Nationalism and the Puzzle of Reversing State Size *

Lars-Erik Cederman[†]
Luc Girardin[‡]
Carl Müller-Crepon[§]

September 29, 2023

Abstract

Having increased for centuries, territorial state sizes began declining towards the end of the 19th century and have continued to do so until today. We argue that processes triggered by ethnic nationalism are the main drivers of this development. Our empirical approach relies on time-varying spatial data on state borders and ethnic geography since the 19th century. Focusing on deviations from the nation-state ideal, we postulate that state internal ethnic fragmentation leads to reductions in state sizes and that the cross-border presence of dominant ethnic groups makes state expansion more likely. Conducted at the systemic and state levels, our analysis exploits information at the interstate dyadic level to capture specific nationalist processes of border change such as ethnic secession, unification, and irredentism. We find that while nationalism exerts both integrating and disintegrating effects on states' territories, it is the latter impact that has dominated.

*Earlier versions of this article were presented at the STANCE conference at Lund University, University of California at Irvine, and at the 2020 Annual Convention of the American Political Science Association. We are grateful for constructive feedback from the audiences at these events as well as from David Siroky and the referees and editors of this journal. We also acknowledge generous financial support from the Advanced ERC Grant 787478 NASTAC, "Nationalist State Transformation and Conflict."

[†]ETH Zürich, lcederman@ethz.ch
[‡]ETH Zürich, luc.girardin@ethz.ch
[§]London School of Economics and Political Science, c.a.muller-crepon@lse.ac.uk

In a pioneering study, Lake and O'Mahony detect a major shift from integration to disintegration in the scale of governance around the world.[1] Having increased for centuries, the territorial size of states started to decline in the early 20th century, a trend that has persisted to the present day. Scholars offer several interpretations of this reversal. Enlarging the scope to include colonial holdings, Griffiths attributes a key turning point to the end of World War II, where emerging norms of self-determination favored secession as opposed to conquest.[2]

While the puzzle of reversing state size may no longer be as perplexing as it was to modernization theorists, Marxists and Realists, who each expected a steady trajectory toward larger polities,[3] there is little consensus on its resolution. In fact, despite being a fundamental transformation of the main governance units over the past two centuries, a systematically tested explanation of border change is missing in the literature on state size. Though a wealth of powerful analytical ideas exists at the macro level, scholarship has yet to articulate a more specific account of the processes that drive border change and states' size.

Confronting this challenge head-on, we argue that the puzzle's resolution requires us to consider the effect on state borders exerted by nationalism, defined as a doctrine that requires state and national borders to be congruent.[4] Integrating theoretical ideas from the nationalism literature that have so far been scattered in separate studies, we develop an account that links state-to-nation incongruence with well defined border-change processes and the effects they have on state size. *Ethnic fragmentation* of states often foments calls for secession which will shrink a state if successful. *Territorial fragmentation* of ethnic groups

[1] Lake and O'Mahony 2004.

[2] Griffiths 2016. See also Fazal and Griffiths 2014. Griffiths' analysis is inspired by Alesina and Spolaore 2003, who argue that states' size reflects a trade-off between cultural cohesion and economies of scale.

[3] Sharpe 1989.

[4] Gellner 1983.

across state borders motivates unification which increases a state's size. In a combination of both incongruences, dominant groups with minority ethnic kin abroad may realize irredentist border change, a form of territorial transfer which does not affect the average size of states. These three types of border change are consistent with the ideals pursued by nationalists in aligning state borders with ethnic geography.

Rather than measuring nationalist motivations and activities directly—which goes well beyond the scope of this study given current data limitations—we study their structural underpinnings with new geocoded data on state borders and ethnic settlements, digitized from a variety of historical maps and atlases. We rely on this information to explain the changes of state sizes in Europe since 1816, and in the world since 1886. Our spatially explicit data allow us to analyze the observable implications of ethnic nationalism at the level of the system as a whole, specific states, and down to dyadic border change processes. While previous studies have examined select processes such as secession, unification, and irredentism, to our knowledge, none has brought them together in a unified empirical framework.

Our empirical analysis indicates that after a period of state enlargement through the unification of Germany and Italy, ethnic nationalism has exerted downward pressure on state size since the end of the 19th century. At the global systemic level, we establish that the continuous shrinking of states corresponds to a decrease in their ethnic heterogeneity and ethnic groups' territorial fragmentation. At the state level, we find that ethnically heterogeneous states tend to shrink and states whose main ethnic group is split by state borders are prone to grow. We then disaggregate all observed border changes into ethnic and non-ethnic instances of secession, unification, and territorial transfer. Regression analyses show that ethnically heterogeneous states experience more ethnic secession and states with a territorially fragmented main ethnic group exhibit a higher likelihood of growth through unification and irredentism. Finally, returning to the macro level, we find that ethnic border change has driven early growth of European states and subsequent shrinkage of states both in European and beyond. Non-ethnic border changes have had a net positive effect on average state size for more than a century.

In sum, our evidence shows that nationalism's transformation of the state has produced both integration and disintegration, but that its overall effect has been dominated by the latter process. Focusing on structural drivers of the changes in state size, our analysis also demonstrates that, especially in Europe, ethnic border change has played an important role in the ethnic homogenization of states. While ethnic cleansing and assimilation have also contributed considerably to the ethnic homogeneity of states existing today, potential reverse causality induced by such ethnic change does not explain our findings.

Was there a trend reversal in state size?

We start our investigation of the reversing trend in state size with descriptive evidence. The dotted line in Figure 1 reports average state size as the arithmetic mean, replicating the findings of Lake and O'Mahony.[5] Including all sovereign states around the world while excluding their colonies, they report a doubling of average state size until the mid-19th century, followed by a steady decline throughout the 20th century.

Rather than relying on Lake and O'Mahony's time-series data we use two alternative geocoded datasets that allow for explicit spatial analysis of both political and ethnic boundaries. First, to reach sufficient historical depth, we describe the development of the European state system in terms of average state size since the early 19th century based on the CShapes-Europe dataset (see the lower, solid curve). The coding of this new dataset involved backdating the CShapes 2.0 dataset for European cases using information from the Centennia historical atlas[6] and Abramson's border data.[7] The CShapes 2.0 dataset covers

[5] Lake and O'Mahony 2004.

[6] Reed 2008.

[7] Abramson 2017.

all sovereign states and their dependencies around the world from 1886 through to 2017.[8] A second step widens the scope to the global state system, using the CShapes 2.0 dataset to capture the full territorial size of the colonial empires (see the upper dashed line). In the main analysis, we add together all territory belonging to each sovereign state including its core area and any colonial holdings. In contrast to the other two datasets, this means that the area of the United Kingdom includes that of the entire British Empire.[9]

Confirming Lake and O'Mahony's puzzle of the "incredible shrinking state,"[10] the CShapes-Europe data show how European state territories increased until the late 19th century before starting to decline in the following century. These shifts in state size largely stem from the birth and death of states (see the Appendix). Within the European state system the German and Italian unification processes brought a precipitous decline in the number of states, mirroring the increase in average state size followed by a steady increase of the number of European states and culminating in the creation of post-Communist states in the 1990s.

Does this puzzling trend hold at the global level? Given the lack of data coverage prior to the late 19th century, there is only a slight increasing trend before decline sets in around 1900 which lasts until the early 21st century. This declining trajectory reflects major geopolitical upheavals, such as the collapse of the land empires at the end of World War I, and even more dramatically, the dissolution of the colonial empires following World War II. The collapse of the USSR and the other communist states at the end of the Cold War marks a smaller but distinctive step downward in the curve. As in the European case, the trend in the number of states largely mirrors changes in state size (see the Appendix).

---

[8] Schvitz et al. 2022.

[9] The Appendix shows our findings are robust to using alternative global data that exclude states' colonial dependencies.

[10] Lake and O'Mahony 2004.

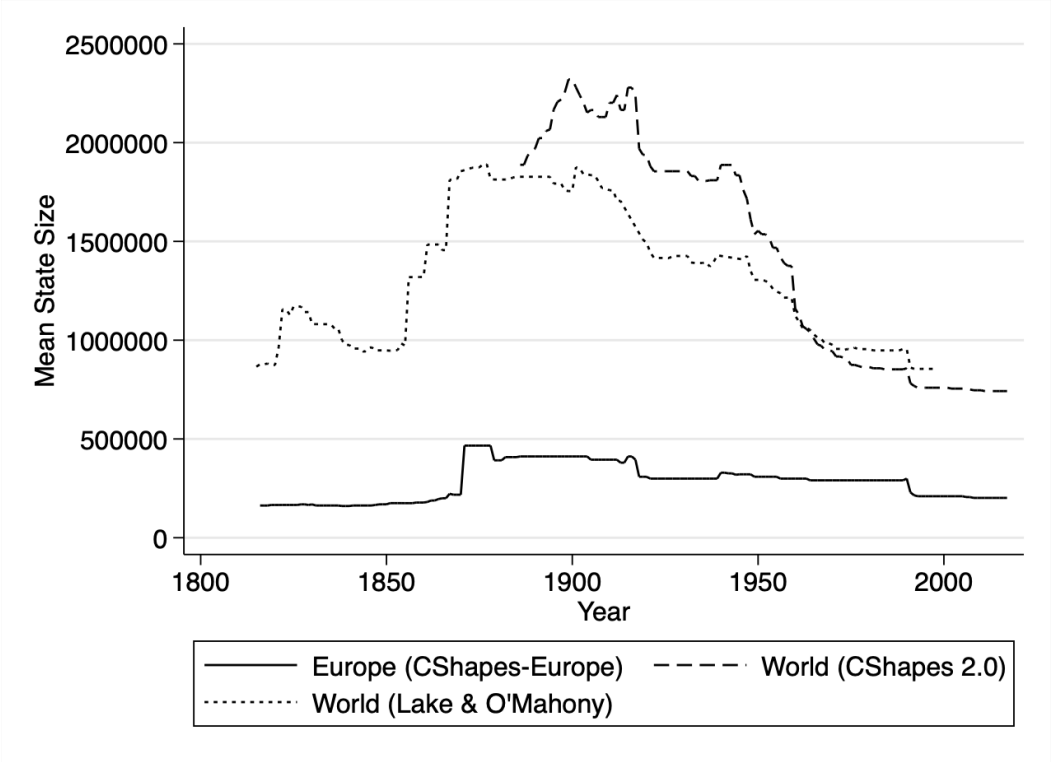

Figure 1: Trends in mean state size in Europe and beyond.

Existing explanations of state sizes and their trends

The literature on state size focuses mostly on warfare and economies of scale. We briefly review these explicit theories of state size leaving aside the extensive literature on underlying causes of border change. We will draw on them to build our argument on how nationalism affected processes of border change.

Geopolitical theories of European state formation suggest that states have grown steadily through persistent warfare since the Middle Ages.[11] As the European state system spread to the rest of the world through colonization, the effective areas controlled by centralized territorial states further increased.[12] Based on this geopolitical perspective, one would expect state size to continue to grow, or at the very least since warfare has become less frequent, stabilize. Thus, the persistent decline in state size during the 20th century challenges "bellicist"

[11] Tilly 1992; Elias 1982.

[12] Roshwald 2015.

theorizing, which expects state sizes to further increase as more powerful states continue to grow.[13] Arguably, the main reason for this anomaly relates to Tilly's and other bellicist scholars' materialist orientation that makes it difficult to appreciate the revolutionary impact of nationalism on the size and shape of states.[14]

Stressing economic production as opposed to geopolitics, economists similarly see value in large-scale governance. This field of literature typically postulates a trade-off between economies of scale and decreasing returns to scale imposed by logistical limitations and preference heterogeneity rather than geopolitical constraints.[15] In a widely cited article, Friedman suggests that territorial state sizes reflect an optimal allocation of net tax revenues based on land and labor.[16] Building on these ideas, Alesina and Spolaore view state size as resulting from a cost-benefit choice between economies of scale and efficiency-reducing preference heterogeneity resulting from cultural diversity.[17] The influence of the latter becomes more important as trade openness and democratization reduce the value of economies of scale associated with large polities. This suggests that trade-driven globalization and democratization can account for the trend toward smaller and more ethnically homogeneous states.[18]

Adopting a more empirical approach, Lake and O'Mahony's aforementioned study addresses this issue using the first systematic panel data on territorial size of sovereign states that covers the 19th and 20th centuries.[19] Based on these pioneering data, they detect a clear

[13]The same can be said about liberal and Marxist perspectives emphasizing the importance of economies of scale see, for example, Sharpe 1989.

[14]Brubaker 2010; Stuurman 1995; Cederman et al. 2023.

[15]Though see Bean 1973.

[16]Friedman 1977.

[17]Alesina and Spolaore 2003.

[18]Alesina and Reich 2015b.

[19]Lake and O'Mahony 2004.

trend toward larger states in the 19th century, followed by a declining trajectory in the 20th century. To account for this puzzling reversal, a number of possible explanations are considered including long-term changes in economies of scale, international economic openness, preference heterogeneity and political regime type. Finding little evidence for any of these accounts—partly because of a lack of time-varying ethnicity data—Lake and O'Mahony propose an account that contrasts the emergence of large federal democracies in the 19th century to that of smaller, more unified ones in the 20th century, while admitting that this conjecture amounts to little more than an "uncaused cause" (p. 700). While this explanation is in principle compatible with our nationalist account, this replaces the old puzzle with a new one: why did the size of democracies change in the first place? Furthermore, Lake and O'Mahony's study leaves room for further research that includes colonial dependencies, which are associated with the most momentous transformation in the scale of governance in modern history.

More recent data collection efforts cover the colonial dimension more fully.[20] Inspired by Alesina and Spolaore's theory, Griffiths explains the decline in state size by referring to international norms and self-determination in particular. Drawing on Fazal's insight that "state death" has become exceedingly rare in the 20th century,[21] Griffiths argues that we have now entered "the age of secession". Fazal suggests the key turning point was the major wave of secession through decolonization triggered after World War II.[22]

While these accounts offer important clues about the present trend of state size decline, they say less about the processes that drive it. Alesina and Spolaore are "not interested in 'nations' as distinct from 'nation-states'," thereby explicitly rejecting the influence of nationalism.[23] However, their theoretical focus on ethnic heterogeneity is well suited to a

[20] Griffiths 2016; Griffiths and Butcher 2013.

[21] Fazal 2007.

[22] See also Fazal and Griffiths 2014.

[23] Alesina and Spolaore 2003, 3.

macro explanation of nationalism-driven state transformation.[24] Although being more open to considering nationalism as a potential explanation for the shrinking of states, Griffiths highlights normative change after World War II including self-determination movements and decolonization [25] Importantly, this interpretation therefore overlooks how nationalism and the idea of national self-determination started to transform European state borders already in the 19th century and continued to do so worldwide more intensively following World War I.[26]

In sum, Lake and O'Mahony's puzzle of the "incredible shrinking state" remains unsolved, especially because the impact of nationalism is yet to be adequately addressed. As we have seen, much of contemporary theorizing about state size has focused on relatively abstract explanations stressing rational choices leading to optimal outcomes rather than considering macro-historical trends driven by structural, evolutionary, and partly unplanned processes, such as nationalism.

Nationalism and its impact on state size

Before explaining the reversed trend in state size, we first consider why states expanded before the onset of nationalism. As previously mentioned, the geopolitical process that consolidated the European state system eliminated scores of units, to a large extent through warfare as argued by the "bellicist" school.[27] Conquest and other types of war-related territorial acquisition increased the average territorial size of states. This process entailed border

[24]More recent studies of nation building by Alesina and colleagues focus on nationalism, although not in relation to state size. see, Alesina and Reich 2015a.

[25]Griffiths 2016. See also Fazal and Griffiths 2014.

[26]Manela 2007.

