modelers may want to jointly assess the reciprocal relationship between state borders and ethnic geography, thus moving beyond the partial effects estimated here.

Our empirical results show that ethnic boundaries substantively affected borders and border change since 1886. We estimate that an ethnic boundary between two locations increases the likelihood of an interstate border between them by 34 percentage points. Conditional on past state borders, ethnic boundaries increase border probabilities by 17 percentage points. Showing that secessions drive the ethnic reshaping of states, we find that peripheral ethnic minorities are at high risk to be subject to secessionist claims, conflict, and final break away. Our results also suggest that the ethnic alignment of state borders is ongoing macrohistorical process. The Russian invasion of Ukraine and secessionist demands across Europe underscore the continuing influence of nationalist revisionism. Looking beyond Europe, we have found similar dynamics of ethnonationalist border change in Asia but less so elsewhere.

In sum, we find that ethnic geography has an important and continuing impact on the shape of European states. In consequence, the common treatment of states (and other political units) as fixed and exogenous entities comes at the risk of selection and reverse causality biases. Selection bias might, for example, deflate estimated effects of ethnopolitical exclusion on conflict (Cederman et al., 2013) if previous secessions caused lower levels of ethnic exclusion and conflict. Reverse causation might inflate estimated effects of ethnic diversity on economic performance (Alesina & Ferrara, 2005) if economic development sparked centripetal *and* centrifugal nationalism (Gellner, 1983, chapter 7), secessions, and thus lower ethnic diversity. Knowing about units' origins is therefore an important prerequisite to inferring the consequences of at least some of their attributes.

Our analysis of post-1886 Europe being primarily structuralist, we caution against deterministic extrapolations. While the *potential* of ethnic centrifugal forces merits full recognition, previous research offers perspectives on how to contain them through ethnic power-sharing and regional accommodation (Cederman et al., 2013). More radical if perhaps utopian, dissociating states from nations altogether may succeed in depoliticizing ethnic divides (Mamdani, 2020). Internationally, territorial integrity norms could rein in nationalist excesses (Zacher, 2001), even though the recent revival of nationalist forces could endanger such progress.

ACKNOWLEDGMENTS

We thank the editor and three anonymous reviewers for their valuable comments and suggestsion and Nils-Christian Bormann, Michael Kenwick, Melissa Lee, Kan Li, Paul Poast, and seminar and conference participants at APSA 2020, the University of Oxford, London School of Economics and Political Science, Harvard University, and the Perry World House Borders and Boundaries Conference, as well as members of the International Conflict Research group for their helpful feedback and Nicole Arnet, Camiel Boukhaf, Nicole Eggenberger, Benjamin Füglister, Sebastian Gmüer, Irina Siminichina, Tim Waldburger, Benjamin Wallin, and Roberto Valli for their invaluable research assistance. We are very grateful to Philipp Hunziker who provided crucial inputs to this work. All remaining errors are ours. We acknowledge generous financial support from the Advanced ERC Grant 787478 NASTAC, "Nationalist State Transformation and Conflict."

REFERENCES

Abramson, Scott, and David B. Carter. 2021. "Systemic Instability and the Emergence of Border Disputes." *International Organization* 75(1): 103–46.

Abramson, Scott F. 2017. "The Economic Origins of the Territorial State." *International Organization* 71(1): 97–130.

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

Alesina, Alberto, and Eliana La Ferrara. 2005. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature* 43(3): 762–800.

Alesina, Alberto, and Enrico Spolaore. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economics* 112(4): 1027–56.

Alesina, Alberto, and Enrico Spolaore. 2005. *The Size of Nations*. Cambridge, MA: MIT Press.

Anderson, Benedict. 1991. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. 2nd ed. London: Verso.

Beissinger, Mark R. 2002. *Nationalist Mobilization and the Collapse of the Soviet Union*. Cambridge: Cambridge University Press.

Branch, Jordan. 2013. *The Cartographic State: Maps, Territory, and the Origins of Sovereignty*. Vol. 127. Cambridge: Cambridge University Press.

Zeynep Bulutgil, H.. 2015. "Social Cleavages, Wartime Experience, and Ethnic Cleansing in Europe." *Journal of Peace Research* 52(5): 577–90.

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

Cadiot, Juliette. 2005. "Searching for Nationality: Statistics and National Categories at the end of the Russian Empire (1897-1917)." *Russian Review* 64(3): 440–55.

Carter, David B., and Hein E. Goemans. 2011. "The Making of the Territorial Order: New Borders and the Emergence of Interstate Conflict." *International Organization* 65(2): 275–309.

Cederman, Lars-Erik, Kristian Skrede Gleditsch, and Halvard Buhaug. 2013. *Inequality, Grievances, and Civil War*. Cambridge: Cambridge University Press.

Cederman, Lars-Erik, Seraina Rüegger, and Guy Schvitz. 2022. "Redemption through Rebellion: Border Change, Lost Unity and Nationalist Conflict." *American Journal of Political Science* 66(1): 24–42.

Cederman, L.-E., Girardin, L., & Müller-Crepon, C. (2023). "Nationalism and the Puzzle of Reversing State Size." *World Politics*, 75(4), 692–734.

