Nationalism’s rise to power across the world.  
An event history analysis of nation-state formation, 1816-2001

Andreas Wimmer and Yuval Feinstein
Department of Sociology, UCLA

Summary: Comparative historical sociology has debated for decades how to best understand the rise and global spread of the nation-state as the core principle of political organization of the modern world. We test major theories with a new dataset that contains information on 145 of today’s states from 1816 to the year when they achieved nation-statehood. Event-history analysis shows that the likelihood of nation-state creation is influenced on the one hand by diffusion mechanisms within empires and among neighbors and on the other hand by a power shift that allows nationalists to overcome the established regime. This shift is more likely to occur the longer nationalists had mobilized the population, if the political center is weakened by wars, if empires lack the global military and economic power to fight independence movements effectively, or if autonomous states lack governmental capacity to co-opt, control or suppress nationalism. We found no robust evidence for the effects of industrialization, the advent of mass literacy, or increasing administrative penetration and direct rule as the classical theories of economic, cultural or political modernization maintained. We conclude that the global spread of the nation-state is driven by proximate and contextual political factors, rather than domestic structural forces that operate over the long durée.

1 Introduction

Nationalism and the nation-state remain defining features of today’s world. In his inaugural address as the 44th president of the United States, Barack Obama spoke eleven times in the name of “our nation”, addressed “the American people” in four different sentences, and evoked twelve times “America” as a collectivity uniting the government and its citizenry. Few of the millions of Americans who listened to his speech will have noticed this. In old and well established nation-states such as the United States, nationalist principles of political legitimacy are so much taken for granted that they recede to the background of everyday consciousness and become “banal” in Michael Billig’s felicitous terms (Billig 1995)—ready to be activated in times of crisis such as war or economic depression. In other, less established states, nationalism is politically even more salient, as the wars in the wake of the break-up of Yugoslavia and the Soviet Union demonstrated.

Three hundred years ago, the political world looked rather different. Rulers ruled in the name of dynastic succession: a prince was entitled to assume the mantle of power upon the death of his father; or of theocratic authority: the spiritual leaders of a community were guiding their flocks in worldly...
matters as well; or of a world-spanning civilizational mission: imperial elites spread true faith, brought progress to “backward” peoples, or later during the 20th century, advanced a revolutionary project in the name of the world’s working classes. By 2001, empires had dissolved, theocracies dethroned and only a handful of countries, mostly in the Middle East, were still governed as absolutist monarchies. All other states were ruled in the name of a nationally defined people and thus conformed to standard definitions of what makes an independent state a nation-state (Gellner 1983; Calhoun 1997). The story might not yet have come to its end, as the recent independence of East Timor, Montenegro, and Kosovo illustrate. The current struggles for independent statehood of Kurdish, Tibetan, West-Saharan, Tamil, and many other nationalist movements demonstrate the continuing political appeal of the ideal of the nation-state in today’s globalized world.

**Figure 1**
The global spread of the nation-state, 1816-2001

This process is of a discontinuous nature, as figure 1 illustrates, because many nation-states were created as part of waves following the collapse of empires. The first wave led to the dissolution of the Spanish empire in the New World; the second wave occurred after the First War with the break up of the Habsburg empire, another wave after the Second War when the Middle East—which France and Britain had recolonized after the fall of the Ottoman empire—as well as South and Southeast Asia were decolonized; the next followed around 1960, as the British and French colonial empires receded; the fifth wave rolled over the Soviet empire during the early nineties. Not all transitions to the nation-state start from empire, however, since there are many independent states (some Western European states,
Thailand, Ethiopia, Nepal, Japan and so forth) where the transformation was brought about without a mobilization against “alien rule”.

Since the Second World War, comparative historical sociology has invested considerable intellectual energy to understand this transformation of the state-system, from Hans Kohn’s early study of the “idea of nationalism” (Kohn 1944) to the elaboration of various now “classical” theories, mostly by scholars working in the tradition of British historical sociology. Among these classical statements we find various modernization arguments: Ernest Gellner maintained that industrialization requires a flexible and mobile labor force best provided by the educational system of a nation-state (Gellner 1983). Similarly, the communicative integration brought about by urbanization and the spread of markets gives rise to feelings of national belonging, according to Deutsch 1966). Benedict Anderson put the finger on cultural modernization processes and especially the spread of literacy in vernacular languages, which made the imagining of nations possible, to paraphrase his now proverbial formulation (Anderson 1991). Charles Tilly, Michael Hechter and Michael Mann moved political modernization center stage, arguing that the rise of the bureaucratic state at constant war with its neighbors (Mann 1995), and the conforming shift from indirect to direct rule (Tilly 1990; Hechter 2000) brought about the mobilization of the subject population under the banners of nationalism and eventually the nation-state.

Other theories of nationalism and the nation-state are less tied, or even opposed, to the modernist paradigm. Anthony Smith attacked modernism for overlooking the continuity between modern nationalism and older ethnic sentiments that shape the latter and fill it with a transformative politico-emotional energy (Smith 1986). Various diffusionist accounts have emerged since Kedour’s early writings had highlighted the disjunction between imported nation-state ideologies and local political reality (Kedourie 1960): John Meyer and co-authors emphasize the pressure of adopting the nation-state emanating from a hegemonic global culture that includes the nation-state form as one of its organizational templates (Meyer et al. 1997). Rogers Brubaker’s more regionally focused study shows how nation-state ideologies and institutions spread after the end of empire in Eastern Europe (Brubaker 1996). Others highlight the domino effects that nation-state creation might have on its neighbors due to competing claims over territory and populations (Wimmer 2002; 2006).

This rich literature also displays certain weaknesses. First, with the notable exceptions of Anderson (1991) and Breuilly (1993), many theories are meant to explore universal processes and factors that explain the rise of the nation-state in the modern world as a whole, although they are empirically tailored to Western cases. This would be less discomfiting if nationalism and the nation-state had remained confined to the area of its origin—thus producing a universe of cases conforming to the empirical focus of these theories.