[27]Tilly 1992. While other researchers highlight alternative explanations such as economic factors, Abramson 2017, and medieval institutions, Grzymala-Busse 2019, the long-term increase in size of European states is undisputed, see Cederman et al. 2023.

adjustments that occurred largely independently of the underlying ethnic map.[28] For example, peace agreements redistributed territory for reasons relating to dynastic claims and balance-of-power considerations rather than the ethnic identity of affected populations,[29] This dynamic explains why states were growing for centuries in early modern Europe well into the 19th century. Colonialism extended this long-term consolidation of the state system beyond Europe's borders, mostly through conquest.

While the emergence of nationalism in the late 18th century did not immediately change the expansionist logic, it introduced a new principle of political legitimacy. The traditional territorial sovereignty was thus complemented by, and partly substituted for, the doctrine of popular sovereignty according to which political power derives from the people rather than from the monarch.[30] After the French Revolution, this partial transition from territorial to popular sovereignty gradually put pressure on state borders that violated the nation-state principle, which requires that state and national borders coincide.[31]

In Western Europe, state-led nationalist assimilation of minorities to a large extent eliminated ethno-nationalist incongruence, especially where such nation-building started early and benefited from an "ethnic core" around which a nation-state could be constructed.[32] However, even under favorable circumstances, such processes could take a long time, as shown by the case of France.[33] Furthermore, assimilation has often been met with less success in other parts of the world and at later stages of world history. In these cases, state borders typically adjusted to the ethnic landscape, rather than the other way around. As

[28] White 2004; Hintze 1975.

[29] Holsti 1991.

[30] Mayall 1990; Hinsley 1973; Yack 2001; Roshwald 2015.

[31] Gellner 1983.

[32] Smith 1986.

[33] Weber 1976.

borders shifted in response to geopolitical tensions given threats of, and actual resorts to, violence, this process of transformation led to both the expansion and shrinking of existing states depending on the fit between political and ethno-nationalist borders.[34]

The spread of nationalism through 19th century Europe illustrates this process.[35] While state formation in Western Europe enabled mostly successful nation building around ethnic cores, further east there was little congruence between states and nations. These disjunctions included the areas that would later become unified as Germany and Italy. Both states overcame the political fragmentation of their large ethnic communities that were previously split into tiny political units loosely organized under the heading of the Holy Roman Empire. The shock of Napoleon's armies triggered a process of nationalist mobilization, which was initially mostly cultural, but became increasingly politicized, and finally produced border change through unification.

In Eastern Europe, however, the situation was precisely the opposite as huge empires were divided into a large number of ethnic groups, some of which were in the process of emerging as ethnic nations in the 19th century. Here, the diffusion and politicization of nationalist principles took even longer time, partly depending on the relatively low level of literacy and the fact that the elites were mostly inspired by German rather than French nationalist influences. Thus, far from operating as a "light switch," nationalism had to develop in steps through cultural awakening, politicization, and mass mobilization.[36]

Based on these historical facts, we would expect nationalist politics to first produce an expansion of state size in the European middle belt of small states and principalities, before starting to produce a downward trend as the large European land empires disintegrated. Since the rest of the world was to a large extent under the domination of European colo-

[34]Miller 2007; Sharpe 1989.

[35]Schieder 1991; Roshwald 2015.

[36]Hroch 1985.

nial empires, a powerful reduction of state size could be expected beyond Europe once the thirst for self-determination had been awakened there, especially after Woodrow Wilson's promotion of this principle at the Versailles peace conference following World War I.[37]

Having outlined how nationalism spread as a historical macro process, we now shift the analytical focus down to the state level. According to Gellner, nationalism 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."[38] To separate these two situations, we label the former case *territorial fragmentation* of an ethnic group and the latter *ethnic fragmentation* of the state. Figure 2 illustrates these two deviations from the nation-state ideal. To the left, a perfect nation-state in which the state can be found, here shown as a box with solid boundary that coincides with the dashed boundary of its only ethnic group. In the middle, we depict the situation of ethnic fragmentation of a state, and to the right territorial fragmentation of the state's dominant ethnic group.[39]

How do the two types of state-nation discrepancy affect state borders? In ethnically fragmented states, territorial losses through secession should be more likely. In contrast, a tendency toward territorial expansion can be expected where the core segment's ethnic group is divided by state borders such that an important part of it extends into neighboring state territory. Generally, the more fragmented an ethnic group is, the more reasons for territorial correction there are. In both cases, ethnic nationalism builds up pressure on borders by engendering revisionist grievances. Such claims can be advanced by stateless ethnic groups, states themselves, or both in cases of irredentism. In the first case, the stateless and excluded groups are particularly likely to take secessionist action.[40]

---

[37] Manela 2007; Cederman 2013.

[38] Gellner 1983, 1.

[39] We define the dominant group as the ethnic group that has the most direct access to the state's executive power (see below for operationalizations).

[40] Germann and Sambanis 2021.

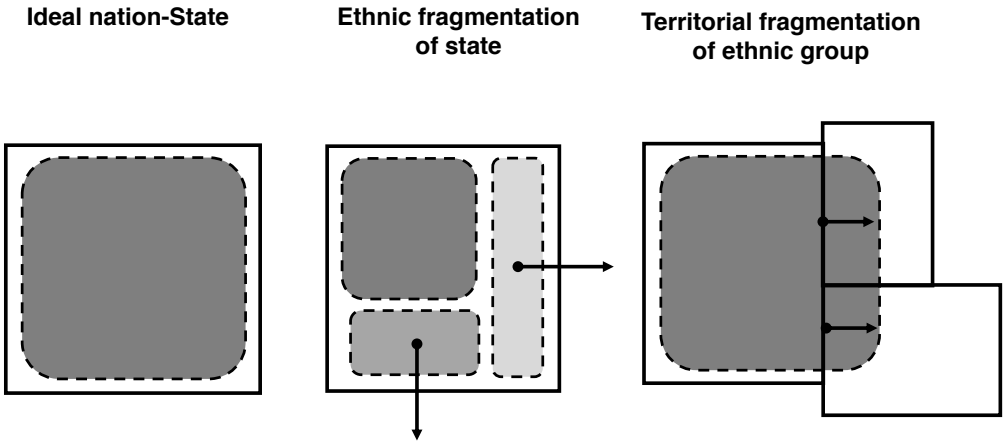

Figure 2: An ideal nation state and deviations from this ideal conceptualized as ethnic fragmentation of the state and territorial fragmentation of its dominant ethnic group.

In contrast, the second, integrating process usually features the unifying state as the main driving force, but it is also possible that elites representing external kin groups support unification.[41] Yet, without resources and organization, grievances and claim-making are unlikely to produce sufficiently powerful collective action to effect border change.[42] Nevertheless, commonly felt resentment with the status quo may boost mobilization.[43] In addition to grievances, common ethnic identities also tend to facilitate nationalist mobilization processes.[44]

---

[41] Weiner 1971.

[42] Tilly 1978.

[43] Petersen 2002.

[44] Beissinger 2002.

In sum, our reasoning uncovers a pathway from incompatibilities to border change that produces either smaller or larger states. We express these expected geopolitical effects of ethnic nationalism in the first two hypotheses:

Hypothesis 1 *Ethnically fragmented states are more likely to experience secession and thus shrink than more unified states.*

Hypothesis 2 *States whose dominant ethnic group is fragmented across state borders are more likely to attempt to incorporate their kin and thus expand than states with more unified kin.*

Both hypotheses describe straightforward changes with immediate consequences for average state size: secession leaves the rump state diminished and the new state is also smaller than the initial common state. Likewise, the incorporation of a kin-populated state clearly increases state size, since the resulting unified state is by definition larger than the absorbing and absorbed states. In irredentist configurations, which feature states trying to absorb territory inhabited by ethnic kin in neighboring states, Hypotheses 1 and 2 are directly linked to each other.[45]

Taken together, Hypotheses 1 and 2 show that ethno-nationalist sorting processes can produce both the integration or disintegration of state territory. To address the main puzzle of state size, then, we need to establish when and where either of these developments occurred and in what proportions. Most scholars of nationalism believe that, globally, there are generally too few states compared to the number of ethnic nations. This predicament of "musical chairs" leaves many ethnic nations stateless. Gellner claims that there are many more nations than states based on back-of-the-envelope calculations of states and languages.[46]

[45]Chazan 1991; Saideman and Ayres 2000; Kornprobst 2008. As we will see below, however, irredentism does not change the number of states in the system and therefore does not affect the average size of states (as long as it is measured as the arithmetic mean).

[46]Gellner 1983, 2. See also Van Evera 1994; White 2004.

Likewise, Hechter suggests that separatist and secessionist nationalism is more common than the unification variety.[47] However, these conjectures do not extend as easily to the European subsystem, which featured a very large number of political units at the end of the Napoleonic wars that were later eliminated through unification processes during the course of the 19th century. Below, we will examine the empirical net effect of nationalism in the global and European state systems.

As with state borders, ethnic boundaries are of course also subject to change through voluntary and forced assimilation, as well as ethnic cleansing and genocide.[48] Yet, once nationalism takes root, state efforts to "reprogram" citizens become increasingly difficult and have often sparked reactive nationalism instead of successful assimilation.[49] While we acknowledge the importance of "right-peopling" alongside "right-sizing" to the increasing ethnic homogeneity of states, only the latter process changes the size of states, which is our main focus.[50] In response to the potential of reverse causality, we show that our results are robust to artificially "freezing" ethnic geography as observed on early maps from the mid-19th century. Our analysis indicates that state border shifts, rather than ethnic geography, account for the increasing alignment of ethnic geography and state borders over time.

Furthermore, it is not ethnic fragmentation per se that prompts change in state size, but nationalist processes that may or may not occur. Especially after World War II, power sharing has become increasingly frequent as a way to manage ethnic diversity.[51] In such cases states respond to ethno-nationalist pressures through internal political change, rather than border change.

---

[47] Hechter 2000.

[48] McGarry and O'Leary 1993.

[49] Hroch 1985; Hechter 2000.

[50] Cf. O'Leary et al. 2001.

[51] See, e.g., McGarry and O'Leary 1993.

Identifying border change events

Instead of merely tracing ethnic and territorial fragmentation at the macro level, this section explains how state sizes vary based on explicit change processes. Therefore, we adopt an explicitly dyadic perspective. Territorial change being a zero-sum game, in any state pair, there must be a net winner (State A) and a net loser (State B). Furthermore, beyond territorial redistribution between existing states, change can be associated with state birth on the winning side, and state death on the losing side. As shown by Table 1, this logic yields four possibilities, each linked to a specific category of border change depending on the birth of State A and the death of State B (or both).

Table 1: Classifying border change events in terms of state births and deaths.

		Death of State B:	
		no	yes
Birth of State A:	no	Transfer	Absorption
	yes	Secession	Collapse/Merger

Thus we refer to *transfer* if territory is shifted from State B to State A without involving any state birth or death. Should State A lose its independence in the process, *absorption* has occurred, for instance through conquest or purchase. The opposite scenario is *secession*, which gives birth to State A.[52] Finally, in case of simultaneous state birth and death, either State B experiences *collapse* which creates a new State A (and possibly other states), or State B joins a newly formed State A (again possibly together with other states) in a process of *merger*.

The key characteristic of merger and collapse is that border change breaks the institutional continuity such that no state survives the transformation. This classification hinges critically on our ability to identify state births and deaths. This may seem like a relatively easy task, but in practice scholars differ in their definition of state continuation. For instance,

[52] In this paper, we define secession widely, whether it follows from central, peripheral or external initiative. Note that our definition includes territories leaving colonial empires. For more restrictive definitions, see Coggins 2014; Wood 1981.

we do not treat the end of the Romanov Empire and the Soviet Union as state collapses in this specific sense, but rather as a series of secessions that left the respective Russian rump states alive as geopolitical entities.

In the empirical analysis to follow, we will rely on the continuing usage of state capitals to identify predecessor and successor states and thus differentiate secession from collapse and absorption from merger.[53] Since the empirical application of this rule leads to no observed mergers or collapses, these are referred to here purely as theoretical categories that ensure conceptual completeness.

Having introduced the main types of border change, we further differentiate between ethnic and non-ethnic versions of each change. Apart from bringing together concepts that are usually discussed in isolation and less systematically, this second conceptual step allows us to identify which cases of border change follow an ethno-nationalist logic, and which do not. This is of crucial importance to the empirical analysis that follows. In particular, we rely on the intersection of the basic types of border change with ethnic settlement areas to derive the four types of ethnic border change displayed in Table 2. Each of them constitutes a subset of the corresponding border changes in Table 1.

Table 2: Classifying ethnic border change events in terms of state births and deaths.

		Death of State B:	
		no	yes
Birth of State A:	no	Irredentism	Unification
	yes	Ethnic secession	Ethnic collapse/merger

The types of ethnic border change can be defined as follows:

[53] Capital relocations within stable state borders are not coded as the birth of a new state.

- *Irredentism* is a case of territorial transfer where State A incorporates a kin subpopulation of its dominant ethnic group by extracting territory from State B. Russia's annexation of Crimea in 2014 is the most recent example of this category.[54]

- *Unification* is a case of absorption where State A and B's dominant ethnic groups are the same. Prominent historical examples are the unification of Germany and Italy in the 19th century.[55]

- *Ethnic secession* is a case of secession where State A and State B have different dominant ethnic groups. During the past two decades, this is by the far most common secessionist setting. Decolonization featured many examples of this type.[56] The decline of the large land empires, such as the Ottoman Empire, also spawned a series of ethnic secession events in the Balkans including Greece, Bulgaria, and Serbia.[57]

- *Ethnic collapse* resembles ethnic secession but proceeds without any surviving states. *Ethnic merger* is a unified state that is new rather than "inherited" from a previous state. This applies for example to Prussia or Piedmont.[58]

Correspondingly, it is possible to define non-ethnic transfer, non-ethnic absorption (e.g., conquest), non-ethnic secession and non-ethnic collapse or merger as the non-ethnic residuals of each of the four main categories of border change shown in Table 1.

---

[54]Irredentist politics is a wider category that includes claim making and support for kin; Brubaker 1996; Coggins 2014. Other definitions of irredentism focus only on the homeland state. See, for example, Siroky and Hale 2017; Kornprobst 2008.

[55]These processes also included irredentist events to the extent that some territories had to be "liberated" from neighboring states, such as France, Prussia and Denmark in the German case, and the Habsburg Empire in the Italian case. Otherwise, unification processes are usually voluntary, although its leadership can be contentious within the aggregate group Griffiths 2010.

[56]Griffiths 2016.

[57]Roshwald 2001.

[58]Again, the use of the capital-based state continuity rule discounts this category empirically. We do however acknowledge the empirical relevance of collapse and merger depending on the definition of state survival.

In the following, we focus mainly on the three types of ethnic border change featuring state continuity, namely irredentism, ethnic secession and unification, while leaving aside empirically unobserved merger and collapse. To further clarify the logic of nationalist state transformation, Figure 3 illustrates the the three main types of ethnic border change (see each column). Where the upper row depicts the configurations prior to border change and the lower row shows the outcome of the respective process. Additionally, state survival is indicated by an arrow from the upper to the lower row, state birth is shown by a star, and a cross signifies state death.

