3 Operationalizing State-Nation Congruence

To lay the groundwork for the empirical analysis to come, this chapter explains how we operationalize and measure state-nation incongruence. Parts II and III then test the observable implications of the theory of nationalist state transformation that we introduced in Chapter 2.

3.1 Motivating Our Approach

A comprehensive, quantitative account of how nationalism affected border change and conflict across post-Napoleonic Europe would require data and methods beyond the reach of contemporary scientific resources. One would need to collect systematic and comparable information on the rise and spread of nationalist ideas and ideologies and elite-level and popular definitions of national communities and their boundaries, as well as insights into the degree to which perceived incongruence shaped individual actors' preferences, emotions, and decisions. While in many ways desirable, such an approach is infeasible due to more than constraints in terms of time and money. To begin, much of the necessary, relevant historical data have simply not been recorded, and archival and secondary sources provide an incomplete picture.[1] High-capacity states and movements in Western Europe, successful instances of nationalist mobilization, and writing by or about a select few "great men" are likely overrepresented.[2] Feasibility aside, comprehensive data of this kind would also run into analytical challenges. Rather than causing conflict and border change, nationalist ideologies and boundaries may themselves shift in response to recent or anticipated future political

[1] But see Shelef 2016, 2020, for a pioneering attempt to systematically trace shifting definitions of national homelands based on media discourse post-1945.

[2] Modern nationalism has rarely been an exclusively elite-driven and top-down phenomenon. A full understanding of its explosive consequences needs to complement elite-level processes with a healthy dose of history given later, as, e.g., Eley 1991 and Hobsbawm 1992 have argued.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

upheaval. Such causal feedback loops could invalidate any inferences drawn from the quantitative analyses of this book.

Therefore, our strategy to (at least partially) circumvent these practical and analytical constraints is to restrict our focus to ethnic and linguistic boundaries as historically important building blocks of nationalism, especially in the European context.[3] The practical advantage of this tack is that information on linguistic affiliations and population distributions has been collected in a systematic and comparable fashion since the mid nineteenth century. The main innovations in this regard were the introduction of regular population censuses and the increasingly sophisticated spatial illustration of ethnic population distributions on ethnographic maps.[4] As we discuss in greater detail later, we rely on dozens of historical maps to derive explicitly spatial and time-varying data on ethnic boundaries. Combining ethnic boundaries with geographic information on state borders enables us to identify the regions across historical Europe where our theory expects state-nation incongruence to be perceived and acted upon.

In response to the problem of methodological statism, we assume that ethnic and linguistic boundaries are temporally more stable and harder to manipulate by self-serving political elites and cultural entrepreneurs than the salience of these and other identity markers in nationalist discourse and mobilization. If this assumption is true, using state-ethnic incongruence as a structural approximation of potential state-nation incongruence as perceived by elites and popular masses is less likely to be haunted by the issue of reverse causation described earlier. Making this assumption by no means implies a view of ethnicity as a natural, essential, and historically never-changing category. Ethnic groups, languages, and dialects are products of human history; their emergence and evolution were shaped by social, economic, cultural, and political processes in the past. The only claims we make are that some identifiable cultural raw material existed before the onset of modern nationalist politics and that such historical realities meaningfully constrained nationalist entrepreneurs in their efforts to politically activate, reshape, or even invent national boundaries. In short, we adopt a position of constrained constructivism – expecting nationalist mobilization to be more consequential where it resonates with the plausibly perceived realities of its target audience.[5]

Before we delve into details about data sources and operationalization, we note some obvious weaknesses and likely objections to this

[3] Barbour and Carmichael 2000.
[4] Hansen 2015.
[5] Snow and Benford 1988.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

approach. First, the assumption about sticky ethnic boundaries and constrained constructivism may be rejected as historically implausible and conceptually inconsistent with the dominant modernist paradigm.[6] As far as plausibility is concerned, the best available evidence from linguistic and historical scholarship suggests that many of the European languages or dialects drawn upon by later nationalisms were split from their respective roots hundreds and sometimes thousands of years before the nineteenth century.[7] Regarding compatibility with modernist theory, even the most renowned pioneers of that paradigm acknowledge the (at least occasional) relevance of "culture," "barriers to communication,"[8] and "popular proto-nationalism."[9] We make minimalist assumptions about the relevance of ethnic building blocks. In contrast to ethnosymbolism,[10] we do not require a strong sense of group consciousness and internal cohesion, nor do we assume continuity between premodern cultural symbols and practices on the one hand and modern nationalist narratives on the other. Some shared identity markers may be enough for ethnicity to matter.

Second, some readers may perceive our empirical approach to be too static and structural to address the most interesting open questions about the causes and consequences of modern nationalism. Relying on ethnic groups and their relatively sticky boundaries could be seen as standing in the way of investigating the timing of nationalist mobilization,[11] changing definitions of national communities and their boundaries, or the contextual conditions and actor constellations that make mobilization for conflict and border change most likely.[12] Fully aware of these limitations, we remain convinced that the systematic application of our simplified structural approach yields relevant insights that go beyond existing work on nationalism, conflict, and border change. In addition, several of our analyses combine structural conditions with contextual factors, which allows us to tackle issues of timing and changes of ethnic boundaries.[13]

Third, our approach is likely to simultaneously over- and underpredict some important consequences of nationalism. Many ethnic groups have failed to "bark" and never developed national consciousness.[14] At