Coggins, Bridget. 2014. *Power Politics and State Formation in the Twentieth Century: The Dynamics of Recognition*. Cambridge: Cambridge University Press.

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

Cranmer, Skyler J., and Bruce A. Desmarais. 2011. "Inferential Network Analysis with Exponential Random Graph Models." *Political Analysis* 19(1): 66–86.

De Luca, Giacomo, Roland Hodler, Paul A. Raschky, and Michele Valsecchi. 2018. "Ethnic Favoritism: An Axiom of Politics?" *Journal of Development Economics* 132(May): 115–29.

Desmet, Klaus, Michel Le Breton, Ignacio Ortuño-Ortín, and Shlomo Weber. 2011. "The Stability and Breakup of Nations: A Quantitative Analysis." *Journal of Economic Growth* 16(3): 183.

Englebert, Pierre, and Rebecca Hummel. 2005. "Let's Stick together: Understanding Africa's Secessionist Deficit." *African Affairs* 104(416): 399–427.

Englebert, Pierre, Stacy Tarango and Matthew Carter. 2002. "Dismemberment and Suffocation: A Contribution to the Debate on African Boundaries." *Comparative Political Studies* 35(10): 1093–118.

Fagan, Moira, and Jacob Poushter. 2020. NATO Seen Favorably Across Member States. Report Pew Research Center.

Fazal, Tanisha M. 2004. "State Death in the International System." *International Organization* 58(2): 311–44.

Fazal, Tanisha M. 2007. *State Death: The Politics and Geography of Conquest, Occupation, and Annexation*. Princeton: Princeton University Press.

Fifield, Benjamin, Michael Higgins, Kosuke Imai, and Alexander Tarr. 2020. "Automated Redistricting Simulation using Markov Chain Monte Carlo." *Journal of Computational and Graphical Statistics* 29(4): 715–28.

Friedman, David. 1977. "A Theory of the Size and Shape of Nations." *Journal of Political Economy* 85(1): 59–77.

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

Germann, Micha, and Guy Schvitz. 2023. "Representing Self-Determination Claims in Space: The GeoSDM Dataset." Mimeo, University of Bath.

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

Goddard, Stacie E. 2006. "Uncommon Ground: Indivisible Territory and the Politics of Legitimacy." *International Organization* 60(1): 35–68.

Goemans, Hein E. 2006. "Bounded Communities: Territoriality, Territorial Attachment, and Conflict." In *Territoriality and Conflict in an Era of Globalization*, eds. Miles Kahler and Barbara F. Walter. Cambridge: Cambridge University Press, pp. 25–61.

Goemans, Hein E., and Kenneth A. Schultz. 2017. "The Politics of Territorial Claims: A Geospatial Approach Applied to Africa." *International Organization* 71(1): 31–64.

Griffiths, Ryan D. 2015. "Between Dissolution and Blood: How Administrative Lines and Categories Shape Secessionist Outcomes." *International Organization* 69(3): 731–51.

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

Hansen, Jason D. 2015. *Mapping the Germans: Statistical Science, Cartography, and the Visualization of the German Nation, 1848-1914*. Oxford: Oxford University Press.

Hardin, Russell. 1995. *One For All: the Logic of Group Conflict*. Princeton: Princeton University Press.

Hastings, David A., Paula K. Dunbar, Gerald M. Elphingstone, Mark Bootz, Hiroshi Murakami, Hiroshi Maruyama, Hiroshi Masaharu, Peter Holland, John Payne, Nevin A. Bryant, et al. 1999. "The global land one-kilometer base elevation (GLOBE) digital elevation model, version 1.0." *National Oceanic and Atmospheric Administration, National Geophysical Data Center,* digital data base.

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

Hechter, Michael. 2013. *Alien Rule*. Cambridge: Cambridge University Press.

Herb, Guntram Henrik. 2002. *Under the Map of Germany: Nationalism and propaganda 1918-1945*. New York: Routledge.

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

Hirsch, Francine. 1997. "The Soviet Union as a Work-in-Progress: Ethnographers and the Category Nationality in the 1926, 1937, and 1939 Censuses." *Slavic Review* 56(2): 251–78.

Hirsch, Francine. 2000. "Toward an Empire of Nations: Border-making and the Formation of Soviet National Identities." *Russian Review* 59(2): 201–26.

Hobsbawm, Eric J. 1990. *Nations and Nationalism Since 1780*. Cambridge: Cambridge University Press.

Hroch, Miroslav. 1985. *Social Preconditions of National Revival in Europe: A Comparative Analysis of the Social Composition of Patriotic Groups among the Smaller European Nations*. Cambridge: Cambridge University Press.

Kertzer, David, and Dominique Arel. 2002. Census and identity. In *The Politics of Race, Ethnicity, and Language in National Censuses*, eds. Jack Caldwell, Andrew Cherlin, Tom Fricke, Frances Goldscheider et al. Cambridge: Cambridge University Press.