The overall effect on state size is now clearly visible: while ethnic secession reduces state size, unification generates the opposite result. In between, irredentist events merely shift the border between two states but does not lead to any change in average state size.[59]

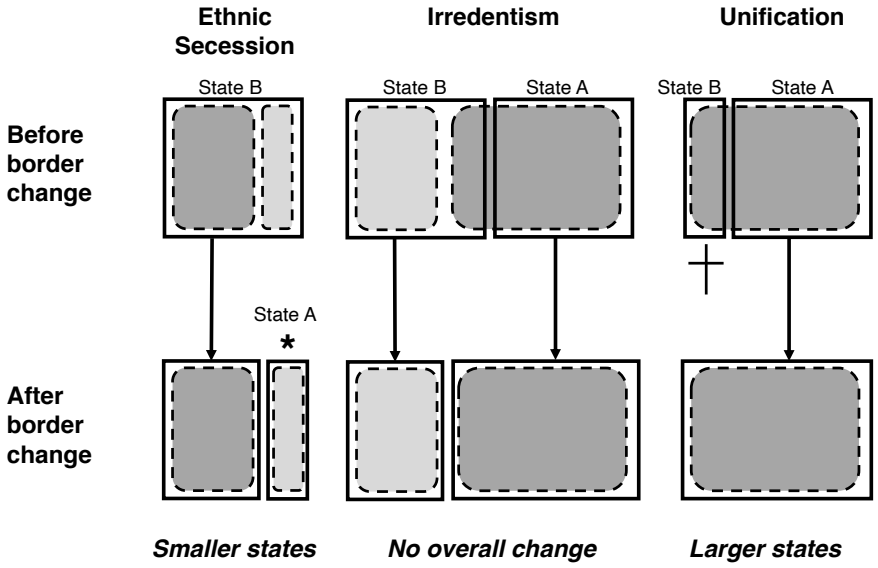

Figure 3: A schematic illustration of ethnic border change processes.

Finally, it is necessary to specify under what conditions these processes of territorial change can be expected to occur. We focus on the realized outcomes with respect to bor-

[59]To depict ethnic collapse, the vertical arrow in the ethnic secession column is replaced by a second star signifying the birth of a second state. Similarly, ethnic merger can be illustrated by replacing the vertical arrow from the unification column with a second cross showing that no state survives the transition.

ders, rather than mere claims and violent or peaceful attempts to effect border change.[60] Furthermore, we test arguments directly related to ethnic nationalism. In the case of ethnic secession, focus is placed on the extent to which minorities that are distinct from the ruling group are present.[61] Yet rather than trying to measure the political access of all group segments in our data resources, our structural analysis uses the ethnic fragmentation of the state as a proxy for ethnic power relations.

Obviously, successful secession has many other causes. Given current data limitations and the need to focus the data collection effort on ethnicity and border change, however, most of these additional causes cannot be tested in this paper and must be left for future research. Such alternative accounts include internal drivers such as the internal ethnic composition of the states in question,[62] and economic inequality[63] as well as external factors relating to great power politics and norms of sovereignty.[64] Our approach assumes that ethnic geography is causally antecedent to these processes. We summarize these arguments into a formal hypothesis that relates to the notions of ethnic and territorial fragmentation introduced above:

Hypothesis 3 *Ethnic secession (or ethnic state collapse) is more likely in more ethnically fragmented states than in more unified states.*

Along similar lines, ethno-nationalist principles yield a clear implication with respect to unification and irredentism. As argued by Horowitz, both processes are closely related, and

[60] Walter 2006; Germann and Sambanis 2021.

[61] E.g., Beissinger 2002; Coggins 2011; Hale 2000.

[62] See, Roeder 2007; Brancati 2006.

[63] See, Sambanis and Milanovic 2014.

[64] See, for example, Coggins 2014; Horowitz 1985; Griffiths 2016.

in some settings, they can be seen as substitutes for each other.[65] The more an overlapping ethnic group is fragmented into several polities, the more likely unification is to occur. In one of the few systematic studies of political unification, Griffiths shows that linguistic homogeneity is a necessary condition for such processes, as opposed to security threats.[66] Irredentism resembles unification in that border-straddling kin groups are likely to trigger expansionist state behavior. Furthermore, groups whose kin dominate neighboring states are especially likely to advance irredentist claims.[67] Abstracting away from explicit power relations, we thus expect that the territorial fragmentation of dominant aggregate groups to foster irredentist border change. At the same time, other factors are also known to spawn irredentist drives, such as economic competition and political institutions,[68] as well as the permissiveness of international norms and interests.[69]

Based on this line of reasoning, we introduce two further hypotheses:

Hypothesis 4 *Unification (or ethnic merger) is more likely in states whose dominant ethnic group is more rather than less fragmented across state borders.*

Hypothesis 5 *Irredentist border change is more likely in states whose dominant ethnic group is more rather than less fragmented across state borders.*

---

[65] Horowitz 1991. Both the German and Italian unification processes in the 19th century featured unification and irredentist border change.

[66] Griffiths 2010.

[67] Saideman and Ayres 2000.

[68] Siroky and Hale 2017.

[69] See, e.g., Horowitz 1985; Saideman and Ayres 2008.

Empirical analysis

After introducing relevant datasets, we first explore border change processes at the global systemic level before examining explicit processes at the state level. Finally, the focus we returns to the systemic level to assess whether nationalism exerts a state-shrinking effect.

Data and measures of ethnic and territorial fragmentation

Measuring deviations from the nation-state ideal requires data on both state borders and the settlement areas of ethnic groups. The latter are defined irrespective of state borders and are referred to as *aggregate groups*.[70] We utilize two main geocoded data resources on state borders, the CShapes-Europe data covering Europe from 1816, and the CShapes 2.0 Dataset for the global state system from 1886.[71]

To address the lack of available data on ethnic settlement areas, we utilize historical maps. For this study, we rely on geocoded information from the Historical Ethnic Geography (HEG) dataset, which presents historical snapshots of the settlement areas of the main ethnic groups in Europe from the 19th to early 21st century. These groups are mostly defined along linguistic, but occasionally also along religious lines. While necessarily abstracting from the presence of multiple overlapping ethnic markers and variation in individuals' attachment to them, the historical maps capture the most relevant ethnic cleavages at the time without incurring the risk of reading history backwards.

We create the HEG data by aggregating information across 73 large scale historical maps from across the continent. We draw on this information to produce "best-guess" polygons by ethnic group and time period. Figure 4 illustrates the procedure using the example of ethnic Hungarians. We first overlay all historical maps that cover a region in a given time

[70] Cederman et al. 2022.

[71] Schvitz et al. 2022.

period. In a second step, these maps are rasterized in order to yield for each raster cell the fraction of the maps containing a settled group. In a third and final step, a 50% threshold is set to produce the "best-guess" polygon.[72]

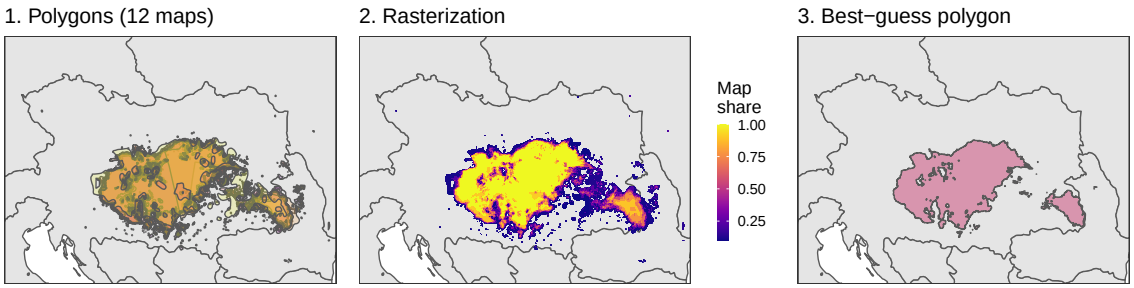

Figure 4: Generating polygons for the Hungarian settlement area in the HEG Dataset.

We guard against bias from systematic measurement error in three main ways. First, even though most map makers aspired to scientific objectivity and mapped ethnic groups using statistical information, e.g. from censuses, some maps may reflect latent nationalist claims.[73] To limits such biases to relatively small geographic areas, we scrutinize maps with significant discrepancies from other maps and drop the two maps that exhibit clear nationalist biases.[74] Second, our best-guess polygons produce an average measure of ethnic boundaries across all maps from a given period, which greatly reduces the influence of any one map. Third, even though our maps are depict the relevant ethnic groups as perceived at the time, we may map ethnic groups at an excessively aggregate level, thus missing out on meaningful sub-group differences, e.g. between dialects of the same language. Yet, such differences are strictly nested within the groups we map. As long as such potential overaggregation is not systematically biased, it will therefore if at all bias against our hypothesis

[72] A 50-50 split between ethnic groups leads to overlapping polygons.

[73] See, e.g., Hansen 2015.

[74] Our final data include maps from 64 authors from 18 nationalities. Biased maps include the case of a map from 1918 by the Lithuanian National Committee with a clearly oversized Lithuanian settlement area compared to 24 other maps of the same area, and maps drawn by German nationalist and national socialist Paul Langhans whose geographic journal was boycotted by geographers of the time for its political biases.

that ethnic fractionalization led to and was reduced by ethnic secessions an in favor of the conjecture that groups' territorial fractionalization increases state growth.

Still, systematic reverse causality and omitted variables could bias the inferences we draw below. In order to reduce this risk, wherever possible, we conduct our analysis using temporally lagged ethnic settlement data. We also show that our results for the European continent are robust to using only the time-invariant set of earliest ethnic maps produced before 1886 which are by design unaffected by border change that happened thereafter. Furthermore, as previously noted, the Appendix analyzes the extent to which ethnic settlement areas in our data change to get more aligned with state borders over time and finds relatively moderate changes. Unfortunately, however, there are currently no historical ethnic data available for the entire world going back to the 19th century. Therefore, the exploration of nationalist state transformation beyond Europe, we are forced to back-project the *Atlas Narodov Mira*.[75] Since the back-projection comes with a significant danger of bias from reverse causality, we test for the stability of the results in an analysis of the post-1946 period (see Appendix).

Based on the spatial information on the overlap between ethnic settlement regions and state territories, we can now define the two main operational measures of fragmentation used for a first test of Hypotheses 1 and 2. To operationalize state S's ethnic fragmentation, we use the standard measure given by the Herfindahl fractionalization index

$$\mathsf{ethfrac}_S = 1 - \sum_{i=1}^{n_S} (\frac{p_i}{P_S})^2 \tag{1}$$

[75] Bruk and Apenchenko 1964. We use a geocoded version of the *Atlas* called the Geo-Referenced Ethnic Groups (GREG) dataset by Weidmann et al. 2010. The atlas has seen much use in the social sciences, although its data quality has drawn criticism, see e.g. Fearon 2003. One particular problem with the ANM is the risk of using overaggregated ethnic categories, which biases, however, against our argument that ethnic fractionalization is linked to shrinking states but in favor of the conjecture that groups' territorial fractionalization increases state growth. See the Appendix for alternative findings based on GeoEPR data. See Cederman et al. 2022.

where n_S is the number of ethnic group segments i in this state and p_i the population size of each segment and P_S the total size of the state's population.[76]

We use a similar fractionalization formula to compute the extent to which the associated aggregate group AG is divided by state borders.[77] Territorial fragmentation of the associated aggregate group with a population size of P_{AG} and n_{AG} group segments, each with a population size of p_i, can be written as

$$\mathsf{terrfrac}_{AG} = 1 - \sum_{i=1}^{n_{AG}} (\frac{p_i}{P_{AG}})^2. \tag{2}$$

Exploring systemic trends in ethnic and territorial fractionalization

To illustrate the extent to which the states in the European and global state systems deviate from the nationalist ideal, we plot the spatially weighted averages of ethnic and territorial fractionalization over time.[78] Starting with the European state system, Figure 5 reveals that states' internal ethnic fragmentation has indeed declined dramatically since the beginning of the 19th century from over 0.6 down to about 0.2 at the beginning of the 21st century. This decline proceeded in major leaps, especially after World War I and the Cold War. The level of ethnic fractionalization correlates closely with average state size from the late 19th century (see the shaded area). Territorial fragmentation also declined until after the end of the Cold War when the breakup of the Soviet Union led to the territorial fragmentation of ethnic Russians.

[76]The population data were drawn from Goldewijk et al. 2011, who provide a back-projection of spatially disaggregated population data.

[77]Cederman et al. 2022 were the first to introduce this concept, but their operationalization uses territory rather than population as measure of segment size.

[78]Observations are weighted by area in order to not give too much weight to much smaller unified statelets.

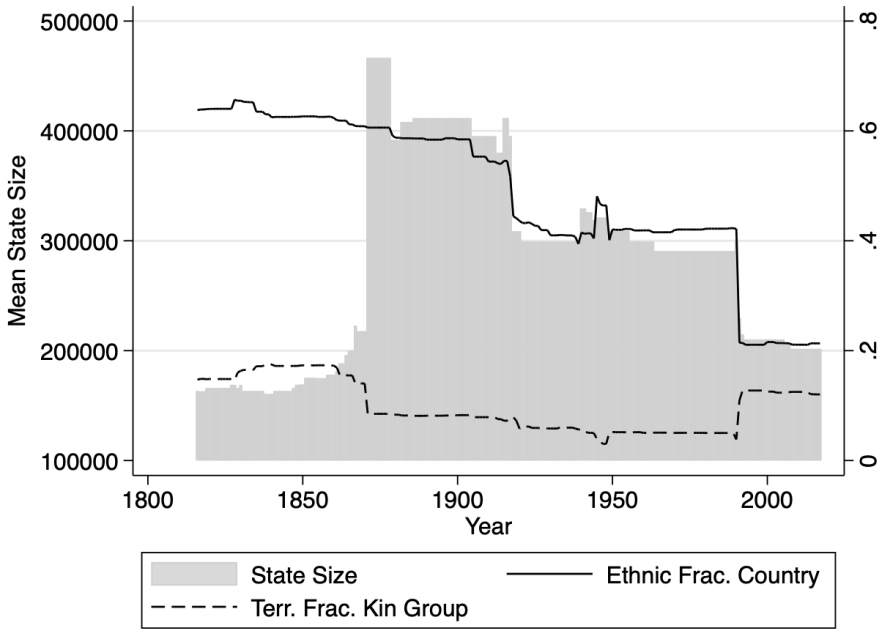

Figure 5: European trends in ethnic and territorial fractionalization.

At the global level, ethnic fractionalization in the 19th century resembles that of the European system at around 0.6 (see Figure 6). However, the global decline is less pronounced than in Europe, ending at a considerably higher average of 0.34. The most notable shift toward state-level ethnic unity appears to have occurred following World War II. Since the CShapes 2.0 dataset includes colonial holdings, the effects of decolonization are clearly visible in our results. There is also a major drop in ethnic unity following the end of the Cold War. In contrast, territorial fractionalization starts at a lower level in the global system, but climbs gradually to 0.13, a similar level to the European subsample. The upward trend throughout the 20th century indicates that although the decolonization process reduced internal ethnic heterogeneity, it produced more cases of border-transgressing ethnic settlements, an effect that is particularly pronounced in Sub-Saharan Africa.[79]

Based on the finding that state ethnic heterogeneity has decreased continuously over the past two centuries, it would seem plausible that the fragmenting effect of ethnic nationalism is responsible for the puzzling decline in state size. However, this macro correlation does not

[79] See, e.g., Michalopoulos and Papaioannou 2016.

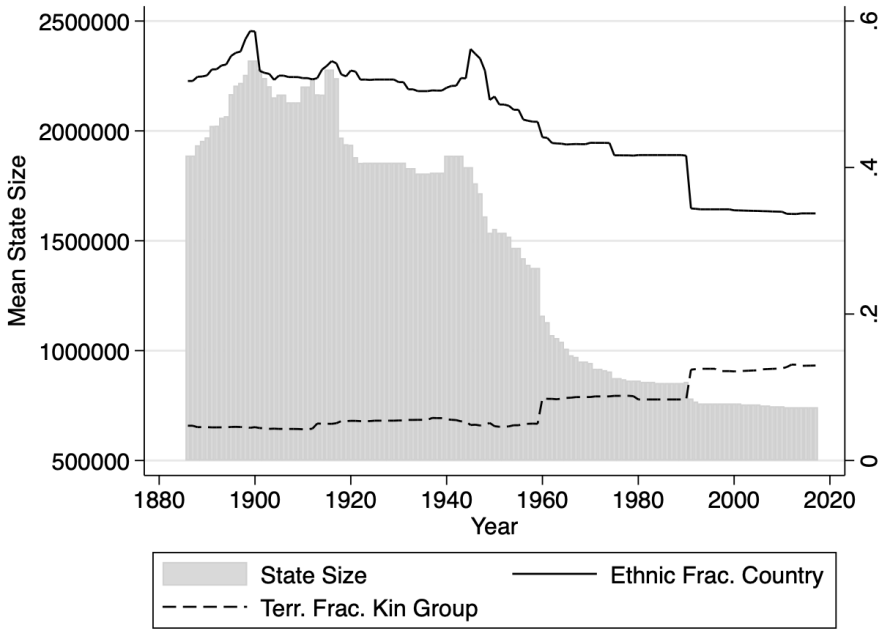

Figure 6: Global trends in ethnic and territorial fractionalization.
(Area-weighted measures based on CShapes 2.0 and back-projected ANM/GREG data.)

imply causation, and may be far removed from capturing the underlying causal mechanism of nationalism. In the following section we therefore turn to statistical state-level analyses.