[6] Mylonas and Tudor 2021, 2023.
[7] Bouckaert et al. 2012; Chang et al. 2015; Barbour and Carmichael 2000.
[8] Gellner 1964, 1983.
[9] Hobsbawm 1992.
[10] Smith 1986.
[11] Mylonas and Tudor 2021.
[12] Shelef and Koo 2022.
[13] See Chapters 8–10.
[14] Gellner 1983, 43.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

the same time, some varieties of nationalisms are clearly nonethnic but still have real historical "bite."[15] Strong ethnic bonds are neither necessary nor sufficient for the formation of modern nationalism, as Gellner famously argues with reference to the Estonian national awakening.[16] But accepting these premises does not imply that ethnicity is causally irrelevant or inessential for nationalism, as Gellner[17] and his most sympathetic readers sporadically seem to infer.[18] Rejecting plausible causes of nationalism based on cherry-picked examples requires strong assumptions of deterministic causation that are at odds with social realities and constructivist notions of historical contingency. Under probabilistic views of causation, general conjecture drawn from exceptionally rapid national awakenings dissolves into non sequiturs. For what it is worth, our probabilistic take on shared ethnicity as one (but not the only) root of national consciousness remains perfectly consistent with everything Gellner has to say about the Estonians.[19] While Estonians may not have had any ethnic consciousness, or even a name for their group, Estonian dialects were spoken in the Baltics a good millennium before Johann Gottfried Herder and homegrown writers like Friedrich Reinhold Kreutzwald began to activate and transform linguistic commonalities along nationalist lines.[20]

Despite these three important limitations, our structural approach seems productive when it comes to addressing the four main gaps in the nationalism scholarship identified in Chapter 1. Geographic data on ethnic and state borders capture the inherent territoriality of nationalism. Using ethnic groups as likely building blocks of nationalist mobilization for border change allows us to move beyond methodological statism. We steer clear of ahistorical modernism by using historical sources as the raw material of our data collection efforts. The temporally open-ended nature of historical change reflected in these data prevents us from conceiving the nation-state as the teleological endpoint of a history read backward, a mistake that arises when today's states are projected into the past. In addition, both our Europe-wide and global measures allow for a systematic evaluation of our theory.

Combining geographic information on state and ethnic borders enables us to operationalize Gellner's congruence principle in explicitly territorial terms.[21] Because many ethnic groups are split by state borders

[15] Mylonas and Tudor 2023.
[16] Gellner 1996. See also Section 1.3.
[17] Gellner 1996.
[18] See, e.g., Mylonas and Tudor 2023.
[19] Gellner 1996.
[20] Raun 2002.
[21] Gellner 1983.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

and most states contain more than one ethnic group, the intersection of ethnic and political geography results in ethnic segments – consistent and comparable units of analysis – nested within states. For each unit, we can identify whether the two requirements for nationalist congruence – home rule and national unity – hold. Wherever political borders cut through an ethnic settlement area, all constitutive segments of the group suffer from division. And where ethnic boundaries separate different ethnic segments within the territory of the same state, some are likely to find themselves under alien rule. As argued in Chapter 2, the former constellation can be expected to produce calls for unification, whereas the latter is likely to spur secessionist demands. Irredentist demands can arise where division meets partial alien rule. With these contours in mind, we turn to a more detailed description of our data sources and operationalization choices.

3.2 Ethnic Groups in Space and Time

Based on our understanding of ethnic groups as nations-to-be outlined in Chapter 2, we turn to their measurement. The problem to be addressed is that existing ethnic-group datasets are either nonspatial, such as the Minorities at Risk dataset,[22] or they do not offer sufficient historical depth, as is the case with *Atlas Narodov Mira*,[23] the World Language Mapping System,[24] and GeoEPR data.[25]

To fill this gap, we introduce a new dataset, Historical Ethnic Geography (HEG) that provides coverage of Europe from the nineteenth century to today.[26] Compared to existing data, the HEG data combine information from many ethnographic maps while also capturing local ethnic diversity. In addition, and importantly for our purposes, the data are time variant, based on historical information, and independent of changing state borders. Before describing this new data resource, we offer a brief discussion of how its first important ingredient – historical maps of ethnicity – originally emerged.

Historical Ethnographic Maps

In a public lecture held on January 18, 1871, the day after the coronation of German Emperor Wilhelm I, Georg Mayrzahl described the

[22] Gurr 1993.
[23] Bruk and Apenchenko 1964.
[24] WLMS 2006.
[25] Vogt et al. 2015.
[26] Europe is defined expansively to include the Caucasus, the Levant, and Northern Africa. We follow this expansive definition to include the Ottoman Empire in our sample.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

population of the newly unified Germany in numerical terms.[27] As recounted by Jason Hansen, Mayrzahl, an economist and statistician, reported the demography of the German Empire's forty million inhabitants by gender, mortality and age, and religion, while in particular drawing attention to the difference between citizens of the new state and "members of the German nation."[28] He noted that among the former were eight percent for whom German was not their mother tongue. According to his account, one-third of all fifty-three million *Volksgenossen* (German nationals) were living outside Germany's new political borders.[29]

If Mayrzahl had had access to a modern slide deck and projector, he would likely have made his point by showing an enlarged map of German speakers across Europe. He probably would have projected a map akin to the ethnographic map of Germany and its neighbors produced by geographer Heinrich Kiepert four years earlier (Figure 3.1). At the time this map was drawn, Kiepert had spent decades researching the geography of German language borders with census data and ethnographic field research in the German-French Alsace.[30]

Having made their first appearance in the middle of the nineteenth century, maps like Kiepert's aimed at documenting the geography of Europe's ethnic diversity. The proliferation of ethnic maps was driven by two major developments. The first involved technological innovations like the development of ever-more detailed census questionnaires, the appearance of precise geographical maps, and the invention of new printing methods and synthetic dyes. These made it possible to categorize populations based on language and religion and to represent their settlement areas on increasingly precise maps. Second, the rise of nationalism and the pursuit of self-determination created a demand for maps that identified and located various ethnic groups in Europe.[31] Initial efforts by German and Austrian geographers in the 1840s were quickly followed by the efforts of cartographers from Russia, the Balkans, and other parts of Europe. Together, they formed a scientific community dedicated to classifying and mapping ethnic groups.