Second, more empirically focused research on the particular trajectories of nation-state creation tends to be segmented along regional and disciplinary lines. Thus, the political science literature on decolonization (Spruyt 2005) and “nation-building” (Bendix 1964) in the post-colonial world developed quite independently from the classical debates about the origins of the nation-state in the West. Yet another strand of scholarship, led by historically minded scholars, has investigated how the land-based Ottoman, Habsburg or Soviet empires collapsed and were transformed into a series of competing nation-states (Karpat 1973; Brubaker 1996; Barkey and von Hagen 1997; Roshwald 2001; Saideman and Ayres 2008). Given the world-covering nature of the process of nation-state creation, however, one wonders whether an integrated view that encompasses post-colonial, post-imperial as well the Continental European trajectories might be within reach.
Third, existing research either proceeds as single case studies, or compares a handful of nation-state formations in the tradition of comparative historical sociology (e.g. Greenfeld 1992), or offers a combination of sweeping generalizations and anecdotic evidence (a tendency deplored by Breuilly 2005; Wimmer 2008). A fourth methodological approach is to trace the various trajectories of nationalist mobilization and the establishment of nation-states across territories and continents in a global historical narrative (Mann 1993; Breuilly 1993). Quantitative scholarship is limited to David Strang’s article on the timing of de-colonization (Strang 1990) and Philip Roeder’s analysis of the institutional conditions that lead to independent statehood in the 20th century (Roeder 2007).

The authors believe that it is time to broaden the methodological horizon and to add a more robust quantitative voice to the scholarly choir. Examining recurring patterns on the basis of a global dataset would help overcoming some of the limitations mentioned above. It might identify those classical statements that also apply to processes of nation-state formation outside the old world. It might identify those local processes, empire-specific causal sequences, or epochal trends that are recurring often enough to be identified as patterns of a general nature. While we agree with Charles (Tilly 1995:1600) that “the attempt to build transhistorical models … is doomed to eternal failure”, we also share his conviction that multivariate analysis with its built-in attention to variation and, we should add, its probabilistic notion of causality, might contribute considerably to our understanding of the general circumstances under which the nation-state is more likely to emerge.

Obviously, there are good reasons why quantitatively minded scholars have shied away from this task. As in much research in comparative history, there is a discrepancy between scholarly interest in Tilly’s now proverbial “big structures, large processes, and huge comparisons” (Tilly 1989) on the one hand, and the scarcity and bad quality of data on the other hand. This problem is exacerbated in the particular domain at hand: Since modern nation-states are the prime data-collecting agencies on which cross-national research relies, answering the question of why such states emerge in the first place is confronted with formidable data problems. In recent years, various scholars have attempted to overcome this difficulty by collecting data on pre-independent territories. Among these efforts, Acemoglu’s and his co-authors’ exploration of the colonial determinants of contemporary economic growth is certainly the best known (Acemoglu et al. 2001). Mahoney and associates as well as Wimmer and Min (2006) in sociology and Steven Wilkinson (see Chandra and Wilkinson 2008) as well as Roeder (2007) in political science have embarked upon similar projects.

Our own efforts have led to a new dataset with information on 145 of today’s states from 1816 to the year when they achieved nation-statehood—virtually the entire world, with the exception of the mini-states in the Carribean, Pacific, and Europe, as well as 12 larger territories for which data on core variables are not available. The dataset includes autonomous states which never experienced a colonial interlude (such as Switzerland or Thailand), land-based empires and their dependent territories, and colonies. It takes the boundaries of states in 2001 as fixed units of observation for the entire period under consideration and traces developments in these territories over time, independent of which states governed them. It contains data relating to the levels of industrialization and adult literacy, the directness of rule, the ethno-demographic make-up, the types and numbers of wars fought, the political and economic power of the political center, the strength of the nationalist movement, as well the natural resources found in a territory. We like to believe that despite its limitations, which will be openly discussed in the data section below, it represents a considerable advance because it allows, for the first time, to assess the generalizability of major theories of the nation-state.

We use discrete-time event history analysis with the creation of a nation-state as the outcome of interest. Overall, our results demonstrate that (1) diffusion effects at the levels of neighborhoods and
Empires as well as (2) the relations of power between nationalist movements and the established regime are far more important to understand nation-state creation than the classical theories maintained. Their focus on domestic structural conditions such as industrialization, the rise of mass literacy, or the shift to direct rule, does not seem to be adequate to understand the global adoption of the nation-state form over the past two hundred years. Our analysis thus lends support to models of institutional diffusion (Kedourie 1960; Strang 1990; Wimmer and Min 2006) and political history approaches (Mann 1995; Roeder 2007), which both emphasize the contextual conditions that empower a nationalist project. In the concluding section, we develop a historical institutionalist model of nation-state creation in line with the empirical findings. It identifies the shift of political power from representatives of the old regime to nationalist forces as the crucial element that determines which institutional order will be established in a polity, and shows how this power shift is influenced by political factors of both proximate and distant nature.

2 Literature review and hypotheses

Many reviews of the literature on nationalism and the nation-state have been published in the past (Calhoun 1997; Ozkirimli 2000; Smith 1983; Smith 1998). Our aim here is to tease out of often sophisticated and complex arguments those aspects that can be condensed into the form of a testable hypothesis. We are well aware of the epistemological implications of this procedure and do not maintain that we can submit these theories and historical narratives in toto to a Popperian test. We do believe, however, that the hypotheses generated by our interpretation are derived from the major theoretical propositions in a fairly direct and unambiguous way.