Analyzing the effect of nationalist state transformation on state size

This section analyzes the direct effect of the misalignment of state and ethnic geographies on state size. In particular, we assess our theoretical arguments that ethnically heterogeneous states are prone to shrink (Hypothesis 1) and prone to grow if their dominant ethnic group is fragmented (Hypothesis 2). Because states endogenously change their ethnic makeup as they lose or gain territory, we model the process with a Cox proportional hazard model that captures the yearly probability that a state's territory either grows or shrinks as

$$h(t)_{i,t} = h_0(t)\, exp(\beta_1 \,\mathsf{EthnicFrac.}_{i,t} + \beta_2 \,\mathsf{TerrFrac.}_{i,t} + \gamma \mathbf{X}_{i,t} + \varepsilon_{i,t}), \tag{3}$$

where i are "state spells" at age t, that is historical periods during which a state's borders remain unchanged. Depending on the model, the hazard h refers to either the end of the spell by a territorial loss or by a gain. In line with our theoretical arguments, the first model tests whether ethnic fractionalization increases the risk of territorial losses (i.e., $\beta_1 > 0$) and the second assesses whether territorial fractionalization increases the odds of territorial gains ($\beta_2 > 0$). $\mathbf{X}_{i,t}$ denotes a vector of control variables that capture potential common causes of border change and the ethnic make-up of states, in particular their size in terms of territory, population and age,[80] as well as its elevation and ruggedness.[81] A second specification stratifies the model by calendar years to account for systemic shocks over time. The measurement of territorial fractionalization assumes that the dominant group is the largest ethnic segment that intersects with the capital. In the global sample, we operationalize the dominant group as the the largest demographic segment regardless of the location of the capital since the capital-based rule is unreliable in massively multi-ethnic states experiencing strong urbanization.

Table 3 presents a first set of results. The findings are compatible with our theoretical expectations. In line with Hypothesis 1, territorial losses are more likely in ethnically fragmented states. This finding holds in the baseline specification (Model 1) and when stratifying the model by year (Model 2). Interpreted as hazard ratios, the results suggest that increasing the ethnic fractionalization of a state by one standard deviation (.23) increases its risk of loosing territory by a factor of 1.44-1.73. Hypothesis 2 also receives strong support: as expected, high levels of territorial fractionalization are associated with more territorial gains (see Models 3 and 4). In terms of hazard ratios, this indicates that raising the territorial fractionalization of a state's dominant group by one standard deviation (.34) increases its risk of gaining territory by a factor of between 1.88 and 2.90.

[80] Time-variant within state spells.

[81] Terrain ruggedness is measured as the standard deviation of the elevation.

Table 3: Losses and gains, Europe 1816-2017, Cox proportional hazard models

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	2.39940***	1.60299*	0.78867+	0.09049
	(0.59090)	(0.70161)	(0.44308)	(0.46818)
Terr. frac.	1.47573***	1.41518*	1.86185***	3.13774***
	(0.38627)	(0.59367)	(0.45900)	(0.67929)
Observations	7610	7610	7610	7610
Pseudo R^2	0.044	0.071	0.076	0.129
Strata	No	Yes	No	Yes

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also find consistent support for both hypotheses in extending the analysis to the global dataset based on CShapes and backdated ANM/GREG groups. To test Alesina and Spolaore's idea that trade and democracy should reduce the size of states, Table 3 also includes two models that feature a dummy variable for democracy and a log-transformed measure of trade openness.[82] As in the European subsystem, states are more likely to shrink as ethnic fragmentation increases (see Models 1-3). Furthermore, territorial fractionalization is also associated with territorial expansion (see especially Models 5 and 6). Unsurprisingly given the origins of nationalism in Europe and its incomplete spread across the globe, the associated effect sizes are somewhat smaller than in the European sample discussed above. Changing ethnic (territorial) fractionalization by the same amounts as in the previous paragraph is associated with risks of losses (gains) that increase by a factor of 1.2-1.3 (1.3-2.3). While missing data introduce considerable uncertainty, democracy is associated with an increase, rather than a reduction, in state size (see Models 3 and 6). There is some evidence that trade shrinks states, but this evidence is rather weak (see Model 3). Yet, it should be reiterated that all of the global analysis relies on backward projected ethnicity data, which are likely to under-count smaller ethnic units that were eliminated through forced or volun-

[82] Alesina and Reich 2015b.

tary assimilation. This means that the disintegration results are likely to biased downward, whereas the effects of gains could be overstated.[83]

Table 4: Losses and gains, World 1886-2017, Cox proportional hazard models

	(1) Loss	(2) Loss	(3) Loss	(4) Gain	(5) Gain	(6) Gain
Ethnic frac.	0.82385**	1.05917**	0.81302	0.12653	-0.15363	-0.93677
	(0.29587)	(0.36244)	(0.53275)	(0.47839)	(0.47065)	(0.63126)
Terr. frac.	-0.54804	0.50367	0.38994	0.67408	2.38321***	1.69148+
	(0.49031)	(0.55961)	(0.78444)	(0.65656)	(0.66361)	(1.01224)
Democracy			-0.16231			0.53496+
			(0.26946)			(0.31778)
Trade openness, log			0.31797			0.09351
			(0.21478)			(0.18822)
Observations	13849	13849	7421	13849	13849	7421
Pseudo R^2	0.094	0.249	0.322	0.070	0.139	0.199
Strata	No	Yes	Yes	No	Yes	Yes

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These two sets of results from the European and global samples give weight to our two main hypotheses and remain robust to a series of robustness checks (see the Appendix for all results). First, we estimate conventional linear regression models that allow us to take the size of territorial gains and losses into account. Second, we further investigate the timing of our main effects in the European and global samples. We find that ethnic fractionalization and territorial fragmentation did not affect territorial gains and losses in European states between 1490 and 1790. This strongly suggests that our main results are produced by post-French Revolution nationalism rather than ahistorical attributes of ethnicity. Furthermore, our main results hold before and after the Second World War, even though the effect of ethnic fragmentation on territorial losses are notably stronger in the latter case. While the "age of secession" played an important role in the decline of state sizes after 1945,[84] ethnic

[83]Most of the results are robust even without backward projected ethnicity data, see the Appendix for analysis starting in 1946.

[84]Griffiths 2016.

and territorial fragmentation contributed to state territorial losses and gains well before this period.

0.1 Analyzing border-change processes

To test Hypotheses 3, 4 and 5, this section identifies the specific processes of border change and assesses their origins in discrepancies between state borders and ethnic geography. Having theoretically defined the conditions for each type of border change, we now need to operationalize the continuation rule that determines cases of collapse and merger. Since both the CShapes-Europe and CShapes 2.0 datasets provide geocoded and time-varying data on capitals, we use this information as a proxy for state leadership. Our operational rule stipulates that states "survive" a border change event if the capital remains unchanged.

What does the simple state-continuation rule entail? Tables 5 and 6 offer an overview of the event counts for our two main samples. In Europe, ethnic border-change events dominate the categories of secession and absorption, whereas transfers are evenly split between ethnic and non-ethnic events. The global picture since 1886 is similar, with ethnic and non-ethnic transfers relatively evenly split and secession being dominated by ethnic events. The main difference pertains to unification, of which there are only three cases: Vietnam in 1975 and Germany and Yemen in 1990. Based on the capital continuity rule, there are no cases of collapse or merger in either the European or global sample.

Table 5: Number of ethnic border change events as share of all events in Europe, 1816-2017.

		Death of State B:	
		no	yes
Birth of State A:	no	Irredentism/Transfer 51/105	Unification/Absorption 50/61
	yes	Ethnic Secession/Secession 34/40	Ethnic Collapse & Merger/All cases 0/0

Based on five-year periods, Figure 7 combines CShapes-Europe state borders with HEG ethnic boundaries. The light blue bars correspond to non-ethnic border changes. This means

Table 6: Number of ethnic border change events as share of all events worldwide, 1886-2017.

		Death of State B:	
		no	yes
Birth of State A:	no	Irredentism/Transfer 42/102	Unification/Absorption 3/21
	yes	Ethnic Secession/Secession 130/136	Ethnic Collapse & Merger/All cases 0/0

that the total height of the bars reflects the total number of border change events per period. The three other colors mark cases of ethnic unification (green), irredentism (yellow), and, combined, ethnic secession (dark red).

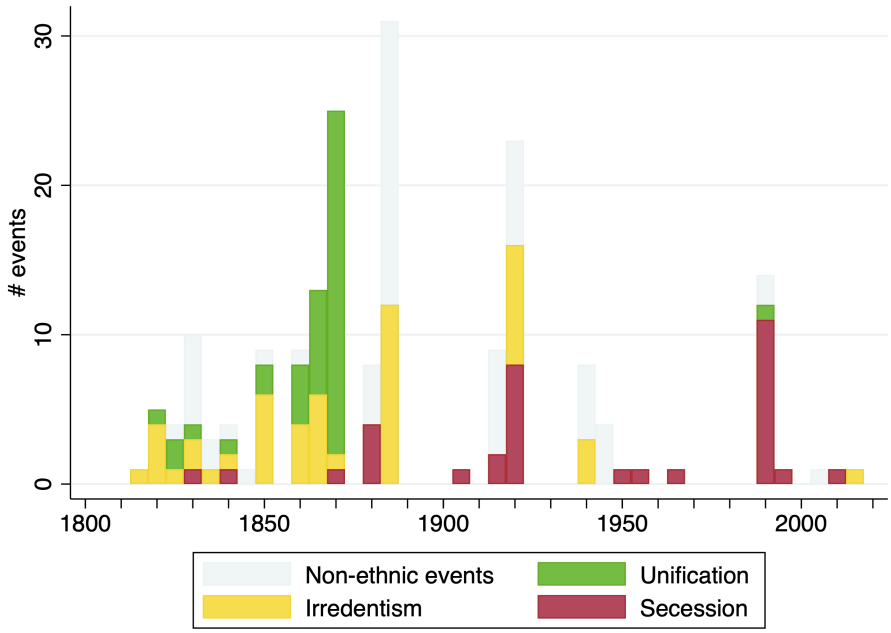

Figure 7: Ethnic border changes in Europe based on CShapes-Europe state borders and HEG ethnic map.

After the unification of Germany and Italy during the second half of the 19th century, unification events became a rarity in European history (see, for example, German reunification in 1989). Instead, ethnic secession increased in importance especially after World War I. Furthermore, as a proportion of all border change events, ethnic border changes have become increasingly significant toward the end of the 20th century.

This picture becomes even starker in Figure 8, which presents ethnic border changes in the global state system from 1886. As in the European setting, ethnic border changes have eclipsed their non-ethnic counterparts since World War II. Before this point, conquest often triggered non-ethnic events, such as unification and irredentism (cf. Figures A8 and A9). Following the end of the World Wars and the Cold War, ethnic secession, which makes up almost all secession cases, dominates the historical trajectory with particularly powerful secessionist waves. Unification, by contrast, is a relatively rare event in the global system. In recent years, cases of irredentism have also been rare, making the Russian annexation of Crimea a prominent exception that is also the most recent border change recorded in our datasets.

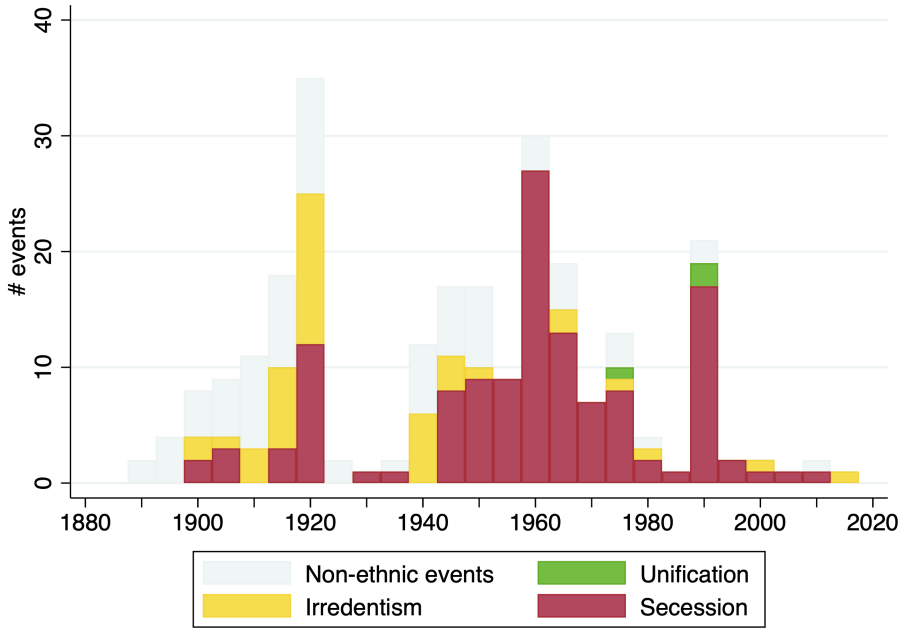

Figure 8: Ethnic border changes around the world based on CShapes state borders and backdated ANM/GREG data.

Nationalist state transformation and border change events

The next step is to use the data on ethnic border change events to test whether discrepancies between ethnic and state geographies affect unification and secession according to our theoretical expectations. Following on from the state-level regression analysis of overall gains

and losses in Tables 3 and 4, we estimate the effect of ethnic and territorial fractionalization on the three main types of border change.

In order to account for these dynamics, we again estimate Cox proportional hazard models where each border change affecting a state's border marks the end of a state period. Failures occur when the state experiences the outcome in question. If a state-period ends with a different border change, it is treated as being censored. We include the same state-level controls as defined above.[85]

The results from the European subsample with CShapes-Europe and HEG data in Table 7 support these conjectures. Once again we report both models with and without stratification by calendar year. Models 1 and 2 show that states' risk of loosing territory through secession increases with higher levels of ethnic fractionalization. As anticipated by Hypothesis 4, states with more territorially fragmented dominant ethnic groups are much more likely to expand. However, stratification by year provides a less precise estimate since many of the unification events in the 19th century happened within the same year.[86] Additionally, the fact that ethnic fractionalization is negatively related to unification is unsurprising given the unifying state must be relatively homogeneous to act as an ethnic magnet for its kin. Finally, in agreement with Hypothesis 5, we find that irredentist events are triggered by states with a leading ethnic group that is highly fragmented (see Models 5 and 6). Again, ethnically fragmented states are much less likely to embark on redeeming their kin from neighboring countries.

Based on global data since 1886, Table 8 presents similar, yet somewhat weaker results. Models 1 and 2 indicate that fragmented states are consistently more likely to experience secession (thus disregarding the few cases of non-ethnic secession, such as Taiwan's divorce

[85]The controls are the state's territorial size and population, its age, as well as its elevation and ruggedness. Controls are all logarithmically transformed and coefficients are not shown.

[86]If unification is based on the winning state (State A), this result looks much stronger (see the Appendix).