For the most part, ethnic maps were drawn based on census data at the level of towns or districts. Other maps, in particular those from the mid nineteenth century, were based on philological research, travel reports, local ethnographic research, and older historical maps.[32] Most

[27] Mayrzahl 1871.
[28] Hansen 2015, 16–18.
[29] According to calculations from 1860.
[30] Hansen 2015, chapter 3.
[31] Kertzer and Arel 2002; Hansen 2015.
[32] Dörflinger 1999; Hansen 2015.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

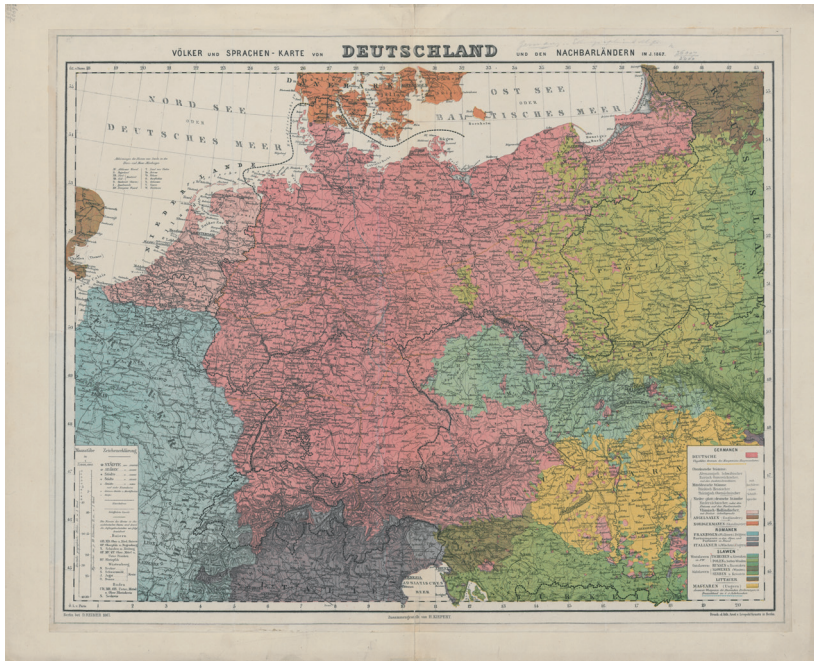


Figure 3.1 "Peoples and Languages of Germany and Neighboring States."

Source: Heinrich Kiepert 1867.

ethnic cartographers relied on native language and mother tongue as the defining attributes of ethnic groups.[33] The production of ethnic maps was generally viewed as a scientific endeavor, motivated by enlightenment-era ideals of measuring and classifying the natural world.[34] The cartographers, therefore, sought to establish common standards and provided detailed justification for their boundary drawing.[35]

But ethnic maps and census data were also used politically. States and nationalist movements disseminated ethnic maps to influence public perception of national homelands and their boundaries.[36] These efforts were most evident at the 1919 peace conference at Versailles following World War I, where all parties relied on their own maps to support their

[33] Cadiot 2005; Hansen 2015.
[34] Livingstone, Withers et al. 1999.
[35] Dörflinger 1999; Hansen 2015.
[36] Herb 2002; Anderson [1983] 1991. See Branch 2013 for parallel consequences of mapping states.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

demands.[37] At the time, this use of maps already raised concerns about political manipulation, especially when the stakes were as high as at Versailles, where the end of the war was negotiated. US geographer Isaiah Bowman noted:

> [e]ach one of the Central European nationalities had its own bag of statistical and cartographic tricks. [. . .]. It would take a huge monograph to contain an analysis of all the types of map forgeries that the war and the peace conference called forth.[38]

Yet while the means of trickery were manifold, the scope of the attempted manipulation was limited. Because the cartographers largely relied on similar data and methods, they could not arbitrarily invent ethnic boundaries without jeopardizing their reputation.[39] 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,[40] and the underlying list of ethnic groups.[41] Although ethnic categorizations may have additionally affected ethnic consciousness,[42] such ethnic malleability was restricted as well. For example, 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 manipulate significantly.