Economic, cultural and political modernization

Ernest Gellner’s articles and books introduced a functionalist theory of nationalism and the nation-state (Gellner 1983). The causal trajectory starts with industrialization, which demands a mobile and flexible workforce to function properly. This in turn requires cultural and linguistic homogenization such that workers can shift from job to job and communicate effectively with each other at the workplace. Such homogenization is provided by the educational apparatus of modern states, and nationalism is the ideology which legitimizes states to pursue such a homogenization policy. Thus, the likelihood of nation-state creation should increase with industrialization (Hypothesis 1).

Benedict Anderson famously described the nation as an “imagined community” (Anderson 1991) that replaced the closely knit, “real” communities of guild, village, or town of the Middle Ages. The imagining of large-scale, anonymous national communities was the product of the reformation and the rise of print capitalism, which enabled and propelled literacy in vernacular languages, replacing complex elite languages such of Latin. Nationalists transformed this communicative space into a communitarian identity and endowed it with a sense of historical depth and destiny (ibid., chapter 2). While originally a matter of educated elites, nationalism was gradually adopted by an increasingly literate population, which eventually led, in a somewhat underspecified way, to the foundation of

---

2 Gellner’s theory also foresees a second pathway to nationalism as a reaction to the new national regime by those minorities who are not easily assimilated into the majority and who come to resent the political domination by majority bureaucrats. Gellner points out that religious minorities (he made allusions to the Jews in Eastern Europe) or racially defined groups (hinting at the case of African Americans) are the most difficult to dissolve into a national melting pot (ibid. chapter 6). Since our dataset includes only successful nation-state projects, we cannot test this line of argument in an adequate way.
modern nation-states. Thus, Hypothesis 2 states that an increase in the literacy rate in vernacular language makes nationalism and the nation-state more likely.

Anderson develops a second line of reasoning in order to account for the fact that in Latin America (and later in South East Asia and Africa), nationalism developed with regard to imperial or colonial provinces—rather than the domain of shared vernacular languages as in Europe (Anderson 1991, chapter 2). Lower level colonial administrators recruited from the local population, he argues, were confined to the horizon established by imperial provinces and thus shared common experiences of a bounded geographic space and built networks among themselves, thus laying the groundwork for the imagining of the nation along provincial boundaries and the eventual independent nation-statehood of these provinces. From the late 19th century onward, the provincially organized school system in Indonesia and French Indochina played a similar role in containing social networks and cognitive spaces to provincial domains (ibid.: chapter 7). We thus expect that state or sub-state units are more likely to develop nationalism and become nation-states than territories that are split between different provinces or states (Hypothesis 3).

Michael Hechter’s political modernization theory focuses on changes in the system of governance over time (Hechter 2000; see already Tilly 1990). He distinguishes between two different pathways: state-building nationalism as well as peripheral nationalism. Both are driven by the reform of techniques of governmental control and resource extraction. Indirect rule via regional elites and notables is replaced by direct rule through a unified and hierarchically integrated bureaucracy. In autonomous states (such as early modern France), the state elites gradually nationalized and homogenized the population and developed nationalism to legitimize their rule. In territories under the rule of ethnic others, such as Croatia under the Habsburgs, the shift to direct rule led to nationalist mobilization by regional elites who resented alien rule, and eventually to imperial collapse and the creation of nation-states. Since similar forces push history down along both pathways, we can state a single hypothesis: The more directly a territory is ruled, the more likely nationalist organization and the nation-state emerge (Hypothesis 4).

Balance of power arguments
A heterogeneous group of authors has focused on more contextual conditions that are largely absent from the classical modernization schemes discussed above. They describe nation-state creation as a form of political revolution or transition the success of which depends on the balance of power between representatives of the ancien régime and nationalists (Mann 1995; Wimmer 2002; Roeder 2007).

Various more specific hypotheses can be formulated. First, a Skocpolian argument (see Skocpol 1979) would maintain that nationalist revolutions are possible when the political center is weakened or has collapsed in the wake of wars, such as the Habsburg empire after the First World War. Stated in the form of a hypothesis, the likelihood of nation-state creation increases with the increase in the number of wars fought on a territory or within the empire to which it belongs (Hypothesis 5 and 6). Second, international relations scholars might argue that the global military and economic standing of pre-nationalist state elites should influence their capability to either co-opt or suppress nationalist movements and thus maintain the status quo and make nation-state creation unlikely (Hypothesis 7).

Third, the balance of power between nationalist challengers and the established order also depends on the relative strength of nationalist movements or elite factions. We suggest to take the time elapsed since the foundation of the first national(ist) organization as a proxy. The longer nationalists had time to propagate their worldview and to mobilize followers around the banner of national independence from the imperial yoke, of overcoming Kleinstaaterei in a unified national state, or of nationalizing the
state from above, the more likely they would be to successfully establish a nation-state (Hypothesis 8). Roeder has focused on the institutional resources at the disposition of nationalist contenders. Ethnic elites that are governing sub-state entities such as a province are more likely to create a nation-state in their own name, because they command the institutional and ideological resources necessary to challenge the center successfully (Roeder 2007). This argument leads to the same empirical hypothesis as the one derived from Anderson’s analysis of nationalism in the colonies: state or sub-state units would be more likely to develop nationalism and become nation-states than territories that are split between different provinces or states (Hypothesis 3 again).³

**Diffusion**

While all of the above approaches emphasize causal forces that are territory-specific, a series of authors have pointed to *diffusion mechanisms* as the main drivers of nation-state formation. This is again a rather heterogeneous group of approaches that locate diffusion mechanisms at different levels. John Meyer and co-authors have argued that the nation-state form is part of a world culture that had emerged over the past two hundred years. It consists of a set of unquestionable, taken for granted normative assumptions about the worth of individual freedom and equality, the legitimacy of rational arguments, the desirability of “efficient” organizational templates, and so forth. This world culture gradually forced political elites and followers to adopt nationalism as a template of political legitimacy and the nation-state as the only acceptable form of state-hood (Meyer et al. 1997). David Strang’s (1990) event history analysis of decolonization supported this view by showing that de-colonization accelerated after 1960 when imperialism was delegitimized in a statement by the UN general assembly. World culture theory contains not only a longitudinal argument—as in Strang’s analysis—but a cross-sectional one as well: The more linkages a territory maintains to the centers of global culture and power, the more likely it is to adopt world-cultural templates. For data reasons, we have to limit ourselves to the longitudinal dimension here and hypothesize that the likelihood that a territory is ruled as a nation-state is positively associated with the increase in the number of territories in the world that have already adopted the nation-state (Hypothesis 9).