Table 7: Border-change events, Europe 1816-2017, Cox proportional hazard models

	(1) Eth. Secession	(2) Eth. Secession	(3) Unification	(4) Unification	(5) Irredentism	(6) Irredentism
Ethnic frac.	5.97399***	7.83802***	-1.42600	-4.58420*	-0.93573	-2.46694***
	(1.45256)	(1.89341)	(0.94696)	(2.03761)	(0.60042)	(0.57699)
Terr. frac.	-2.31241	-3.09197	5.84872***	8.65288+	2.42076***	3.59641***
	(1.76799)	(2.30287)	(1.23267)	(5.14039)	(0.53849)	(0.97605)
Observations	7610	7610	7610	7610	7610	7610
Pseudo R^2	0.335	0.562	0.186	0.266	0.082	0.206
Strata	No	Yes	No	Yes	No	Yes

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

from China).[87] Since there are only three cases of ethnic unification in the post-1886 global state system, we cannot test the respective hypothesis robustly using this sample. Lastly, we find evidence in line with Hypothesis 5 in regards to irredentism, which is positively linked to territorial fractionalization (see Models 3 and 4). With much more ethnic diversity being present in the global sample, especially in decolonized states, country internal cohesion plays a less prominent role in irredentist processes.

Table 8: Border-change events, World 1886-2017, Cox proportional hazard models

	(1) Eth. Secession	(2) Eth. Secession	(3) Irredentism	(4) Irredentism
Ethnic frac.	1.05988+	2.90984**	-1.56927*	-1.97628*
	(0.61079)	(1.07136)	(0.71824)	(0.95093)
Terr. frac.	-8.12664**	-3.58896	2.55279**	3.90667***
	(2.89292)	(2.61256)	(0.81136)	(0.88094)
Observations	13849	13849	13849	13849
Pseudo R^2	0.268	0.682	0.062	0.165
Strata	No	Yes	No	Yes

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In sum, we conclude that ethnic secession, unification and irredentist events appear to fit well into the macro process of nationalist state transformation that we have outlined, thus adding detailed evidence in favor of Hypotheses 3, 4 and 5. Moreover, there is am-

[87] This assumes that both the People's Republic of China and Taiwan are dominated by the same aggregate group. This is a controversial point but conforms with the coding of the *Atlas Narodov Mira*.

ple additional qualitative evidence that confirms these findings. While the collapse of land and colonial empires in many cases were precipitated by warfare and economic factors,[88] nationalism played a decisive role in the disintegration processes. For example, Hungarian and Turkish nationalism undermined the Habsburg and Ottoman empires respectively well before World War I by fueling a combination of secessionism and irredentism.[89] Nationalism also contributed to the collapse of the colonial empires following World war II, bringing about a flood of costly protests and rebellions, nationalism also contributed to the collapse of the colonial empires following World War II.[90] Finally, Beissinger convincingly shows that the collapse of the Soviet Union was triggered by ethno-nationalist mobilization that produced a quickly cascading series of events that unfolded against the backdrop of an overextended and inefficient system.[91]

Before evaluating whether ethnic nationalism produces primarily integration or disintegration and the growth or decline of states, it is necessary to account for how the dyadic change processes add up to changes in state size in the entire system. This is the task of the next section.

Deriving average state size as a function of border changes

Our definitions of border change processes allow us to trace the contribution of each type of border change to the average territorial size of states. To do this, we exploit the fact that the arithmetic mean of state size is dependent only on the number of states in the system if one treats the total area of the system as a constant (except gains from previous territories outside of state control). Thus, only secession and absorption events affect average state

[88] See, e.g., Roshwald 2001.

[89] See, e.g., Weiner 1971; Kann 1974; Roshwald 2001.

[90] For an historical overview, see Hiers and Wimmer 2013.

[91] Beissinger 2002. See also Hale 2000.

size. In contrast, transfers and the subcategory irredentism, do not influence average state size because these border changes only affect the size of specific states rather than the overall surface area of the system (see the Appendix for details).

Given this understanding, we are now in a position to trace the development of mean state size as a function of the specific processes of border change. Having defined ethnic secession and unification as ethnic subcategories of secession and absorption, we are also able to construct trajectories for these cases and for their non-ethnic counterparts, that is, non-ethnic secession and non-ethnic absorption.

This procedure yields three alternative histories: the actual trend in average state size, the hypothetical trajectory based solely on ethnic border changes, and finally, the curve that corresponds to strictly non-ethnic border changes.[92] Whereas non-ethnic change is associated with the developmental theories that we referred to at the beginning of this study, the ethnic trajectory captures the logic of ethnic nationalism.

What do these alternative histories look like? Figure 9 traces the three trajectories in the European state system. The actual trend is identical to the solid curve in Figure 1. The non-ethnic trajectory remains mostly flat, as would be expected given the calming influence of great power cooperation within the Concert of Europe. In contrast, the ethnic trajectory stays much closer to the actual curve. In Europe then, nationalism first had an expanding and then contracting effect on state size, reflecting the fact that the process first affected the city-state belt before destabilizing the land empires in Eastern Europe.

If we turn to the global comparison as shown in Figure 10, the divergence between the two hypothetical scenarios becomes much more pronounced.[93] As a reflection of imperial conquest disregarding ethnic borders, the non-ethnic curve continues to increase until World

[92]These are not fully-fledged counterfactual histories that "rerun history" with and without ethnic border change, and may thus violate the assumption of "cotenability," see Cederman 1996.

[93]See the Appendix for an alternative chart based on geocoded ethnicity data from the GeoEPR dataset, see Wucherpfennig et al. 2011.

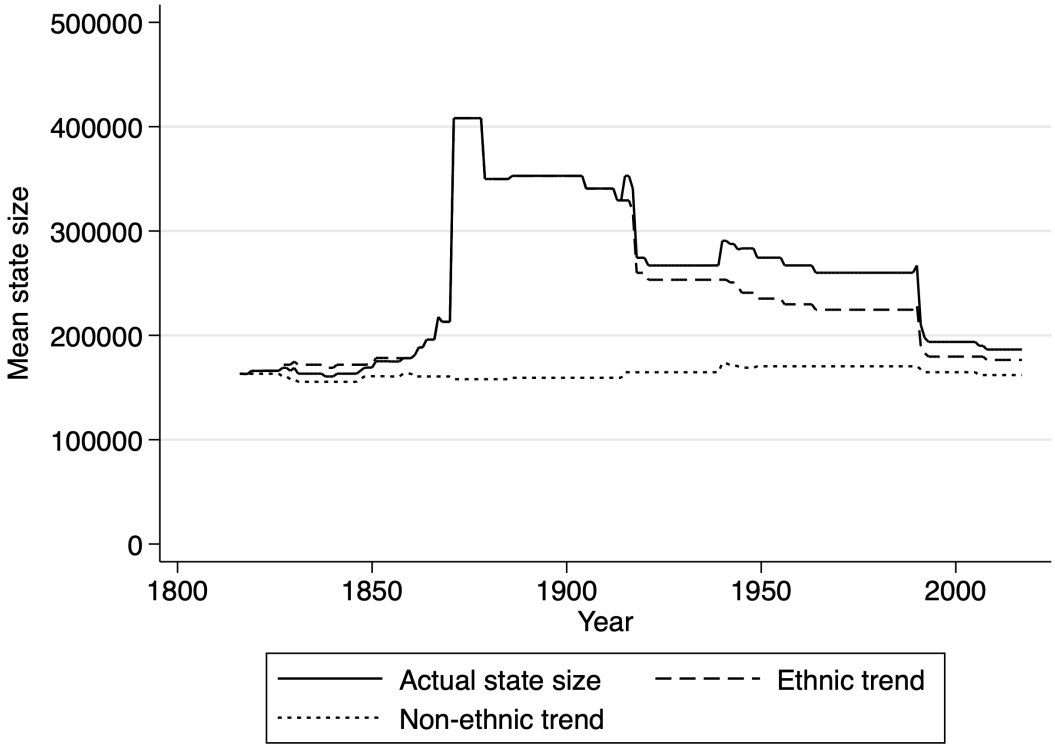

Figure 9: Comparing ethnic and non-ethnic trends in mean state size with actual trend in Europe. (Based on CShapes-Europe state borders and HEG ethnic map.)

War II, after which it starts declining gently while remaining at a high level, much as would be expected from "colorblind" developmental theories.[94] In contrast, the ethnic scenario closely follows the actual trend, which again suggests that ethnic nationalism accounts to a large extent for changes in the scale of governance. These trends offer an intuitive solution to the puzzle that we introduced at the very beginning of this study. As expected, the initial upward trend is followed by a long-term downward trend that persisted throughout the 20th century.

While these findings offer solid support for our theoretical framework, it should be noted that we have made a number of important assumptions along the way. Some of these are further explored in the Appendix, which offers a series of sensitivity analyses with alternative model and data specifications. For example, we show that there is little support for

[94] As argued by Fazal and Griffiths 2014, the declining trend reflects changes in the normative environment of the state system that made conquest a rare event after 1945.

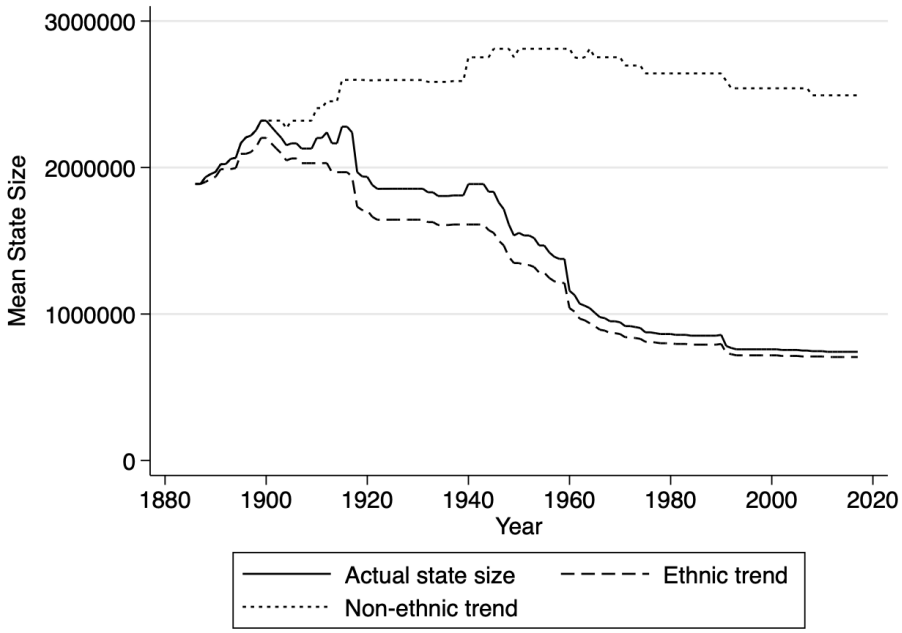

Figure 10: Comparing ethnic and non-ethnic trends in mean state size with actual global trend. (Based on CShapes state borders and backdated ANM/GREG data.)

alternative explanations referring to democracy and trade, and that the main results do not hinge exclusively on the decolonization process in the global sample.

Throughout the study we have treated ethnic boundaries as if they were exogenous. Of course, extensive constructivist scholarship reminds us that this can be a questionable assumption. In the Appendix, we mitigate these concerns to some extent by holding ethnicity constant from 1886 onward instead of using time-varying data. Finally, the Appendix also features an mutual information analysis, showing that over the past 150 years European state borders have adjusted more to ethnic boundaries from 1886 than ethnic boundaries have to state borders in 1886. While this leaves open the issue of reverse causation beyond Europe, it does suggest that that our basic assumption is a reasonable starting point for the continent of Europe.

Finally, it should be stressed that our results do not suggest that ethnic nationalism is the only process affecting state borders. Decolonization would have produced a much better ethnic fit in the absence of the *uti possedetis* norm, which prescribes that post-imperial bor-

ders should follow colonial administrative borders rather than ethnic settlement patterns.[95] When the colonial empires disintegrated, ethnic cohesion increased compared to the colonial period, but most post-colonial states remained highly fragmented, and ethnic groups territorially divided.[96] More generally, historical border precedents continue to influence border demarcations and thus state sizes.[97] Given the inertia of the "territorial integrity norm,"[98], most nationalists have thus been forced to accept compromises involving autonomy and power sharing rather than border adjustments.[99]

Conclusion

This study began by introducing the puzzle of reversing state size. We have argued that nationalism is key to understanding this transformation. Our empirical analysis focuses on the implications of ethnic nationalism, connecting underlying ethnic conditions with processes of border change. As anticipated, both internal and external deviations from the nation-state ideal strongly influence state borders. In particular, ethnically fractionalized states run a much higher risk of losing territory or even collapse than more homogeneous states. Conversely, those states that are led by ethnic groups with significant fragmented kin in neighboring states tend to expand into those areas. However, from the late 19th century, the former state-shrinking effect has dominated, which explains why the 20th century saw a massive reduction in state size. With our newly assembled data on the European and global state systems, we have been able to trace the beginning of this downward trend to the turn of the 20th century.

---

[95] Griffiths 2015.

[96] Michalopoulos and Papaioannou 2016.

[97] Abramson and Carter 2016.

[98] Zacher 2001.

[99] McGarry and O'Leary 1993.

By offering a systematic and precise spatio-temporal perspective, this paper goes beyond existing macro analyses of nationalism that are either entirely qualitative or based on conventional country-level panel data. Explicitly geocoded data enables the analysis of these processes with greater precision than has so far been possible with non-spatial country-level indicators. Rather than treating the state as a black box, the spatial approach allows us to combine different levels of analysis. We have presented a series of analyses covering the system level via the state level down to the level of dyadic processes, which are then related back to the system level.

The dyadic perspective also contributes to conceptual development, especially with regards to processes of border change. Linking territorial gains and losses to state birth and death yields systematic typologies of border change processes and corresponding ethnically-related subcategories, such as ethnic secession, unification and irredentism. Furthermore, this study breaks new ground in terms of data use, both on political and ethnic borders, enabling us to consider the influence of decolonization, albeit the precision of ethnicity data remains more developed for Europe than globally. For this continent, we introduce new spatial data that aggregate a large number of historical ethnicity maps dating back to the 19th century.

While we have made inroads into several analytical dimensions, this study calls for future research, especially as regards the deeper causes of state formation and ethnogenesis. We have treated as exogenous the fact that pre-nationalist units in the city-state belt were smaller than the emerging German and Italian ethnic nations, while the opposite applied to the empires in Eastern Europe and beyond. Future analysis should also attempt to endogenize the demographic processes that gave rise to the ethnic units that operated before and during the process analyzed here.

Furthermore, given the limitations of current available data, we must leave the systematic testing of several alternative explanations of border change processes to future research,

including the influence of administrative units on secessionist behavior,[100] as well as other institutional and economic variables related to irredentism.[101] Furthermore, our approach is based on correlations at the systemic and state levels rather than offering ironclad strategies of causal identification, which are better implemented in more selective settings. At any rate, we have gone beyond simple correlational macro analysis by relying on control variables linked to alternative explanations, fixed-effects estimation, mutual-information analysis, and several alternative datasets (see the Appendix).

More can certainly be done to study the direct influence of ethnic nationalism, beyond the structural link between ethnic configurations and state size. Most importantly, data on nationalist mobilization are needed to pin down the timing of nationalism more exactly beyond our general argument about nationalism's diffusion from west to east in 19th century Europe. Time-sensitive information of this kind would help us to evaluate our explanation compared to alternative ones more definitively. In particular, an explicit account of the timing of ethnic nationalism calls for analysis of specific claims and grievances as well as entire mobilization processes that indicate how nationalism has diffused in different parts of the world.[102] Naturally, the reconstructed trends shown in Figures 9 and 10 should not be confused with explicitly counterfactual analysis that requires more sophisticated, empirically informed simulation modeling.[103]

Another area that urgently calls for future research is the mapping of historical ethnicity worldwide. Instead of relying on imperfect back-projected data, which causes bias as we have already noted, data collection efforts might draw on existing maps, such as Mur-

[100] For example Roeder 2007.

[101] E.g. Siroky and Hale 2017.