Collecting Data from Historical Ethnic Maps

Cognizant of potential caveats in using historical ethnic maps to support our data collection, we cast as wide a net as possible. We looked for ethnic maps in various online collections, as well as in major map collections like those in the US Library of Congress, the British Library, and the National Library of France. Our search yielded a (digital) mass of 350 map scans, although many failed to deliver on their promise and had to be discarded because they did not depict ethnic groups, were of poor quality, had low spatial detail, or were otherwise flawed. We also screened maps for obvious political biases but identified only very few

[37] Palsky 2002.
[38] Palsky 2002, 113.
[39] Hansen 2015. Blatant manipulation had consequences, as when geographers boycotted the journal *Petermann's Geographische Mitteilungen* due to its nationalist editor Köhler 1987, see Herb 2002.
[40] Hansen 2015.
[41] Hirsch 1997; Cadiot 2005. For example, Kertzer and Arel 2002 note that Greek, Serbian, and Bulgarian nationalists used alternative linguistic criteria to claim parts of Macedonia.
[42] Kertzer and Arel 2002; Anderson [1983] 1991.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

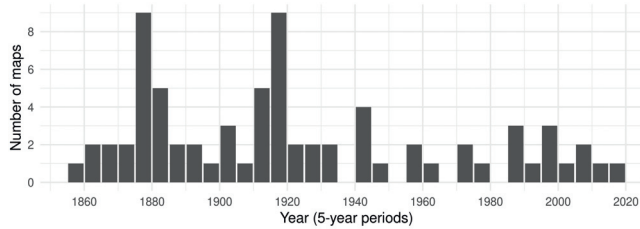


Figure 3.2 Digitized maps by year of creation.
Note: Binned into five-year periods.

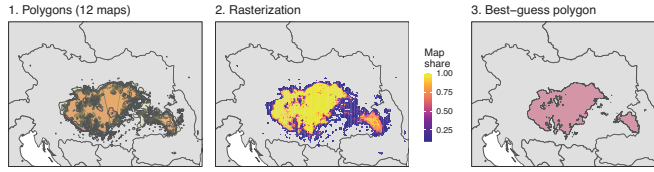


Figure 3.3 Generating polygons for the Hungarian settlement area in the HEG dataset.

cases.[43] The remaining seventy-three maps fulfilled our quality criteria. They were drawn by cartographers from across the continent, covered various parts of Europe at different times (see Figure 3.2), and used sometimes diverging categorizations of ethnic groups.

Having digitized map scans of the ethnic settlement areas allowed us to combine all usable maps into one joint, time-varying atlas of historical ethnic geography. This was done in three steps illustrated by the example of maps of ethnic Hungarians in Figure 3.3.

First, we harmonized the raw data, unifying the names of ethnic groups to decide which groups our final data would contain. This step addresses problems that arise where the levels at which groups are mapped differ between sources. For example, one map may show ethnic "Bavarians" and other German dialects, whereas different maps only show ethnic "Germans." Such problems can be solved by turning to the level at which such groups are most frequently mapped ("Germans" in this particular case). Because maps are not available for every single year and group, we had to define time periods in which any one

[43] Biased maps included a map from 1918 by the Lithuanian National Committee with a clearly oversized Lithuanian settlement area compared to twenty-four other maps of the same area, as well as maps drawn by German nationalist and national socialist Paul Langhans who edited the journal *Petermann's Geographische Mitteilungen*, which was in turn boycotted by contemporary geographers for its political biases.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

map was valid. Doing so involved systematically searching for episodes of large-scale ethnic change, most of which were due to genocide or mass displacement, as junctures that invalidated earlier maps for a given group.

Second, we spatially overlaid groups' settlement areas across all maps from a given time period by using a grid. Such rasterization produced a unified map for each group that shows the share of historical maps depicting a group as living in a particular grid cell. The resulting data from this triangulation has the advantage of combining the information across all available maps. It also enables us to depict local ethnic diversity where maps either disagree in their assessments or show overlapping ethnic settlement areas (as is often the case).

Third, we produced "best-guess settlement areas" that assume that a group lives in a location if at least half the maps suggest it is so for each group in a given time period. This move not only simplifies our spatial analysis but also reflects nationalists' concerns about national homelands – a territorial concept that is inherently binary since any one location either belongs to the homeland of a nation or not, although it can, of course, belong to more than one nation. Although national homelands are socially constructed areas that can, therefore, be contested at the margin, we use our data on ethnic settlement areas as a proxy for the concept in most of our analyses.[44] The 50 percent threshold ensures that our settlement areas include the core of national homelands claimed on the basis of ethnic settlement patterns and do not reflect any areas in which only a few members of a group happen to have been found by a single cartographer.[45]

Going beyond HEG's coverage of Europe, one could, in principle, apply the same procedures to construct a new, global dataset of historical ethnic geography. But such a data project would hardly be feasible due to the lack of high-quality ethnographic maps for many parts of the world before World War II. Instead, our global analyses are relatively limited in their aims as they primarily test the generalizability of our European findings. Given the lack of global historical ethnic data going back to the nineteenth century, our explorations of nationalist state

[44] Shelef 2016.

[45] Including such areas through a "maximalist" approach would risk overexpanding a group's settlement area to the point where even the most committed nationalists would be troubled to construct a national homeland on its basis. It would also expose our measurement to greater potential political biases from single maps, which the threshold reduces significantly.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

transformation beyond Europe are based on a back-projected version of the *Atlas Narodov Mira*[46] and the GeoEPR dataset.[47]

Solutions to the Pitfalls of Mapping Ethnic Groups

To reiterate, measuring ethnic geography across all of Europe over 140 years comes with several caveats. First, political biases pose the most important potential problem associated with the HEG data, especially where politically motivated cartographers produced maps that reflected their expansionary goals rather than the ethnolinguistic reality. As explained earlier, we address this risk at various points of the data collection – dropping maps with overt biases, averaging across maps to reduce the influence of outliers, and relying on a conservative 50 percent threshold coding of ethnic settlement areas. These measures significantly shrink the room for biases that have enough geographical weight to substantively affect our results.