Kitamura, Shuhei, and Nils-Petter Lagerlöf. 2020. "Geography and State Fragmentation." *Journal of the European Economic Association* 18(4): 1726–69.

Lehner, Bernhard, Kristine Verdin, and Andy Jarvis. 2008. "New Global Hydrography Derived from Spaceborne Elevation Data." *Eos, Transactions American Geophysical Union* 89(10): 93–94.

Lindsay, Bruce G. 1988. "Composite Likelihood Methods." *Contemporary Mathematics* 80(1): 221–39.

Livingstone, David N., and Charles W. J. Withers, eds,. 1999. *Geography and Enlightenment*. Chicago: University of Chicago Press.

Mamdani, Mahmood. 2020. *Neither Settler nor Native*. Cambridge, MA: Harvard University Press.

Manela, Erez. 2007. *The Wilsonian Moment: Self-Determinatino and the International Origins of Anticolonial Nationalism*. Oxford: Oxford University Press.

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

Michalopoulos, Stelios, and Elias Papaioannou. 2016. "The Long-Run Effects of the Scramble for Africa." *American Economic Review* 106(7): 1802–48.

Morgenthau, Hans. 1985. *Politics Among Nations: The Struggle for Power and Peace*. New York: Knopf.

Murphy, Alexander. 2002. "National Claims to Territory in the Modern State System: Geographical Considerations." *Geopolitics* 7(2): 193–214.

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

Nugent, Elizabeth R. 2020. "The Psychology of Repression and Polarization." *World Politics* 72(2): 291–334.

O'Leary, Brendan. 2001. The elements of right-sizing and right-peopling the state. In *Right-sizing the State: The Politics of Moving Borders*, eds. Brendan O'Leary, Ian Lustick, Thomas Callaghy, Thomas M Callaghy et al. Oxford: Oxford University Press, pp. 15–73.

Paine, Jack, Xiaoyan Qiu, and Joan Ricart-Huguet. 2024. "Endogenous Colonial Borders: Precolonial States and Geography in the Partition of Africa." *American Political Science Review*, forthcoming.

Palsky, Gilles. 2002. "Emmanuel de Martonne and the Ethnographical Cartography of central Europe (1917–1920)." *Imago Mundi* 54(1): 111–19.

Park, Juyong, and Mark E. J. Newman. 2004. "Statistical Mechanics of Networks." *Physical Review E* 70(6): 066117.

Petersen, Roger D. 2002. *Understanding Ethnic Violence: Fear, Hatred, and Resentment in Twentieth-Century Eastern Europe*. Cambridge: Cambridge University Press.

Ratner, Steven R. 1996. "Drawing a Better Line: Uti Possidetis and the Borders of New States." *American Journal of International Law* 90(4): 590–624.

Roeder, Philip G. 2012. *Where Nation-states Come from*. Princeton, NJ: Princeton University Press.

Sack, Robert David. 1986. *Human Territoriality: Its Theory and History*. Cambridge: Cambridge University Press.

Sahlins, Peter. 1989. *Boundaries: The Making of France and Spain in the Pyrenees*. Berkeley: University of California Press.

Sambanis, Nicholas, and Jonah Schulhofer-Wohl. 2009. "What's in a Line? Is Partition a Solution to Civil War?" *International Security* 34(2): 82–118.

Schvitz, G., S. Rüegger, L. Girardin, L.-E. Cederman, N. Weidmann, and K. S. Gleditsch. 2022. "Mapping the International System, 1886-2017: The Cshapes 2.0 Dataset." *Journal of Conflict Resolution* 66(1): 144–61.

Simmons, Beth A. 2005. "Rules Over Real Estate: Trade, Territorial Conflict, and International Borders as Institution." *Journal of Conflict Resolution* 49(6): 823–48.

Simmons, Beth A., and Michael R. Kenwick. 2021. "Border Orientation in a Globalizing World." *American Journal of Political Science*, 66(4), 853–870.

Siroky, David S., and Christopher W. Hale. 2017. "Inside Irredentism: A Global Empirical Analysis." *American Journal of Political Science* 61(1): 117–28.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.

Tilly, Charles. 1978. *From Mobilization to Revolution*. New York: McGraw-Hill.

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp M. Hunziker, and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Dataset Family." *Journal of Conflict Resolution* 59(7): 1327–42.

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

Weber, Max. 1978. *Economy and Society*. New York: Bedminster.

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): 491–99.

Weiner, Myron. 1971. "The Macedonian Syndrome an Historical Model of International Relations and Political Development." *World Politics* 23(4): 665–83.

White, George W. 2004. *Nation, State and Territory. Origins, Evolutions and Relationships*. Lanham: Rowman & Littlefield.

Wucherpfennig, Julian, Aya Kachi, Nils-Christian Bormann and Philipp Hunziker. 2021. "A Fast Estimator for Binary Choice Models with Spatial, Temporal, and Spatio-temporal Interdependence." *Political Analysis* 29(4): 570–76.

Zacher, Mark W. 2001. "The Territorial Integrity Norm: International Boundaries and the Use of Force." *International Organization* 55(2): 215–50.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: 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* 1–16. https://doi.org/10.1111/ajps.12838