[102] Wimmer and Feinstein 2010 construct a large historical dataset on nationalist mobilization and structure data in terms of contemporary state units rather than changing polities. Such an approach arguably risks introducing hindsight bias.

[103] See, e.g., Bhavnani et al. 2014.

dock's African ethnicity data, among others.[104] Likewise, time-varying data on administrative units, which are currently unavailable for the entire samples used in this study, would allow us to test the influence of historical political borders.[105]

Moreover, to reduce complexity, our analysis does not consider political violence, although it is well known that conflict patterns interact with both increases and decreases in the scale of governance. For instance, Abramson and Carter show that territorial claims are made frequently during times of war and crisis.[106] In the era of nationalism, the bellicist expectation that interstate warfare promotes state formation has to be qualified, especially since it puts pressure on multi-ethnic states.[107] Because most governments fear territorial disintegration, successful secession often occurs through violence see e.g. Walter 2006. Future research will have to study whether there is self-reinforcing dynamic between nationalism and warfare. Finally, to capture changes over time, future research needs to uncover how the evolution of international norms and great power interests interacts with the ethnonationalist processes analyzed in this study.[108]

Given the need for more research, this study does not ultimately resolve the puzzle of reversing state size. It does, however, outline an explanation that arguably brings us closer to its resolution. During the past two centuries, ethnic nationalism has transformed the principle of governance in the state system, first by prompting further integration in Europe in the mid-19th century, and then by starting to fragment existing units. Nationalism triggered several waves of imperial disintegration, starting with the collapse of the European land empires after World War I, and followed by the dismantling of the European colonial

[104] Murdock 1981.

[105] See, e.g., Roeder 2007; Griffiths 2015.

[106] Abramson and Carter 2021.

[107] Spruyt 2017.

[108] Cederman 1997; Coggins 2011; Fazal and Griffiths 2014.

empires and the break-up of the communist states at the end of the Cold War. The irredentist threats targeting Ukraine and Taiwan as well as ongoing centrifugal tensions within the United Kingdom, Spain, Ethiopia, Myanmar and other multi-ethnic states of today indicate that the process of nationalist state formation is ongoing and may yet lead to further decline in state size.

References

Abramson, S. (2017). The economic origins of the territorial state. *International Organization 71*, 97–130.

Abramson, S. F. and D. B. Carter (2016). The historical origins of territorial disputes. *American Political Science Review 110*(4), 675–698.

Abramson, S. F. and D. B. Carter (2021). Systemic instability and the emergence of border disputes. *International Organization 75*(1), 103–146.

Alesina, A. and B. Reich (2015a). Nation building. NBER WP 18839.

Alesina, A. and B. Reich (2015b). What's happening to the number and size of nations? E-International Relations.

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

Bean, R. (1973). War and the birth of the nation state. *Journal of Economic History 33*, 203–221.

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

Bhavnani, R., K. Donnay, D. Miodownik, M. Mor, and D. Helbing (2014). Group segregation and urban violence. *American Journal of Political Science 58*(1), 226–245.

Brancati, D. (2006). Decentralization: Fueling the fire or demapening the flames of ethnic conflict and secessionism. *International Organization 60*, 651–685.

Brubaker, R. (1996). *Nationalism Reframed: Nationhood and the National Question in the New Europe*. Cambridge: Cambridge University Press.

Brubaker, R. (2010). Charles tilly as a theorist of nationalism. *The American Sociologist 41*, 375–381.

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

Cederman, L.-E. (1996). Rerunning history: Counterfactual simulation in world politics. In P. E. Tetlock and A. Belkin (Eds.), *Counterfactual Thought Experiments in World Politics*. Princeton: Princeton University Press.

Cederman, L.-E. (1997). *Emergent Actors in World Politics: How States and Nations Develop and Dissolve*. Princeton: Princeton University Press.

Cederman, L.-E. (2013). Nationalism and ethnicity. In W. Carlsnaes, T. Risse, and B. Simmons (Eds.), *The Handbook of International Relations* (2nd ed.). London: Sage.

Cederman, L.-E., L. Girardin, P. Galano Toro, and G. Schvitz (2023). War did make states: Revisiting the bellicist paradigm in early modern europe. *International Organization*.

Cederman, L.-E., S. Rüegger, and G. Schvitz (2022). Redemption through rebellion: Border change, lost unity and nationalist conflict. *American Journal of Political Science 66*(1), 24–42.

Chazan, N. (1991). Introduction: Approaches to the study of irredentism. In N. Chazan (Ed.), *Irredentism and International Politics*. Boulder: Lynne Rienner.

Coggins, B. (2011). Friends in high places: International politics and the emergence of states from secessionism. *International Organization 65*, 433–467.

Coggins, B. (2014). *Power Politics and State Formation in the Twenthieth Centuiry: The Dynamics of Recognition*. Cambridge: Cambridge University Press.

Elias, N. (1982). *The Civilizing Process: State Formation and Civilization*, Volume 2. Oxford: Basil Blackwell.

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

Fazal, T. M. and R. D. Griffiths (2014). Membership has its privileges: The changing benefits of statehood. *Review of International Studies 16*(1), 79–106.

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth 8*, 195–222.

Friedman, D. (1977). A theory of the size and shape of nations. *Journal of Political Economy 87*(1), 59–77.

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

Germann, M. and N. Sambanis (2021). Political exclusion, lost autonomy, and escalating conflict over self-determination. *International Organization 75*(1), 178–203.

Goldewijk, K. K., A. Beusen, G. van Drecht, and M. de Vos (2011). The hyde 3.1 spatially explicit database of human induced land use change over the past 12,000 years. *Global Ecology and Biogeography 20*, 73–86.

Griffiths, R. D. (2010). Security threats, linguistic homogeneity, and the necessary conditions for political unification. *Nations and Nationalism 16*, 169–188.

Griffiths, R. D. (2015). Between dissolution and blood: How administrative lines and categories shape secessionist outcomes. *International Organization 69*(3), 731–751.

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

Griffiths, R. D. and C. R. Butcher (2013). Introducing the international: System(s) dataset (isd), 1816–2011. *International Interactions 39*(5), 748–768.

Grzymala-Busse, A. (2019). Beyond war and contracts: The medieval and religious roots of the european state. *Annual Review of Political Science 23*, 2.1–2.18.

Hale, H. E. (2000). The parade of sovereignties: Testing theories of secession in the soviet setting. *British Journal of Political Science 30*(1), 31–56.

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

Hechter, M. (2000). *Containing Nationalism*. Oxford: Oxford University Press.

Hiers, W. and A. Wimmer (2013). Is nationalism the cause or consequence of the end of empire? In J. A. Hall and S. Malesevic (Eds.), *Nationalism and War*. Cambridge: Cambridge University Press.

Hinsley, F. H. (1973). *Nationalism and the International System*. London: Hodder and Stoughton.

Hintze, O. (1975). The formation of states and constitutional development: A study in history and politics. In F. Gilbert (Ed.), *The Historical Essays of Otto Hintze*. New York: Oxford University Press.

Holsti, K. J. (1991). *Peace and War: Armed Conflicts and International Order, 1648-1989*. Cambridge: Cambridge University Press.

Horowitz, D. L. (1985). *Ethnic Groups in Conflict*. Berkeley: University of California Press.

Horowitz, D. L. (1991). Irredentas and secesssions: Adjacent phenomena. In N. Chazan (Ed.), *Irredentism and International Politics*. Boulder: Lynne Rienner.

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

Kann, R. A. (1974). *A History of the Habsburg Empire 1526-1918*. Berkeley: University of California Press.

Kornprobst, M. (2008). *Irredentism in European Politics: Argumentation, Compromise and Norms*. Cambridge: Cambridge University Press.

Lake, D. A. and A. O'Mahony (2004). The incredible shrinking state: Explaining change in the territorial size of nations. *Journal of Conflict Resolution 48*(5), 699–722.

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

Mayall, J. (1990). *Nationalism and International Society*. Cambridge: Cambridge University Press.

McGarry, J. and B. O'Leary (1993). *Introduction: The Macro-political Regulation of Ethnic Conflict*. London: Routledge.

Michalopoulos, S. and E. Papaioannou (2016). The long-run effects of the scramble for africa. *American Economic Review 106*(7), 1802–1848.

Miller, B. (2007). *States, Nations, and the Great Powers: The Sources of Regional War and Peace*. Cambridge: Cambridge University Press.

Murdock, G. P. (1981). *Atlas of World Cultures*. Pittsburgh, PA: University of Pittsburgh Press.

O'Leary, B., I. S. Lustick, and T. Callaghy (2001). *Right-Sizing the State: The Politics of Moving Borders*. Oxford: Oxford University Press.

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

Reed, F. (2008). Centennia Historical Atlas.

Roeder, P. G. (2007). *Where Nation-States Come From: Institutional Change in the Age of Nationalism*. Princeton, New Jersey: Princeton University Press.

Roshwald, A. (2001). *Ethnic Nationalism and the Fall of Empires: Central Europe, Russia and the Middle East, 1914–1923*. London: Routledge.

Roshwald, A. (2015). *On Nationalism*, Volume 3. Cambridge: Cambridge University Press.

Saideman, S. M. and R. W. Ayres (2000). Determining the causes of irredentism: Logit analyses of minorities at risk data from the 1980s and 1990s. *Journal of Politics 62*(4), 1126–1144.

Saideman, S. M. and R. W. Ayres (2008). *For Kin or Country: Xenophobia, Nationalism, and War*. Columbia: Columbia University Press.

Sambanis, N. and B. Milanovic (2014). Explaining regional autonomy differences in decentralized countries. *Comparative Political Studies 47*(13), 1830 – 1855.

Schieder, T. (1991). *Nationalismus und Nationalstaat: Studien zum nationalen Problemen im modernen Europa*. Göttingen: Vandenhoeck und Ruprecht.

Schvitz, G., L. Girardin, S. Rüegger, N. B. Weidmann, L.-E. Cederman, and K. S. Gleditsch (2022). Mapping the international system, 1886-2019: The cshapes 2.0 dataset. *Journal of Conflict Resolution 66*(1), 144–161.

Sharpe, L. J. (1989). Fragmentation and territoriality in the european state system. *Internaitonal Political Science Review 10*(3), 223–238.

Siroky, D. S. and C. W. Hale (2017). Inside irredentism: A global empirical analysis. *American Journal of Political Science 61*(1), 117–128.

Smith, A. D. (1986). *The Ethnic Origins of Nations*. Oxford: Basil Blackwell.

Spruyt, H. (2017). War and state formation: Amending the bellicist theory of state making. In L. B. Kaspersen and J. Strandsbjerg (Eds.), *Does War Make States? Investigations of Charles Tilly's HIstorical Sociology*. Cambridge: Cambridge University Press.

Stuurman, S. (1995). A millennium of european state formation. *International Review of Social History 40*, 425–441.

Tilly, C. (1978). *From Mobilization to Revolution*. New York: McGraw-Hill.

Tilly, C. (1992). *Coercion, Capital, and European States, AD 990-1990*. Oxford: Basil Blackwell.

Van Evera, S. (1994). Hypotheses on nationalism and war. *International Security 18*(4), 5–39.

Walter, B. F. (2006). Information, uncertainty, and the decision to secede. *International Organization 60*, 105–135.

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

Weidmann, N. B., J. K. Rød, and L.-E. Cederman (2010). Representing ethnic groups in space: A new dataset. *Journal of Peace Research 47*(4), 87–119.

Weiner, M. (1971). The macedonian syndrome: An historical model of international relations and political development. *World Politics 23*(4), 665–683.

White, G. W. (2004). *Nation, State, and Territory: Origins, Evolutions, and Relationships*. Lanham: Rowman Littlefield.

Wimmer, A. and Y. Feinstein (2010). The rise of the nation-state across the world, 1816 to 2001. *American Sociological Review 75*(5), 764–790.

Wood, J. R. (1981). Secession: A comparative analytical framework. *Canadian Journal of Political Science 14*(1), 107–134.

Wucherpfennig, J., N. B. Weidmann, L. Girardin, L.-E. Cederman, and A. Wimmer (2011). Politically relevant ethnic groups across space and time: Introducing the geoepr dataset. *Conflict Management and Peace Science 28*(5), 423–437.

Yack, B. (2001). Popular sovereignty and nationalism. *Political Theory 29*(4), 517–536.

Zacher, M. W. (2001). The territorial integrity norm: International boundaries and the use of force. *International Organization 55*, 215–250.

Supplementary Material: Nationalism and the Puzzle of Reversing State Size

A1 Changes in ethnic boundaries vs. state borders

This section assesses to what extent ethnic settlement areas became more aligned with state borders, a process that would carry the risk of biasing our results. We do so by tracing how similar ethnic geographies observed over the entire time span were to state borders observed in 1886. In addition, we ask the reverse question of how similar state borders observed over time were to ethnic geographies in 1886.

We measure the similarity between ethnic and state geographies using a *Mutual Information* measure from information theory that assesses the amount of information one partitioning A (ethnic settlement areas) carries about another partitioning B (states) of the same set of points.[1] These points are the centroids of a hexagonal grid that covers the European landmass. Our normalized mutual information (MI) metric is defined as

$$MI(A,B) = H(A) - H(A|B) \tag{4}$$

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

where $H(A)$ and $H(A|B)$ are the (conditional) entropies of the partitioning A. While MI returns the quantity of information A carries on B in bits, MI_{norm} adjusts that information with the entropy of partitionings A and B, yielding a measure that strictly varies between 0 (no mutual information) and 1 (full mutual information).

Figure A1 plots the results from contrasting stable and time-variant ethnic and state borders. As one would expect, the upper left quadrant shows constant mutual information between state borders in 1886 and ethnic geographies in 1886. More interestingly, the upper right quadrant shows a slight increase in the information time variant ethnic data carries on time-invariant state borders from 1886. This suggest that there is a marginal adjustment of ethnic geographies to historical state borders over time. However, changes observed in the bottom left quadrant are much larger. Here, we fix our ethnic data to reflect the year 1886 and let only state borders vary. We see that, over time, the information that these borders contain on past ethnic geographies increases substantially from .68 to .82. Changes in state borders thus account for most of the increasing similarity of state and ethnic geographies observed in the lower right quadrant. This asymmetry show that it makes sense to study how state borders adjust to ethnic boundaries.

[1] Vinh et al. 2010.

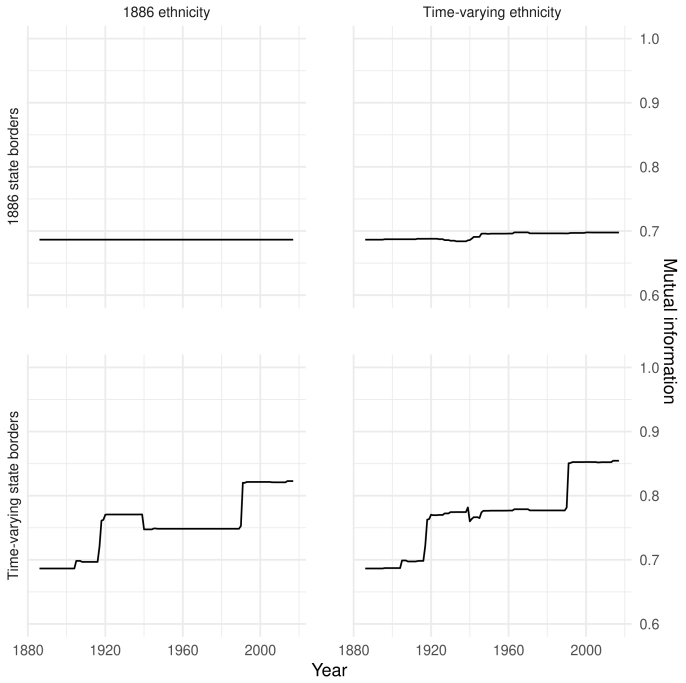

Figure A1: Mutual information among (time-varying) data on ethnic geographies and states (CShapes-Europe and HEG data).