A second caveat is the risk of reverse causality. The panel nature of the data addresses a simple form of this problem where state borders induce ethnic change. Yet a more complex version of this bias emerges where territorial claims lead to the shaping and reshaping of ethnic identities so that these claims then become self-fulfilling prophecies. This risk concerns some analyses more than others. Overall, we find little evidence that such dynamics bias our results, which are robust to relying on only the earliest (and thus, least likely reversely caused) maps from before 1886 or examining the effects of groups that are ethnolinguistically so distinct that it would have been nigh impossible to construct them out of thin air to substantiate territorial claims.

Third, if taken at face value, the HEG dataset tends to underestimate ethnic diversity in many contemporary Western European countries, particularly in urban settings. This underestimation is because the ethnic maps we use as source material, for the most part, depict regionally concentrated groups and not the spatially dispersed, multiethnic populations that descended from migrants who came to Europe after World War II.[48] While this constraint clearly limits the data's value in studying some forms of contemporary nationalism – particularly its anti-immigrant

[46] Bruk and Apenchenko 1964.

[47] Wucherpfennig et al. 2011. We use a geocoded version of the *Atlas* called the Geo-Referenced Ethnic Groups (GREG) dataset by Weidmann, Rød and Cederman 2010. The atlas has seen much use in the social sciences, although its data quality has drawn criticism, see, e.g., Fearon and Laitin 2003. Since the back projection comes with a significant danger of bias from reverse causality, we complement our main validation exercises with analyses based on data restricted to the post-World War II period.

[48] E.g., Dražanová 2020.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

variety – we do not expect our argument about the nationalist reshaping of states to apply to the "new" minorities in Europe's nation-states. This limited scope is simply due to their dispersed settlement patterns and the lack of a homeland within their host states that these groups could plausibly claim even if politically mobilized. Their political mobilization mostly focuses on political inclusion rather than separation.

3.3 States and Ethnic Segments within Them

States' territories and changing borders constitute the second important ingredient in our analyses. When intersected with our data on ethnic settlement areas, state borders define ethnic segments, which serve as our main unit of analysis. As clarified in Chapter 2, ethnic segments should be seen as structural proxies for the real actors since many ethnic segments remain unmobilized. The following sections introduces our data on state territories as well as the derivation of ethnic segments.

States and Their Borders

The mapping of states is as difficult as the concept of statehood is contested. We build directly on the minimalist definition of states, which was introduced in Chapter 2, as organizations "that maintain a quasi-monopoly of violence over a fixed territory."[49] This conceptualization recognizes states as organizations with a government, typically centered in a states' capital and territory delimited by borders. When focusing on Europe, we consider only state territories on the European continent and not their overseas colonies. But in our global sample, our definition of states leads us to include contiguous and overseas colonies as regular parts of their territories.[50]

With these conceptual issues in mind, we draw on and extend several existing datasets on states and their territories. To reach sufficient historical depth for our main analyses, we build on three datasets. The first, CShapes 2.0, covers all sovereign states and their colonial dependencies around the world from 1886 to 2017.[51] Furthermore, the *Centennia Historical Atlas*[52] and Scott Abramson's border data cover Europe beginning

[49] Abramson 2017.

[50] We do not follow the conventional "saltwater rule" that differentiates colonial holdings from the metropolitan core territory. This would introduce a theoretical inconsistency in our treatment of nationalists seeking independence of a national homeland that is separated from the state's capital by salt water, a river, mountain ranges, or solely an ethnic boundary.

[51] Schvitz et al. 2022.

[52] Reed 2008.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

in the Middle Ages.[53] For our main analyses of European state territories, we combine and harmonize these datasets into our own CShapes Europe dataset, which covers all state territories and their capitals from 1816 to 2017 (see Figure 3.4). Further analyses expand our focus to a global sample either by using the CShapes 2.0 dataset since 1886 or by increasing its historical reach to premodern and early modern Europe using Abramson's data from 1100 to 1790.

In using these datasets, our main and most contentious assumption is that a state's quasi-monopoly of violence was realized in a uniform manner across the entirety of the state's territories at any given point. This assumption represents, of course, a simplified picture that can diverge starkly from reality in some regions and eras. Regionally, low levels of state capacity lead us to overestimate the degree of state-nation incongruence where states are weak and unable to rule, extract, and homogenize. We return to this important nuance in a number of analyses. First, the analyses of violent conflict between states and ethnic minorities conducted in Chapters 7–9 confirm that states do not hold the quasi-monopoly of violence at all times. Instead, they are frequently challenged to the point of losing control over some of their territory altogether, which our border data register as border change. Second, zooming in on regional variation in state capacity more directly, Chapter 9 assesses the effects of railroad expansions as one of the most important dimensions of direct rule in the late nineteenth and early twentieth centuries.

We derive ethnic segments by combining ethnic settlement patterns with historical state borders. This analytical step directly follows our territorial conceptualization of state-nation congruence in that we simply overlay state borders with our ethnic map. Doing so produces ethnic segments as the spatial intersections between ethnic groups and state territories, both of which can vary over time. This implies that each ethnic segment is part of a larger ethnic group and of a state territory. Figure 3.5 illustrates the operationalization of ethnic segments by using Belgium as an example. The stylized map shows three ethnic segments – Dutch (Flemish), French (Walloon), and German – nested within the territory of the Belgian state. All three segments are simultaneously nested within broader ethnic groups whose settlement areas extend far beyond the Belgian state territory. As such, the Dutch, French, and German segments in Belgium are clear cases of ethnic division.