Global diffusion is a well-researched topic thanks to John Meyer and associates. Fewer authors have written about the imperial arena or the neighborhood as the locus of diffusion processes. Benedict Anderson (1991, chapter 7) described how elites from all over French Indochina were drawn together in the newly founded schools where they were trained as civil servants to staff the administrations of the various provinces of Indochina. In these environments, they built networks with each other, came in touch with nationalism as a political doctrine and learned about the nation-state institutions of the center. While Anderson emphasizes how these educational institutions aligned the emerging national imagination along provincial lines, we are more interested in the diffusion processes that operate across territories within an empire. One example may suffice here. In 1810, the Cortes of Cadiz replaced the absolute monarchy and the Napoleonic troops had finally been defeated. Liberalism, French republicanism and modern nationalism became the major currents of thought also among the Latin American delegates to the Cortes. Natural law arguments in the scholastic tradition, referring to an original contract between conquerors and the Crown, still dominated the thinking for example of the Mexican anti-colonial rebellion unfolding during these years (Brading 1985:43f.). The Latin American delegates quickly learned to replace such arguments with a modern appeal to the right of national sovereignty. As this example illustrates, empires establish a communicative field within which peripheral political elites observe and imitate each other’s nation-building projects or mirror that of the

---

³ Ideally, of course, one would also code the degree of institutional and ideological power over which leaders of sub-state units command. No such data is available on a global scale, however.
imperial center (Fieldhouse 1966). In other words, the more territories within an empire have already become nation-states, the more likely other territories will follow (Hypothesis 10).

Finally, Wimmer (2002, chapter 3) elaborated how nationalism and the nation-state might diffuse between neighboring territories independent of whether or not they are governed by the same empire. The superior political and military mobilizational capacity of the nation-state entices neighboring political elites to adopt this political model as well in order to compete successfully with their neighbors—a domino effect of sorts. This mechanisms is more likely to operate in areas where territorial boundaries and ethnic settlement patterns are not coinciding, thus opening the danger of irredentist claims by nationalizing neighbors on one’s own territory (see also Weiner 1971; Brubaker 1996, chapter 3). A good example is the competitive rivalry between Greek, Bulgarian, Serbian and Hungarian nationalist in the 19th and early 20th century, who emulated and rivaled each other’s nation-state projects in order to establish control over territories with uncertain ethnic affiliation and identities that they hoped to snatch away from a failing Ottoman empire. Hypothesis 11 thus states that the likelihood of nation-state formation increases with the number of neighboring nation-states.

Existing datasets and findings of the quantitative literature
Quantitative tests of these various approaches, theories and hypotheses have so far been limited to two publications, both with a different focus and smaller universe of potential cases than that of the present study. David Strang’s (1990) work aimed to understand the conditions under which the colonial dependencies of Western powers became independent states from 1870 to 1987. He assembled a new dataset relying on various sources, most notably the Stateman’s Yearbook that collected the reports written by British ambassadors for Her Majesty. He found support for hegemonic cycles theories (decolonization is more likely in times of hegemony), world polity theory (the chances of independence increase after the anti-colonial declaration of the UN had been adopted in 1960), balance of power arguments (colonies governed by a metropole with strong naval capability are less likely to become independent), imperial diffusion effects, and so forth. We build on this endeavor by enlarging the empirical and temporal horizon to include the American Continent, the dependencies of non-Western, land based empires (of the Ottomans, Romanovs, Habsburgs and so forth) as well as the former Soviet Union and Yugoslavia. We also include the first fifty years after the Congress of Vienna that Strang had to exclude from consideration because the Stateman’s Yearbook did not exist in these years. Our dataset also improves on data quality by relying on a wide range of sources for historical data and by including a broader range of variables that are relevant for the nationalism and nation-state literature.

In the context of our own research, Strang’s findings in support of a world polity, global diffusion theory are the most relevant. His quantitative analysis reveals that the likelihood of de-colonization increases in the period after 1960, and he attributes this effect to a UN declaration that calls for the decolonization of the world. However, he does not provide any evidence that the 1960 turning point indeed fits the observed data best, nor would it be possible to show that it is this particular event that occurred in 1960 and not another (say the massive number of African states that achieved independence that year) that propelled the trend towards de-colonization. We conclude that identifying a time trend might not be the most convincing strategy to test global diffusion arguments, a problem to which we will return further below.

---

4 David Strang’s pioneering quantitative work has found empirical support for such imperial level diffusion processes for colonies (Strang 1990).
The scope of Philip Roeder’s study complements that of David Strang in focusing mostly on the former Soviet Union. But he also offers an analysis of a global dataset to test his institutional capacity argument, according to which the institutional autonomy of provincial elites allows them to establish cultural/ethnic hegemony within their terrain and provides them with the political resources to successfully challenge the metropole and establish an independent state. Roeder’s empirical analysis is based on a global dataset containing information on 336 sub-state units that existed between 1901 and 2000 (ibid.:323-331). Which one of these units, he asks, has eventually become an independent nation-state rather than survived as a province, or disappeared in the wake of a re-organization of the state? He shows that the likelihood of success increases if the sub-state unit is self-governing, if the central elites are weakened by internal strife and political turmoil, if they exclude the population of the provinces from political participation, if the provincial population is linguistically and religiously different from the core population, and if the province had experienced independent statehood prior to incorporation.