A2 Trends in the number of states

As a complement to the plots showing the trend in average state size, we present corresponding trends in the number of states, both for the European and global samples (see Figures A2 and A3 respectively).

Figure A2: Trend in number of states in Europe (CShapes-Europe)

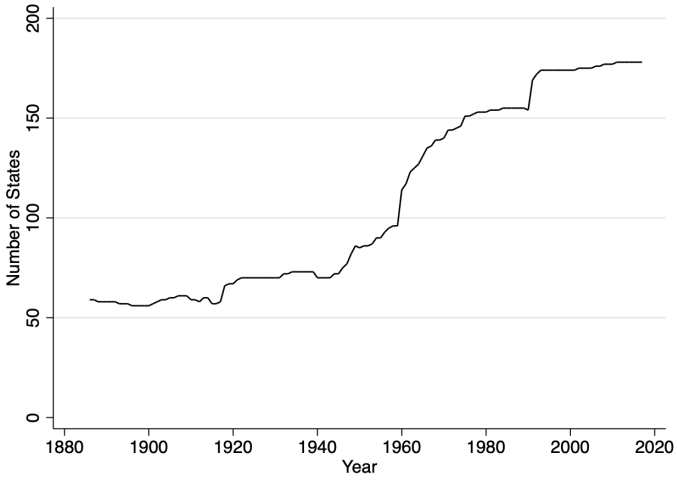

Figure A3: Trend in number of states around the world (CShapes 2.0).

A3 Robustness analysis

We start this section but presenting full versions of Tables 3 and 4 including all control variables, see Tables A1 and A2.

In this section, we first provide some supplementary information about the event-based estimation. Referring to the European and global samples respectively, Tables A3 and A4 display estimations based on all events rather than merely ethnic ones. For obvious reasons, the results are consistent with the analysis restricted to ethnic border-change events.

We continue by testing some alternative model specifications with the main datasets. We focus on the robustness of the main results from Tables 3 and 4. First, we present the result of using linear models rather than Cox models (see Tables A5 and A6).

Since Russia much of the parts that were added to Russia during the 19th century are not part of Europe, we show that the results in our main models are not effected when Russia/USSR is removed from the analysis (see Tables A7 and A8).

We now continue to investigate the robustness of the main findings by modifying the underlying data in various ways. To guard against endogeneity of the historical ethnicity data in HEG, we use the earliest snapshot from this dataset dating back to the late 19th century. However, the findings of this alternative analysis do not offer any surprises in a replication of the gain/loss analysis (see Table A9) or events analysis (see Table A10). Figure A4 displays the ethnic and non-ethnic trajectories. All in all, the results do not deviate noticeably from the main analysis, except for unstable results for unification since both the German and Italian unification processes having been dropped .

There is less that can be done in the global sample, since early ethnicity data are available for the global sample. Yet, the backward projection can be limited by analyzing the ANM data with state borders from 1946. Table A11 shows that this restriction weakens our results for territorial gains, but studying events, the effect of territorial fractionalization is still present for irredentist configurations (see Table A12). Moreover, the results in the

Table A1: Losses and gains, World 1886-2017, Cox proportional hazard models

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	2.39940***	1.60299*	0.78867+	0.09049
	(0.59090)	(0.70161)	(0.44308)	(0.46818)
Terr. frac.	1.47573***	1.41518*	1.86185***	3.13774***
	(0.38627)	(0.59367)	(0.45900)	(0.67929)
State age, log	0.23236+	0.36349	-0.09655	-0.25618
	(0.14122)	(0.24266)	(0.11121)	(0.22178)
State area, log	0.09230	-0.12405	0.43798**	0.55667**
	(0.10451)	(0.16050)	(0.14798)	(0.19228)
State population, log	-0.11957	0.19015	-0.11307	-0.15248
	(0.11050)	(0.17582)	(0.14295)	(0.19021)
Elev. mean, log	-0.18733	-0.07578	-0.53492+	-0.32137
	(0.14763)	(0.19459)	(0.27864)	(0.29413)
Elev. SD, log	0.01493	0.01397	0.55004+	0.41333
	(0.16502)	(0.22779)	(0.30139)	(0.29920)
Observations	7610	7610	7610	7610
Pseudo R^2	0.044	0.071	0.076	0.129
Strata	No	Yes	No	Yes

Standard errors in parentheses
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

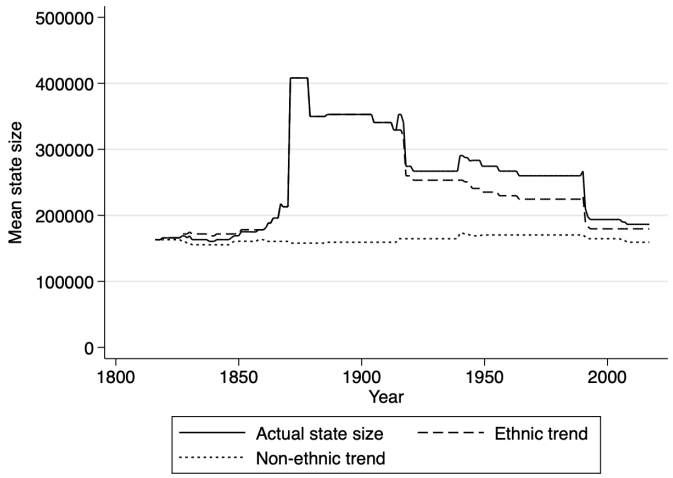

Figure A4: Comparing ethnic and non-ethnic trends in mean state size with actual trend in Europe based on CShapes-Europe state borders with fixed HEG data from the 19th century.

main model become entirely robust if we rely on GeoEPR rather than ANM ethnic data (not shown).

Table A2: Losses and gains, World 1886-2017, Cox proportional hazard models

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	0.82385**	1.05917**	0.12653	-0.02818
	(0.29587)	(0.36244)	(0.47839)	(0.42607)
Terr. frac.	-0.54804	0.50367	0.67408	2.40995***
	(0.49031)	(0.55961)	(0.65656)	(0.67391)
State age, log	0.35840**	0.44516**	-0.19960	-0.01250
	(0.12122)	(0.17230)	(0.12275)	(0.17809)
State area, log	0.16324*	-0.00621	0.21676**	0.02589
	(0.06483)	(0.08487)	(0.07970)	(0.08867)
State population, log	0.04579	0.26089**	0.36677***	0.55497***
	(0.07693)	(0.08994)	(0.11055)	(0.10640)
Elev. mean, log	0.00677	-0.19672	-0.05773	0.17524
	(0.22284)	(0.24773)	(0.29614)	(0.25952)
Elev. SD, log	0.09128	0.13141	0.03294	-0.38205+
	(0.19302)	(0.26464)	(0.27948)	(0.22316)
Observations	13849	13849	13849	13849
Pseudo R^2	0.094	0.249	0.070	0.137
Strata	No	Yes	No	Yes

Standard errors in parentheses

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: All border-change events, Europe 1816-2017, Cox proportional hazard models

	(1) Secession	(2) Secession	(3) Absorption	(4) Absorption	(5) Irredentism	(6) Irredentism
Ethnic frac.	5.82098***	7.60395***	0.57997	-0.41739	-0.01459	-0.80025
	(1.29177)	(2.04073)	(0.51434)	(1.06949)	(0.45773)	(0.55517)
Terr. frac.	-2.02323+	-2.91595	3.80763***	4.69574*	1.71031***	3.09542***
	(1.21034)	(1.98802)	(0.74635)	(1.91946)	(0.48348)	(0.86542)
Observations	7610	7610	7610	7610	7610	7610
Pseudo R^2	0.297	0.513	0.150	0.222	0.060	0.123
Strata	No	Yes	No	Yes	No	Yes

Standard errors in parentheses

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Nor do we find any important differences when we replace the CShapes-Europe data with data on state borders from Centennia (see Tables A13 and A14). The trend analysis also does not deviate from the main analysis (see Figure A5).[2]

[2]This analysis relies on hard-coded state continuation coding since Centenna does not offer information on capitals.

Table A4: All border-change events, World 1886-2017, Cox proportional hazard models

	(1) Secession	(2) Secession	(3) Absorption	(4) Absorption	(5) Irredentism	(6) Irredentism
Ethnic frac.	1.10271^{+}	2.61620**	0.78428	1.12358	-0.64469	-1.05741*
	(0.63625)	(0.95776)	(0.73307)	(0.99196)	(0.48184)	(0.52467)
Terr. frac.	-4.98254*	-3.25433	3.12049*	4.14394*	0.81679	2.58066***
	(2.35186)	(2.27974)	(1.25313)	(2.10925)	(0.68912)	(0.69424)
Observations	13849	13849	13849	13849	13849	13849
Pseudo R^2	0.247	0.665	0.154	0.465	0.050	0.136
Strata	No	Yes	No	Yes	No	Yes

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: Log Losses and gains, Europe 1816-2017, OLS

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	0.41883***	0.69800**	0.23824***	0.06232
	(0.11074)	(0.21346)	(0.06826)	(0.31735)
Terr. frac.	0.20872***	0.12329	0.24133***	0.58476**
	(0.05732)	(0.14916)	(0.06306)	(0.18564)
Observations	7610	7610	7610	7610
Pseudo R^2				
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
State FE	No	Yes	No	Yes

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6: Log Losses and gains, World 1886-2017, OLS

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	0.24551*	2.12885***	0.06352	-0.87337
	(0.10638)	(0.56879)	(0.08033)	(0.76168)
Terr. frac.	0.02212	-0.31767	0.07969	0.96816*
	(0.05215)	(0.28347)	(0.05501)	(0.38510)
Observations	13849	13849	13849	13849
Pseudo R^2				
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
State FE	No	Yes	No	Yes

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A7: Losses and gains, Europe 1816-2017 without Russia/USSR, Cox proportional hazard models

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	2.36506***	1.28625+	0.71349	0.04936
	(0.63528)	(0.72903)	(0.45855)	(0.45136)
Terr. frac.	1.52700***	1.21830*	1.84785***	3.48083***
	(0.38220)	(0.55293)	(0.45442)	(0.80275)
Observations	7409	7409	7409	7409
Pseudo R^2	0.044	0.065	0.072	0.136
Strata	No	Yes	No	Yes

Standard errors in parentheses
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A8: Losses and gains, World 1946-2017 without Russia/USSR, Cox proportional hazard models

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	0.74989**	0.98595**	0.18766	0.06388
	(0.28873)	(0.36703)	(0.47162)	(0.40706)
Terr. frac.	-0.40345	0.54330	0.74892	2.35781***
	(0.46461)	(0.56754)	(0.66917)	(0.67724)
Observations	13718	13718	13718	13718
Pseudo R^2	0.097	0.257	0.074	0.140
Strata	No	Yes	No	Yes

Standard errors in parentheses
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A9: Losses and gains, Europe 1886-2017, Cox proportional hazard models (fixed HEG data)

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	2.69121*	2.38176*	1.17394	0.59641
	(1.09935)	(0.95607)	(0.84622)	(0.74989)
Terr. frac.	1.98000**	5.26485**	2.22522**	4.98087***
	(0.72861)	(1.67102)	(0.79538)	(1.35930)
Observations	4230	4230	4230	4230
Pseudo R^2	0.066	0.228	0.047	0.162
Strata	No	Yes	No	Yes

Standard errors in parentheses
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Another important sensitivity test at the global level uses the CShapes 2.0 dataset and ANM/GREG data without any colonial dependencies (see Table A15 and A16). Figure

Table A10: Border-change events, Europe 1886-2017, Cox proportional hazard models

	(1) Eth. Secession	(2) Eth. Secession	(3) Irredentism	(4) Irredentism
Ethnic frac.	6.02558*	5.17404*	-1.27222	-1.76509
	(2.43037)	(2.47276)	(1.40012)	(1.21075)
Terr. frac.	-0.76602	3.17801	0.44874	2.79375+
	(3.82198)	(4.50762)	(1.07164)	(1.61145)
Observations	4230	4230	4230	4230
Pseudo R^2	0.343	0.462	0.044	0.183
Strata	No	Yes	No	Yes

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A11: Losses and gains, World 1946-2017, Cox proportional hazard models

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	1.39709**	1.69192***	-0.84301	-1.17348
	(0.43049)	(0.50405)	(0.62911)	(0.76171)
Terr. frac.	-0.73229	0.76221	0.20690	0.63823
	(0.85644)	(1.62503)	(1.20567)	(1.21471)
Observations	10142	10142	10142	10142
Pseudo R^2	0.201	0.486	0.052	0.078
Strata	No	Yes	No	Yes

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A12: Border-change events, World 1946-2017, Cox proportional hazard models

	(1) Eth. Secession	(2) Eth. Secession	(3) Irredentism	(4) Irredentism
Ethnic frac.	1.91442**	4.05991*	-3.72845**	-7.04169
	(0.70470)	(1.83266)	(1.19221)	(4.76537)
Terr. frac.	-8.92798*	-3.94026	5.73808***	10.69080*
	(4.20398)	(3.97679)	(1.63766)	(4.59172)
Observations	10142	10142	10142	10142
Pseudo R^2	0.324	0.759	0.269	0.585
Strata	No	Yes	No	Yes

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A6 presents the resulting trajectories. Again, these findings confirm the robustness of our analysis.

Table A13: Losses and gains, Europe 1816-2017, Cox proportional hazard models (Centennia & HEG data)

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	1.43415***	1.21105***	-0.06228	-0.29435**
	(0.27989)	(0.25118)	(0.20731)	(0.11070)
Terr. frac.	-0.29867	0.33964	1.36614***	0.21129**
	(0.27147)	(0.32206)	(0.23466)	(0.06870)
Observations	7780	7780	5933	5933
Pseudo R^2	0.028	0.029	0.020	0.007
Strata	No	Yes	No	Yes

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A14: Border-change events, Europe 1816-2017, Cox proportional hazard models (Centennia HEG data)

	(1) Eth. Secession	(2) Eth. Secession	(3) Unification	(4) Unification	(5) Irredentism	(6) Irredentism
Ethnic frac.	2.22973***	1.84707**	-1.64976*	-1.46082*	-0.21603	-0.15909
	(0.60882)	(0.63725)	(0.70191)	(0.61592)	(0.53129)	(0.43609)
Terr. frac.	-0.87131	0.64852	4.76901***	5.11511***	1.58791***	1.73097**
	(0.65535)	(0.61853)	(0.55991)	(1.17417)	(0.43556)	(0.59742)
Observations	7780	7780	7780	7780	7780	7780
Pseudo R^2	0.097	0.192	0.125	0.257	0.027	0.044
Strata	No	Yes	No	Yes	No	Yes

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To further test the robustness of our findings, we analyze a version of the regression models that uses data from the GeoEPR dataset[3] which are provided in aggregate form.Cederman et al. 2022. (see Tables A17 and A18). Using GeoEPR rather than GREG/ANM, Figure A7 displays a graph with the three trajectories that resemble our previous results.

We also compare the effects we document in the main analysis for post-Napoleonic Europe with the preceding macro-historical period. Table A19 presents the main territorial gain and loss models with data provided by Abramson[4] covering the time between 1490 and 1790. In comparison to the main results, these analyses show no consistent or statistically significant effects of ethnic and territorial fractionalization on losses and gains and the main effects of interest have reversed signs.

Finally, we test an alternative temporal understanding of our main results. Namely, one could argue that they are driven entirely by the "age of secession" that followed World War

[3]Wucherpfennig et al. 2011; Vogt et al. 2015.
[4]Abramson 2017.

Figure A5: Comparing ethnic and non-ethnic trends in mean state size with actual trend in Europe based on Centennia state borders with fixed HEG data.