[53] Abramson 2017.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

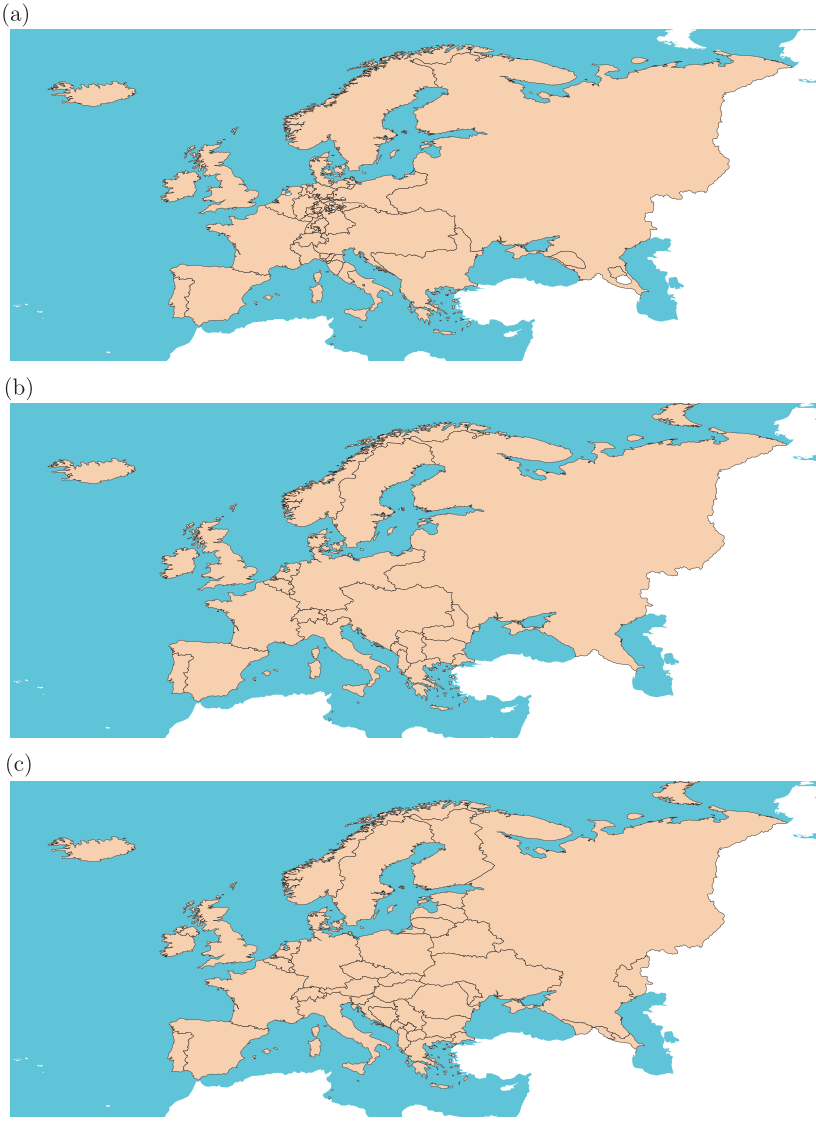


Figure 3.4 European state borders. (a) European states in 1816. (b) European states in 1914. (c) European states in 2017.

Data source: CShapes Europe.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

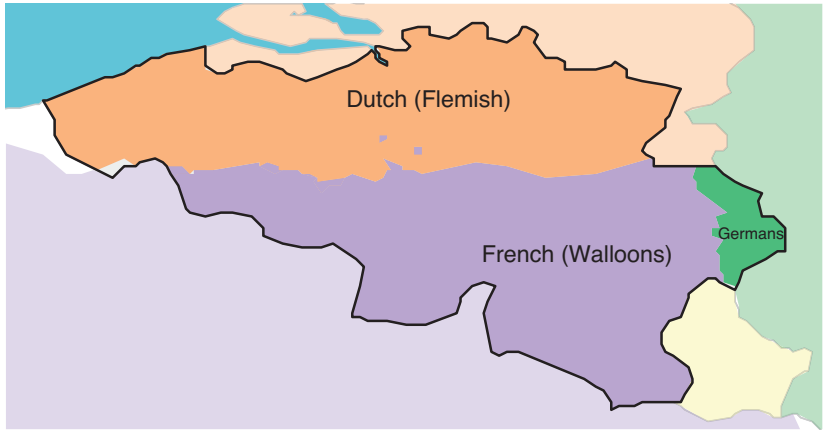


Figure 3.5 Belgian examples of ethnic segments as units of analysis.

Ethnic Segments and Dominant versus Nondominant Groups

Chapter 2 introduced a crucial distinction among ethnic segments dividing them into dominant and nondominant groups. An ethnic segment is the dominant group of a state if it controls that state's government.[54] All other segments in a state are nondominant groups under alien rule. For the most part, we encode this political dominance geographically by assigning dominant status to the largest ethnic segment that overlaps a state's capital. Figure 3.5 illustrates these points. The Belgian state is clearly divided along ethnic lines, and during historical periods of French (Walloon) dominance, Belgian Germans and Dutch (Flemish) appear as nondominant segments under alien (i.e., Walloon) rule.[55] For global analysis of the post-1945 period, the Ethnic Power Relations dataset provides more precise information by encoding ethnic groups' access to executive power in their respective countries' governments.[56]

[54] We use the notion of a dominant group if one group dominates the access to executive power in a state as a *Staatsvolk*. Cederman, Gleditsch and Buhaug 2013 use a broader notion of power access contrasting "included" to "excluded" groups or alternatively as "ethnic groups in power" and "marginalized ethnic groups," see Cederman and Girardin 2007.