Roeder’s bargaining model advances our theoretical understanding of nation-state formation considerably by revealing the importance of the balance of power between nationalist contenders and representatives of the ancien régime. Many of his arguments parallel our own understanding of how the nation-state spread over the world. However, his quantitative work suffers from serious weaknesses that raise doubt about the validity of his specific version of the balance-of-power argument. Most importantly, there are sample selection problems: His list of provinces that enter the event history dataset contains no entries for many provinces that never became nation-states, such as German Bundesländer, US states, Swiss Kantons, etc., although these are obviously enjoying as much autonomy as the states of India that we find in his list. The Ottoman empire has only Bulgaria (after 1878) and Crete (after 1898) and Samos as sub-state units—and none of the Ottoman villayets (none of which developed into a nation-state), and so forth.

3 Dataset and modeling approach

In short, both existing quantitative studies of nation-state creation operate with a limited universe of cases that exclude large portions of the world that were candidates for making the transition to the nation-state. They are also limited by data-quality issues and by analytical problems that raise doubts about the reliability of results.

3.1 Units of observation

Our dataset contains information on 145 territories from 1816 until a nation-state is created. 139 of these territories have made the transition to the nation-state by 2001, while the others are still governed as absolutist monarchies. All territories are referring to the geographic boundaries of countries that existed in 2001. Our coverage is almost global, since we only miss mini-states with less than 20'000 km² surface plus eight larger states on which no data on literacy were available.5 We also exclude the early nation-state creations in Great Britain, France, Paraguay and Haiti because they occurred before our data series start in 1816. Our analysis is thus not about “the origins”—how the earliest nation-states were born in the seventeenth and eighteenth centuries (cf. Kroneberg and Wimmer In preparation)—but about the general mechanisms that might help explain its subsequent adoption in the rest of the world.

5 We could not find early literacy rates for Albania, Belize, Djibouti, Equatorial Guinea, Greenland, Iceland, Lesotho and Namibia. We did include, however, Gambia, Kuwait, Cyprus, Bahrain, Qatar and Mauritius, all of which control less than 20’000 km² of territory.
Some notes on our choice of units of observation may be in order. To clarify, our approach implies that we combine data from various polities in the case of territories that did not correspond to a political unit at a given point in time. Data for Poland in the 1870ies, to give an example, are proportionally combined from the Russian, German and Austrian empires that controlled pieces of what is today Poland.

Choosing today’s state system as an observational grid is less problematic than it seems. First, today’s grid of countries is smaller grained than at any time before, except for Germany, Italy, Yemen and Vietnam that used to be split into smaller countries that then fused. Choosing a grid with smallest possible units helps to avoid the misattribution problem that would arise if our data would be coded on larger units than the ones that eventually achieved nation-statehood. Second, an alternative approach would be to code changing political units of observation, i.e. either the states or the provinces that actually exist in any given year. These political units would then include shifting sets of populations over time, depending on how empires and states expand or contract, provincial boundaries are redrawn, and so forth. “Turkey” in 1816 and “Turkey” in 1925, to give an example, would then refer to quite dramatically different sets of territories and populations. It would also pose the problem that units would appear (Latvia in 1918) and disappear (Latvia in 1945), while the Latvian population and its national aspirations obviously remained. The problem of a shifting case set would only be exacerbated if provinces were taken as units of observation, as in Roeder’s study.

But doesn’t choosing today’s countries as constant units of observation imply that we select on the dependent variable since most of these units actually did become nation-states? Selecting units of observation depending on the outcome is not the same as selecting observations depending on the observed outcome—the mortal sin of quantitative analysis—which our procedure clearly does not imply because we choose existing states of 2001 as units, not existing nation-states. Still, one could argue that our units have an intrinsically higher event propensity than other constant units of observation because most of them are nation-states today. Imagine a division of the globe into constant units of observation using a 100 x 100 km² grid (using GIS technology). This would produce an even higher event probability, simply because many of the remaining non-nation-states are very small (Buthan, Brunei, Qatar), while some existing nation-states are very big (the Soviet-Union, Canada, the USA) and would thus be broken down into a very large number of events. Obviously enough, treating these events as independent from each other is highly problematic, a problem that our research design largely avoids.

3.2 Variables

Dependent variable
We define nation-state creation as the year in which an internationally recognized independent state adopted a constitution that established government in the name of a nationally defined people of equal citizens (all data were adopted from Wimmer and Min 2006). We choose the year in which all of these various definitional elements were present: sometimes independence was achieved before a constitution was established, sometimes the other way round. Note that our definition focuses on the principles of political legitimacy on which a state rests. We are not asking in how far these principles have been realized. Thus, we code autocracies as nation-states as long as the dictator pretends to rule “in the name of the people”. The extent to which citizenship rights have been granted to all citizens is
also not of interest to us here—disregarding the fact the property restrictions on voting rights and the exclusion of women from full citizenship rights usually lasted many decades into nation-statehood.6

We coded only successful nation-states, i.e. states that survived for more than three years and achieved international recognition by at least two states, thus excluding short lived states such as the West Ukrainian National Republic, the Kurdish republic of Mahabad in Iran or the Republic of Uyghurstan in China. This is a reasonable selection principle, since we would otherwise face the difficult problem of dealing with states that were declared independent by politically marginal movements and that never achieved de-facto control over any territory.