Table A15: Losses and gains, World 1886-2017 without dependencies, Cox proportional hazard models

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	0.86827^{+}	1.79298**	-0.19655	-0.50459
	(0.49053)	(0.56103)	(0.42624)	(0.55299)
Terr. frac.	0.09213	1.61329*	1.12021	2.48761**
	(0.48299)	(0.69899)	(0.69201)	(0.88016)
Observations	13841	13841	13841	13841
Pseudo R^2	0.050	0.120	0.044	0.122
Strata	No	Yes	No	Yes

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

II[5] with nationalism before the Great War not having any effect. We test this conjecture by introducing a "post-45" dummy variable and its interaction with ethnic and territorial fractionalization in Table A20. While ethnic fractionalization has a larger effect on territorial losses after 1945, the effect is positive, statistically significant, and only slightly smaller than in the main analysis for the time before 1945. This result shows that the effect of nationalism is not limited to post-war Europe.

[5]Griffiths 2016.

Table A16: Border-change events, World 1886-2017 without dependencies, Cox proportional hazard models

	(1) Eth. Secession	(2) Eth. Secession	(3) Irredentism	(4) Irredentism
Ethnic frac.	2.47469**	3.29942+	-2.27791*	-1.75627
	(0.83270)	(1.69131)	(0.92579)	(1.17970)
Terr. frac.	-0.24692	2.18970	2.36727**	4.52740***
	(1.38135)	(3.63092)	(0.80883)	(1.03249)
Observations	13841	13841	13841	13841
Pseudo R^2	0.229	0.423	0.079	0.225
Strata	No	Yes	No	Yes

Standard errors in parentheses
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

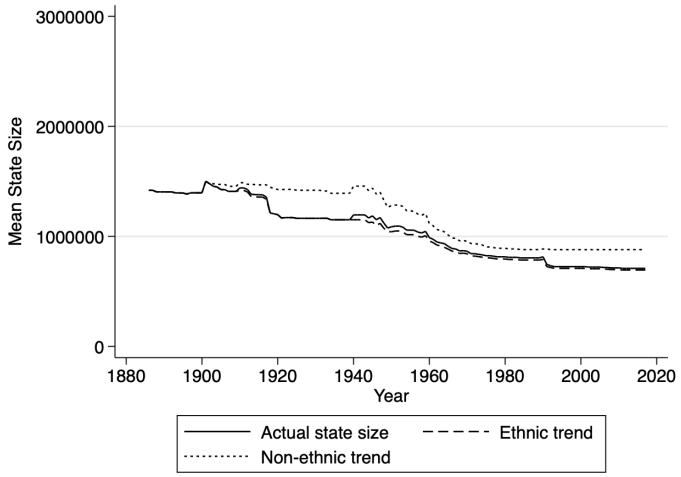

Figure A6: Comparing ethnic and non-ethnic trends in mean state size with actual trend worldwide without colonies based on CShapes 2.0 state borders and backdated ANM/GREG data.

Table A17: Losses and gains, World 1886-2017, Cox proportional hazard models, GeoEPR ethnicity data

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	0.85998*	1.06635**	0.61356	0.66891
	(0.33773)	(0.40912)	(0.51438)	(0.43780)
Terr. frac.	-0.08634	0.00072	0.51937	1.30709*
	(0.32330)	(0.44678)	(0.42526)	(0.51374)
Observations	13587	13587	13586	13586
Pseudo R^2	0.089	0.241	0.071	0.128
Strata	No	Yes	No	Yes

Standard errors in parentheses
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A18: Border-change events, World 1886-2017, Cox proportional hazard models, GeoEPR ethnicity data

	(1) Eth. Secession	(2) Eth. Secession	(3) Irredentism	(4) Irredentism
Ethnic frac.	1.77186*	4.37455***	-0.83257	-0.75341
	(0.74876)	(1.13273)	(0.52685)	(0.79096)
Terr. frac.	-0.20593	-0.94183	0.90202+	1.96095*
	(0.68389)	(0.80570)	(0.46438)	(0.77675)
Observations	13587	13587	13587	13587
Pseudo R^2	0.241	0.695	0.042	0.087
Strata	No	Yes	No	Yes

Standard errors in parentheses

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

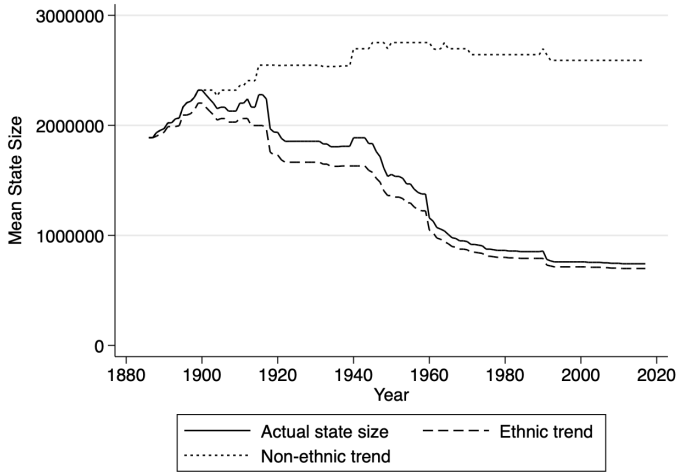

Figure A7: Comparing ethnic and non-ethnic trends in mean state size with actual trend worldwide based on CShapes 2.0 state borders and backdated aggregated GeoEPR data.

Table A19: Losses and gains, Europe 1490-1790, Cox proportional hazard models

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
Ethnic frac.	-0.19309	-0.11826	-0.30652*	-0.22527
	(0.14699)	(0.14204)	(0.14871)	(0.14689)
Terr. frac.	-0.07005	-0.02743	-0.07855	-0.02388
	(0.13233)	(0.12528)	(0.11906)	(0.12603)
Observations	13073	13073	13075	13075
Pseudo R^2	0.004	0.004	0.008	0.008
Strata	No	Yes	No	Yes

Standard errors in parentheses

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A20: Losses and gains, Europe 1816-2017, Cox proportional hazard models: Pre- vs. post-1945

	(1) Loss	(2) Loss	(3) Gain	(4) Gain
EEthnic frac.	1.48584*	1.45058*	0.30747	-0.13635
	(0.61932)	(0.72453)	(0.43071)	(0.40531)
Terr. frac.	1.11214***	1.40458*	1.46572***	2.98163***
	(0.32314)	(0.57863)	(0.40269)	(0.67492)
Post-45 (0/1)	-2.84500***		-2.76389***	
	(0.65087)		(0.74636)	
Ethnic frac. x post-45	3.92543***	4.52244*	2.34889+	1.41894
	(1.03643)	(2.12550)	(1.27036)	(3.68456)
Terr. frac. x post-45	1.48243	3.60142	1.72501	0.58858
	(1.42824)	(2.85232)	(1.53627)	(6.23489)
Observations	7610	7610	7610	7610
Pseudo R^2	0.065	0.075	0.095	0.127
Strata	No	Yes	No	Yes

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A4 Tracing general border change in Europe and the global state system

In this section, we provide complementary bar chars that depict general border changes. Based on the definition of border change, we explore the empirical trajectory of the European state system. Using CShapesHist state borders and HEG ethnic boundaries starting from the late 19th century, Figure A8 shows each border change event broken down into three of the four main categories. We display cases of absorbed states in green, transfers in orange, and secession in red.

The European system saw more cases of absorption in the 19th century, reflecting the German and Italian unification processes, as well as several cases of conquest triggered by Germany in the 1940s. German reunification is clearly visible as a green dot in 1990. In contrast, secession became increasingly common in the course of the 20th century. The opposite tendency applies to transfer events, whose frequency culminated during the world wars, but decreased in frequency since then.

At the global level, Figure A9 tells us that secession dominates even more than within the European state system, which is not surprising since the European analysis excludes colonies. Absorption, in contrast, happens relatively rarely. Furthermore, transfer also becomes rarer, especially during the second part of the 20the century. This development reflects the emergence and consolidation of the territorial integrity norm that made violent

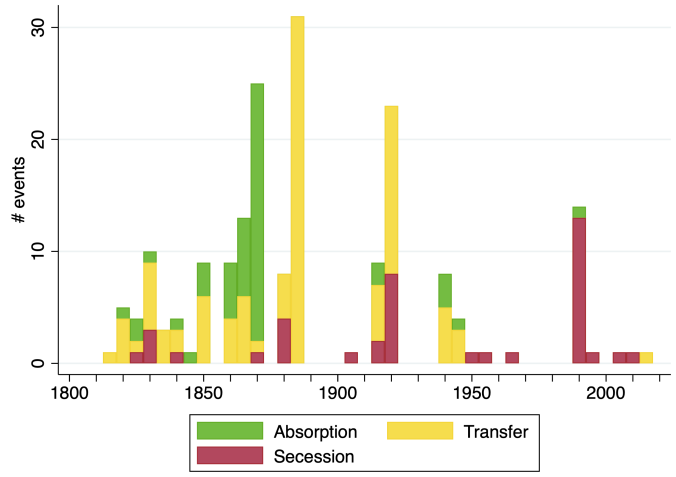

Figure A8: Border change events in Europe based on CShapes-Europe state borders

border change illegal.[6] However, Russia's conquest of the Crimea in 2014 represents an exception from this trend (see the rightmost orange bar).

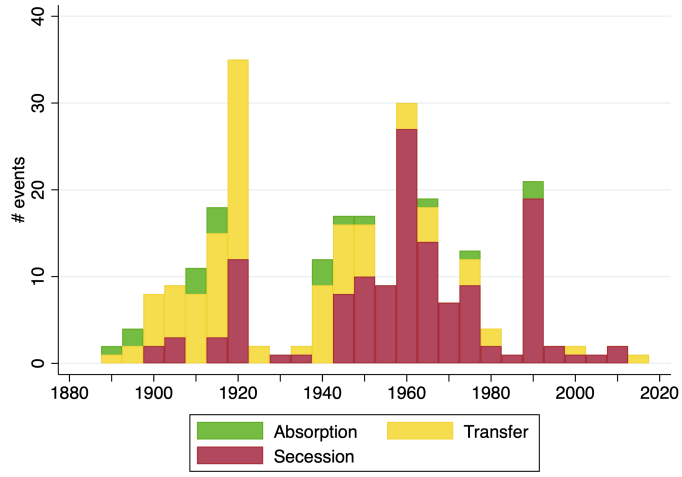

Figure A9: Border change events around the world based on CShapes 2.0 state borders.

A5 Derivation of ethnic and non-ethnic state-size trajectories

This section offers additional details on how to generate the ethnic and non-ethnic trajectories shown in Figures 9 and 10. The analysis relies on the following definition of mean state size:

[6] Zacher 2001.

$$\overline{S}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} S_{it}$$

where S_{it} is the territorial size of state i at time t and N_t the number of states in the system at time t.

From the above expression, it is easy to see that average state size only depends on the number of states because, given the zero-sum nature of border change, the total surface of the system $\sum_i S_{it}$ remains unaffected by such events. The only exception is expansion or retraction of the system with its outer environment, for instance through conquest of previously stateless territory. Holding constant the spatial extent of the state system, only border changes that involve unit processes—state births and deaths—are able to affect the average size of states. Transfer cannot affect this variable.[7]

The other three types of border change can, and usually do, affect the number of states in the system, and therefore also mean state size. In most cases, border changes involve only one state dyad. Absorption typically removes one state from the total number of states: $N_{t+1} = N_t - 1$. Correspondingly, secession usually increments the number of states: $N_{t+1} = N_t + 1$. A strictly dyadic collapse/merger event will leave the total number of states unaffected to the extent that there is only one birth and one death.

A more complicated procedure is needed where secession and absorption involve more than one dyad. The border change events that increase the total number of states (i.e. secessions $SecA_t$ and collapses/mergers of $CollA_t$ of State A), as well as those that decrease this number (i.e. absorption events $AbsB_t$ and collapses/mergers $CollB_t$ suffered by State B), have to be normalized for each gaining and losing state respectively, because the same newborn state can emerge from more than one instance of secession and/or collapse. Likewise, one state death may benefit more than one state, as illustrated by the partitioning of Germany in 1945. (Note that we refer to both collapses and mergers as $CollB_t$ for simplicity.)

We can now compute $SecA_t$ and $CollA_t$ as the sum of the weighted contribution to the birth of all states for each time period t:

$$SecA_t = \sum_i \frac{1}{Birth_{it}} \sum_{j=i} SecA_{jt}$$

$$CollA_t = \sum_i \frac{1}{Birth_{it}} \sum_{j=i} CollA_{jt}$$

where

$$Birth_{at} = \sum_{j=a} SecA_{jt} + CollA_{jt}$$

is the total number of border change dyads that give rise to state a.

Similarly, we compute $AbsB_t$ and $CollB_t$ similarly:

[7] This is obviously not true for other measures of average state size, e.g., the mean of logarithmic size.

$$AbsB_t = \sum_i \frac{1}{Death_{it}} \sum_{j=i} AbsB_{jt}$$

$$CollB_t = \sum_i \frac{1}{Death_{it}} \sum_{j=i} CollB_{jt}$$

where

$$Death_{bt} = \sum_{j=b} AbsB_{jt} + CollB_{jt}$$

is the total number of border change dyads that eliminate state b in time period t

We can now turn to the derivation of the trajectories themselves. Assuming that N_0, the number of states in the system at the beginning of the sample, is known, we can now compute the total (T) number of states N_t^T at time t as a function of cumulative border changes ΔN_t^T provided that we know how many states seceded, $SecA_t$, how many states were born out of collapse/merger, $CollA_t$, how many states were absorbed, $AbsB_t$, and how many states lost their lives through collapse/merger $CollB_t$:

$$N_t^T = N_{t-1}^T + \Delta N_t^T$$

where

$$\Delta N_t^T = SecA_t + CollA_t - AbsB_t - CollB_t.$$

The arithmetic mean of state size in any year t can now be computed as:

$$\overline{S}_t^T = \frac{S_t}{N_t^T}$$

We now provide more information on how to compute the ethnic and non-ethnic trajectories. First, we can separate out the contribution of ethnic border changes, N_t^E, as defined in the main text from those that occur through ethnic processes:

$$N_t^E = N_{t-1}^E + \Delta N_t^E$$

where the annual net change in the total number of states is

$$\Delta N_t^E = ESecA_t + ECollA_t - EAbsB_t - ECollB_t.$$

where $ESecA_t$, $ECollA_t$, $EAbsB_t$, and $ECollB_t$ represent the number of ethnic secessions, collapses/mergers of State A, absorptions and collapses/mergers of State B.

Ethnic border change thus produces the following average state size:

$$\overline{S}_t^E = \frac{S_t}{N_t^E}.$$

Similarly, we can compute the number of states resulting from non-ethnic border change as a residual once the ethnic border change events have been removed from the respective total border change categories:

$$N_t^{\not E} = N_{t-1}^{\not E} + \Delta N_t^{\not E}$$

where

$$\Delta N_t^{\not E} = \not E SecA_t + \not E CollA_t - \not E AbsB_t - \not E CollB_t.$$

This results in a separate trend in average state size assuming only non-ethnic mechanisms of border change are at work:

$$\overline{S}_t^{\not E} = \frac{S_t}{N_t^{\not E}}.$$

References

Abramson, S. (2017). The economic origins of the territorial state. *International Organization 71*, 97–130.

Cederman, L.-E., S. Rüegger, and G. Schvitz (2022). Redemption through rebellion: Border change, lost unity and nationalist conflict. *American Journal of Political Science 66*(1), 24–42.

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

Vinh, N. X., J. Epps, and J. Bailey (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research 11*(Oct), 2837–2854.

Vogt, M., N.-C. Bormann, S. Rüegger, L.-E. Cederman, P. Hunziker, and L. Girardin (2015). Integrating data on ethnicity, geography, and conflict: The ethnic power relations dataset family. *Journal of Conflict Resolution 59*(7), 1327–1342.

Wucherpfennig, J., N. B. Weidmann, L. Girardin, L.-E. Cederman, and A. Wimmer (2011). Politically relevant ethnic groups across space and time: Introducing the geoepr dataset. *Conflict Management and Peace Science 28*(5), 423–437.

Zacher, M. W. (2001). The territorial integrity norm: International boundaries and the use of force. *International Organization 55*, 215–250.