[55] While usually correct, there are historical exceptions. For example, Vilnius, the coordinates of the capital of independent Lithuania in the 1990s, are "missed" by the Lithuanian settlement area by some 500 meters and instead attributed to the Belarussians. In cases like this, we rely on secondary sources to correctly identify politically dominant groups.

[56] Vogt et al. 2015.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

3.4 Border Change and Ethnonationalist Conflicts within and between States

Two main characteristics of nationalist state transformation are at the center of our theoretical and empirical attention: border change and armed conflict within and between states that arise as precursors to or as consequences of border change. We study these outcomes with newly collected data that are aligned with our ethnic segments.

Border Change

The predominant way of measuring border change in the literature employs lists that record when territory changed hands and the states involved.[57] But these lists are nonspatial, making it difficult to trace the extent to which ethnic geography shaped border change or to understand how ethnic segments have been affected by it. We, therefore, code border change spatially by relying on our geocoded state borders, which enables us to trace territorial transfers between states through space and time.

As we further elaborate in Chapter 5, we differentiate between three main types of border change. *Secession* involves a transfer of territory toward a newly created state. *Absorption* comes with a transfer of the entire territory from one state that ceases to exist to another state. *Territorial transfer* leads to a change of territory from one state to another while both states continue to exist.

Importantly, we further distinguish between ethnic and nonethnic variants of such border change by overlaying the territory that changed hands with our ethnic segments. Secession becomes *ethnic secession* if the territory is settled by a nondominant group in the old state. Absorption turns into national *unification* if both states are dominated by the same dominant groups. Territorial transfer amounts to *irredentist* border change if the transferred territory is settled by an ethnic nondominant group in the former state that dominates in the receiving state.

Conflict

In post-Napoleonic Europe, many cases of border change came about as a consequence of violent conflict between or within states. Yet there were at least as many cases of conflict and mobilization that did not result in border change even if that was the goal of one of the parties, just as some border change occurred as a result of nonviolent mobilization

[57] Tir et al. 1998.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

(most notably the breakup of the Soviet Union). To analyze how violent and nonviolent mobilization relates to state-nation incongruence, we differentiate between civil war, interstate militarized disputes, and nationalist territorial claims.[58]

Civil War. We collected data for each ethnic segment and year between 1816 and 2017 that encodes whether an ethnic civil war had started. For years before 1945, we started with rosters of civil wars based on lists provided by Kristian Gleditsch[59] and Meredith Sarkees and Frank Wayman.[60] For the post-1945 period, this data collection effort relied on existing data from Uppsala Conflict Data Program/Peace Research Institute Oslo (UCDP/PRIO)[61] linked to ethnic groups via the ACD2EPR dataset.[62]

Militarized Interstate Disputes and Wars. To analyze interstate conflict in Europe since 1816 and across the globe since 1946 for our main analyses, we relied on the dyadic Militarized Interstate Disputes (MID) dataset provided by Zeev Maoz and colleagues.[63] These data code dispute initiation at the level of directed country–dyad years.[64]

Territorial Claims between States. To measure irredentism, we use data on the onset of territorial claims between a claimant and target state with coverage from 1816 until 2001. Using data collected by Bryan Frederick, Paul Hensel, and Christopher Macaulay,[65] we identified claims that relate to territory that includes "significant portions of ethnic, religious, linguistic, or other identity groups linked to the challenger state." We expanded this prior coding to explicitly identify the specific ethnic segment in state B that is targeted by the claimant state A.[66]

[58] See Section S7.1 for further details.
[59] Gleditsch 2004.
[60] Sarkees and Wayman 2010.
[61] Gleditsch et al. 2002.
[62] Wucherpfennig et al. 2012.
[63] Maoz et al. 2019.
[64] In Chapter 4, we look at interstate wars in premodern Europe that were compiled in Peter Brecke's list of interstate wars. Brecke 1999.
[65] Frederick, Hensel and Macaulay 2017.
[66] Additional types of political violence and nonviolent mobilization are discussed in subsequent chapters. These include self-determination movements in Chapter 6 and episodes of ethnic cleansing in Chapter 10.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

3.5 Right-Sizing, Right-Peopling, and the Homogenization of States

By describing our data, we offer a synoptic account of the nationalist state transformation that has occurred in Europe since the mid nineteenth century. Conceptualizing this transformation as an increasing alignment between state borders and ethnic geographies, our account highlights border change as an important driver of this transformation. This overview adds to existing literature that has predominantly focused on states' efforts to right-people their populations through ethnic cleansing and assimilation.[67]

Prior empirical studies differ in how they capture the nationalist transformation of states and ethnic groups that results from border change or conflict. Focusing on states, most analysts measure states' ethnic homogeneity over time. They expect the formation of nation-states – in particular those driven by secessionism – to come with greater levels of ethnodemographic uniformity.[68] Other scholars focus on ethnic groups and the degree to which they are divided by state borders. Processes of nation-state formation through unification and irredentism in particular increase groups' territorial unity inside the same state.[69]

Each of these classic concepts captures one of the two core dimensions of state-to-nation alignment. Yet neither measure tells the whole story. For example, central Europe featured many ethnically homogeneous German states. Once unified, ethnic Germans were less fragmented by state borders but lived in states that were slightly more ethnically diverse than before. Conversely, the dissolution of the Soviet Union increased the ethnic homogeneity of states in Eastern Europe but led to the political fragmentation of ethnic Russians.