Following this coding principle, several territories did experience more than one episode of nation-state formation. The Baltic territories, for example, were independent national states from 1918 to the Second War, when they were swallowed again by the Soviet Union, and regained nation-state status again in 1991. Other, more problematic cases are those in which a territory achieved nation-state status as part of a larger state, which subsequently broke apart into smaller nation-states. Such was the case of Gran Colombia (giving rise to Colombia, Venezuela, Ecuador and eventually Panama), the Central American Republic (fragmenting into Guatemala, Honduras, Costa Rica and El Salvador), as well as Yugoslavia and Czechoslovakia. In 24 such cases, we code a second event of nation-state creation and treat these as independent from the previous ones. To check for robustness of our results, we ran all our models with a restricted definition of the outcome that excludes these repeated events.

Independent variables
To test the economic modernization argument, we coded the length of railway tracks (in km) per 1000 square kilometer. Some of these data were adapted from the monumental historical compendium of economic data assembled by Mitchell Various years). Other data were coded from primary sources—which are amazingly rich thanks to the enthusiasm that the history of railways has sparked among lay and professional scholars.7 We usually had to track down each railway line and the year it was opened to the public, and then break down the total track length by today’s territories if they crossed them. This variable shows a reasonable high correlation of r=0.45 (or 0.65 for years before 1970) with the percentage of the workforce employed outside of agriculture, which would offer a more direct measurement of the industrialization hypothesis if data coverage was better. But doesn’t the automobile, which increasingly replaced railcars, make the length of railway tracks a poor proxy variable for industrialization? Those industrialized countries that ceased to build railways after the First or Second World War often achieved nation-statehood before they started to rely on the automobile. And while some Communist dependencies or colonies led their railways systems decline (as in Latvia or Mauritius), many more continued to build railways after the Second World War and well into the seventies and beyond.8 We conclude that the length of railway tracks represents a reasonably adequate proxy for industrialization.

6 We do not define states that in their constitutions exclude segments of its population from the citizenry as modern nation-states, however. Thus, the Herrenvolk democracies of Rhodesia and Apartheid South Africa, the slave-holding United States before the civil war, and Australia before granting citizenship to Aborigines are not considered modern nation-states. To test for the robustness of our results, we also ran all our models on a dependent variable that coded these cases differently, i.e. taking years of independence as the creation of a nation-state. The results remained unaffected.

7 For the Habsburg empire, we relied on Strach 1906; Oberegger 2008; Heinersdoff 1975; for the successor states of Yugoslavia on Oberegger 2008; for the Romanov empire on Perl 1872, Roll 1915, and Rautavuori 2008; for Germany on Roll 1915; for the former Soviet Union republics, on Central Statistical Administration 1957 and Sakari and Likka 2003; for the Ottoman empire, on Bonne 1998; Karkar 1972.

8 This is the case, for example, in Zimbabwe, Mozambique, South Africa, Angola, Bangladesh, Malaysia, Armenia, Slovakia, Croatia, Uzbekistan, Tadjikistan, and so forth.
To test Anderson’s main argument about the role of mass literacy in generating nationalist imaginaries, we took advantage of the considerable amount of research on the history of mass literacy that has emerged in the wake of Cipolla’s ground-breaking work (Cipolla 1969). We assembled data on adult literacy rates for all our territories, relying on several dozens published country-studies, government censuses, historical research on particular regions, existing quantitative datasets, and so on (the sources are listed in Appendix A). We found estimations within a 10 years range from the year of nation-state creation in all cases. We used the best estimates available for the beginning of the time series in 1816 and usually several later datapoints to interpolate, a justifiable procedure given the slow pace of literacy rate development (and the rarity of backsliding). Despite our best efforts, problems of reliability and comparability remain: Figures up to the 1870ies (when census taking becomes widespread) define literacy as the ability to sign fluently with one’s name, while later figures refer to the ability to write a letter or a couple of sentences (functional literacy; on the reliability of literacy estimates based on signatures, see Reis 2005, fn. 9-12); data for pre-colonial East and South-East Asia, which had high levels of literacy in vernacular languages before colonization, are hard to estimate (cf. Reid 1990); colonial census often comprise only the population under control and within the reach of the embryonic colonial administration; and so forth. Despite these difficulties and in view of the very substantial variation of literacy data over time and across territories, we believe that data-quality is sufficient to justify the inclusion of this variable in our models.

The degree of directness of rule that plays a crucial role in Hechter’s theory of nationalism and the nation-state can be approximated by calculating central government expenditure (of both an autonomous state or an imperial center) for a particular territory, assuming that the more a government spends, the denser and deeper administrative penetration. This represents an adequate proxy for other state-focused political modernization arguments as well (e.g. Tilly 1990; Breuilly 1993; Mann 1995; Wimmer 2002). We again used data from Mitchell and complemented this with additional sources for the Ottoman and Spanish empires as well as the Soviet Union. Territories that were split between various states were given a composite figure based on values proportional to the territorial share of each state. Imperial dependencies (but not colonies) were given the same values throughout an empire, assuming that land-based empires were more uniform in modes of territorial control than sea-born colonies. Thus, Bosnia and Iraq under the Ottomans received the same values on the literacy variable, while Kenya and Malaysia under British rule were each given a different figure. Pre-colonial territories that were not governed by a centralized state were assigned a value of 0 (because, by definition, they cannot be ruled directly), all other polities (including pre-colonial and pre-modern autonomous states such as Dahomey, Burma, or Morocco) were given their proper values or coded as missing values.\(^9\)

\(^9\) In a handful of cases where no data are available for these early years, we chose the data for the most similar society available. These cases are Vietnam and Korea (approached through Chinese early 19th century data), Laos and Cambodia (which were given Thailand’s early figures), Morocco, Libya, Algeria and Malaysia (which were given Tunisian data from the middle of the 19th century), the Cisleithanian provinces of Austria (which had comparable rates to Prussian Westphalia). For a couple of British colonies in Africa, we assumed the same growth rate over the 20th century as in Ghana, the best researched case. All pre-colonial territories in societies that had no indigenous writing culture were assigned the value of 0.