First, to operationalize state-to-nation congruence in a unified and one-dimensional manner, we add a third, information-theoretic measure of the mutual information the geography of states and ethnic nations provide to each other. This measure increases when ethnic settlement areas are less divided by state borders and as states become more ethnically homogeneous. Offering a single indicator for states' alignment with ethnic groups and vice versa, the measure combines the two dimensions that make European nation-states.[70]

[67] O'Leary et al. 2001.

[68] E.g., Alesina, Baqir and Easterly 1999; Wimmer 2018.

[69] E.g., Cederman, Rüegger and Schvitz 2022. Both concepts are commonly measured via Herfindahl's concentration index, which denotes the chance that two individuals drawn from the same state belong to the same ethnic group or the chance that two group members live in the same state. For further details, see Chapters 5 and S5.

[70] *Mutual information* measure from information theory that assesses the amount of information one partitioning A (ethnic settlement areas) carries about another partitioning

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

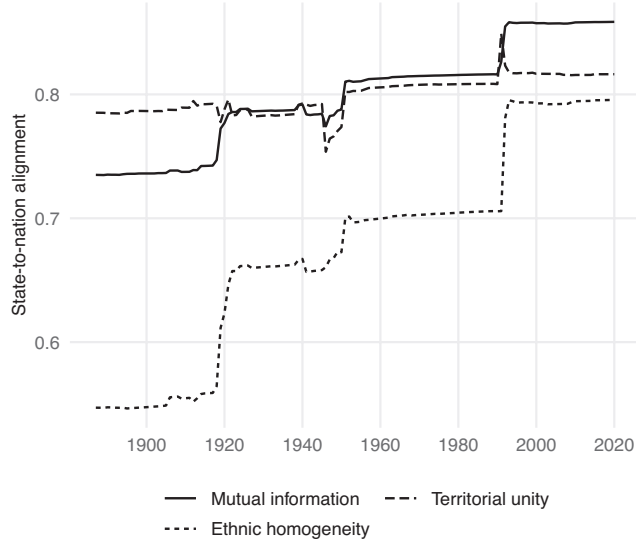


Figure 3.6 Increasing state-to-nation congruence, Europe 1886–2020.

We next measure the ethnic homogeneity of states, territorial unity of groups, and mutual information between states and ethnic nations in our data on European state territories and ethnic geography for each year between 1886 and 2020.[71] All three measures range between zero and one, with one denoting full alignment between states and nations on the respective dimension. Figure 3.6 shows the resulting trends of increasing state-to-nation alignment. The results reveal that states' ethnic homogeneity increased from 0.55 (which approximately corresponds to today's United States) to 0.80 (which is close to contemporary Sweden). Ethnic nations' high levels of territorial unity have remained comparatively constant, reflecting that there are more ethnic groups than states and that many of those groups have no state of their own. Combining both dimensions, our measure of mutual information increases from 0.74 in 1886 to 0.86 in 2019.

B (states) of the same set of points. These points are the centroids of a hexagonal grid that covers the European landmass. Our normalized mutual information (MI) metric is defined as

$$MI(A,B) = H(A) - H(A|B), \tag{3.1}$$

$$MI_{norm}(A,B) = \frac{MI(A,B)}{(E\{A\} * E\{B\})^{.5}}. \tag{3.2}$$

[71] Given the lack of time-varying global data on ethnicity, we cannot conduct this exercise for the whole globe.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

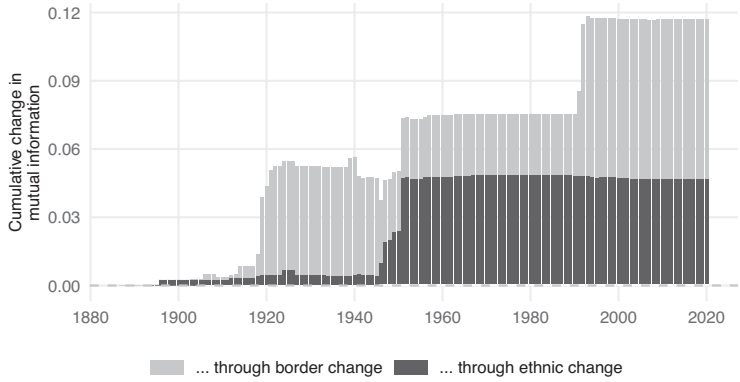


Figure 3.7 Cumulative change in state-nation congruence, Europe 1886–2020: contributions of right-sizing and right-peopling.

Data source: Based on the yearly varying HEG data.

The next step disaggregates the changes in each year into changes that resulted from border change and changes from shifts in ethnic geography, most of which were caused by violent ethnic cleansing,[72] although we reiterate here our neglect of diversifying migration flows. Figure 3.7 displays the results. As of 2020, 44 percent of the alignment between European states and ethnic groups in our data is attributed to change in ethnic geography while the remaining 56 percent is attributed to border change.[73] The figure furthermore reveals strong time dependence. Although border change contributed most to the making of nation-states after World War I and the fall of the USSR, World War II killed and displaced millions among ethnic minorities thus increasing state-nation congruence through ethnic change.

This exercise in macrolevel accounting provides the first measurement of the relative importance of border change in the transformation of states into nation-states in Europe over the past two centuries. It sets the stage for our analysis of the drivers of border change in Part II, and related armed conflicts and ethnic cleansing in Part III.

[72] See Chapter 10.
[73] See Chapter 5.

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press

https://doi.org/10.1017/9781009527767.005 Published online by Cambridge University Press