\(^10\) Our prime sources for calculating figures for the Spanish empire were Klein 1998; for the Ottoman empire Shaw (1978) and Akar (1999) for government expenditure, Cem Behar (1996) and Karpat (1985) for population data; for the Soviet Union Plotnikov (1948/1954) and Svodnii otdel gosudarstvennogo byudzheta (Various years) and for population data Side (1992) and Kozlov (1988). The Soviet budgets included basically the entire economy, given the nature of the Soviet economy (on Soviet budgets, see Hutchings 1983). To make these figures comparable to others, we excluded all expenses related to production and distribution of goods from the provincial budgets, including those related to pensions, health care, and the like.

\(^11\) Nominaly dependent, but de-facto self-rulled territories under the Ottomans were also assigned their proper values or missing if not available.
All figures were standardized as per capita figures, converted into US constant dollars using time-varying conversion rates (XY), and then additionally adjusted for purchasing power differences using Angus Maddison’s GDP estimations (Maddison 2003). This variable shows a very high correlation with the measurement of direct rule developed by Lange 2005) for British colonies (r=0.82 for 19 data points, using non-standardized values). In addition, our measurement of government expenditure seems to capture historical shifts in the mode of government quite adequately (such as for example the decentralization of the Hapsburg monarch after the Ausgleich in 1867).

All our diffusion variables were generated by counting the number of territories governed as nation-states within the neighborhood, within the same imperial domain, or in the world. Territories that were shared between various empires or states were given proportional values based on respective territorial shares. We coded these variables also as percentage as well as absolute numbers to take into account that territories that have many neighbors or pertain to large empires have more opportunities for “contagion” than isolated islands or imperial polities with less than a handful of adjacent territories. Another coding recorded the number of nation-states established during the previous five years in order to capture the dynamic nature of diffusion processes in more adequate ways. This is the version of the variable that we use in subsequent analysis, but our results also hold using the two other codings.

To evaluate whether wars in the territory or the empire affect the creation of nation-states, as a Skocpolian argument would have it, we used the Wimmer/Min dataset of wars in all territories of the world from 1816 to 2001 (Wimmer and Min 2006), with additional codings for the New World from 1789 onward. This dataset allows distinguishing between inter-state wars, civil wars, and nationalist wars of independence and we coded a large number of different war variables to test whether specific types of wars are more effective in weakening the political center and thus shift the balance of power in favor of nationalist revolutionaries.

To test Roeder’s theory of institutional capacity and Anderson’s provincial horizon of identity argument, we coded for each year whether a territory was conforming to an autonomous state or a sub-state unit (1) or not (0). We counted provinces, colonies, mandate territories, vilayats and sanjaks, Russian governorates etc. as relevant sub-state units. Various re-organizations of colonial and imperial provinces had to be taken into account. Pre-colonial territories that were divided between various indigenous states or that had no centralized polity at all were coded as 0.

We also used different codings of the direct rule variable to accommodate different possible interpretations. In one variant, we assigned all pre-colonial, pre-modern states 0 in order to exclude non-bureaucratic regimes from our definition of direct rule, thus effectively interpreting it as “modern, bureaucratic forms of direct rule”. Another variant defined de-facto autonomous territories nominally controlled by an empire (mostly Algeria, Bahrain, Egypt, Kuwait, Libya, Tunisia, and the UAE under Ottoman rule) as 0, while assigning pre-modern or pre-colonial states their values or coding them as missing. This represents a coding of direct rule that does not distinguish between modern and pre-modern forms of rule.

We smoothed values when they jumped up during major wars, in order to exclude the temporary war expenses.

Obviously, we coded all territories in the neighborhood, empire or world, including those on which we lack data on other independent variables and which are thus not included in the rest of the analysis. For the imperial diffusion variables, we extended membership in the empire five years after independence. To calculate the neighborhood variables, we used the matrix of contiguity provided by the Correlates of War project, using a maximum distance of 150 miles to define neighbors separated by water (COW 2008).

Territories that were split between empires receive composite values according to the territorial share of the empires.
The strength of nationalist challengers is proxied by the years elapsed since the foundation of the first national organization. To count as a first national organization, it needs to be organized along modern principles, i.e. define membership formally (thus excluding clientelist networks or informal factions) and institutionalize leadership roles independently of individuals (thus excluding the personal followings of political leaders). In addition, the organization must claim to represent, speak in the name of, or advance the interests of the national community in the name of which the territory eventually became governed. Note that our definition is relatively broad, including organizations that are not nationalist in the strict sense of the term, i.e. are not exclusively devoted to bringing about national independence for a particular territory. They all do establish, however, an “imagined community” that comprises the ethnic core population of a territory and thus introduce nationalism into the political arena.

To test the political center’s capacity to resist nationalist movements into account, we relied on the Correlates of War Project’s Composite Index of National Capabilities, which combines energy consumption, military expenditure, the number of soldiers, steel production, urbanization, and population size into an often used index (Singer 1987). This is a rough, but reasonably adequate proxy for a central elite’s overall economic and military power that they can use to co-opt or repress nationalist movements and maintain a pre-national status quo. All dependent territories were assigned the value of the imperial or colonial center, pre-colonial, state-less territories were coded as 0, while those autonomous states not listed in the COW dataset were coded as missing values.

Correlation matrix and summary statistics are provided in Table 1 below.

| Table 1: Means, Standard Deviations, and Correlations among Variables Affecting Nation-State Creation |
|---------------------------------------------------|---|---|---|---|---|---|---|---|---|---|---|
| R: Length of railway tracks per km² | L | Y | E | S₁ | S₂ | NS₁ | NS₂ | NS₃ | Wₑ | Wₜ |
| R: Length of railway tracks per km² | .515 | .553 | .023 | .346 | .272 | .280 | .001 | .061 | .086 | .060 |
| L: Percent literates among adults | .710 | .281 | .566 | .590 | .598 | .016 | .089 | .269 | .022 | .002 |
| Y: Years since first national organization founded | .205 | .601 | .662 | .650 | .073 | .117 | .135 | .012 | .070 |
| E: Central government expenditure for territory | .402 | .476 | .481 | .084 | .054 | - | .011 | .044 |
| S₁: 1st Cubic spline on year | .911 | .911 | .254 | .160 | .055 | .133 | .049 |
| S₂: 2nd Cubic spline on year | .958 | .244 | .139 | .030 | .063 | .049 |
| NS₁: Total No of nations-states in world | .200 | .138 | .047 | .002 | .031 |
| NS₁: No of NS created in the empire past 5 years | .183 | - | .193 | .085 |
| NS₃: No of NS created among neighbors past 5 years | .030 | - | .138 |
| P: Center’s share of global power | .017 | .164 |
| Wₑ: Number of wars fought in the empire | .244 | .028 |
| Wₜ: Number of wars fought in the territory | .164 |

Mean | 5.6 | 20.2 | 11.7 | .448 | 188 | 21.7 | 40.0 | .258 | .084 | .074 | .535 | .056 |
Standard deviation | 16 | 27 | 24 | 1 | 46 | 30 | 26 | 1 | .352 | .080 | 1 | .251 |
N | 174 | 176 | 176 | 983 | 176 | 176 | 175 | 176 | 176 | 164 | 175 | 1751 |

16 Most of the information is based on Woronoff (Various Years). Additional sources were consulted for many territories.
3.3 Modeling approach and time specification

This dataset was set up as an unbalanced panel with territories as units. All independent variables were lagged one year to avoid endogeneity problems. We use discrete-time event history models, estimated via a logistic regression analysis of territory-years, to estimate the effect of covariates on the likelihood of nation-state creation. Following standard practice, we cluster standard errors on territories to account for the non-independence of observations within territories. To control for unobserved heterogeneity across territories, we also ran models with territory fixed effects, allowing the constant to vary independently for each territory.

Cross-national quantitative research is haunted by the instability of results caused by robustness problems (the combination of variables affects their individual effects), the problem that parameters might have different effects in different sub-samples, and unobserved heterogeneity exacerbated by data problems (see most recently Young 2009). In view of these difficulties, our strategy was to find theoretically and historically/substantially meaningful models that are at the same time robust to a very wide range of model specification, parsimonious (using the Bayesian Information Criterion as a guideline), do not depend on only few observations (which can be discovered through post-estimation diagnostic techniques), hold across time and continents (using these as control variables as well as running models with subsamples).

One of the major challenges in event history analysis is how to conceptualize the effects of the passage of chronological time. The most primitive models assume that the baseline hazard rate, i.e. the “risk” of the event occurring, remains constant over time. This is not a realistic assumption given that the contextual conditions of 1816 are hardly the same as those in 2001 and that our independent variables capture only a portion of these differences. How can one capture the changing baseline hazard rate of nation-state creation over time?

Partisans of the global diffusion approach might suggest that the deepening and broadening reach of world culture itself describes the change in the baseline event risk appropriately. To see whether global diffusion variables are indeed better at predicting nation-state creation than simple linear time, we compared the fit of regression models that include only a global diffusion variable but no baseline time trend to the fit of alternative models that include only a secular time-trend but no global diffusion variable. The diffusion variables never fitted the data as well as the simple chronological time (not shown). In substantial terms, this means that the advancement of chronological time captures the fact that the likelihood of nation-state creation increased better than the global diffusion term. This does not mean the end for the global diffusion argument, to be sure, because a cross-sectional argument could still be defended: territories with more intense connections to the global centers might well be more likely to become nation-states. We unfortunately lack the data to test this hypothesis. In the following model that tests diffusion arguments, we include a global diffusion variable although it is highly collinear with chronological time terms and thus produce negative significant coefficients that represent a statistical artifact and are substantially meaningless.

The above analysis led to the conclusion that we cannot substitute global diffusion variable for a baseline hazard trend. To inductively find the best fitting time model, we experimented with natural cubic splines, higher order polynomials on a linear time trend, period effects, discrete time models

---

17 Another procedure was to regress the predicted hazard rates from a discrete-time model (using decades) first on the global diffusion variable and then on linear time, using OLS regression. The results showed that the secular time trend has a stronger association with the discrete-time hazard rates.
using 10 to 20 year periods, and so forth. We found that the natural cubic splines with three knots fits the data very well and represents the additional advantage of flexibly adapting themselves to different model specifications. We thus include them in all the models that follow. Substantively, these cubic splines describe a constant hazard rate until the Second World War and a sharply increasing baseline risk thereafter. This increase in the baseline risk might have many causes. It may be due to the post-war rise to global hegemony of the United States, champion of de-colonization and the self-determination of peoples. Or it may relate to the unprecedented growth of the global economy that made many more nation-state projects feasible. Or it may capture decreasing popularity of the colonial project in France and Britain after, and as a consequence of, the Second World War. Note that none of the effects that we report in the following depend on the specification of the baseline hazard rate with cubic splines.

4 Results

The correlation matrix in table 1 shows that many of the developmental variables associated with major theories of nation-state formation are highly collinear (especially years since foundation of first national organization, industrialization, literacy, and global diffusion). We therefore introduce them in separate models to test their explanatory power independent from each other. Otherwise, we group variables that are related to the same theoretical approach into the same models. We maintain variable in subsequent models if results are significant and substantially plausible until we arrive at what we consider the best fitting explanation of nation-state creation. We include Continental dummies in all our main models to account for possible specificities of world regions. The results of this procedure are displayed in table 2.

4.1 Explaining nation-state creation

Model 1 includes the length of railway track as our proxy to test Gellner’s industrialization hypothesis. Even without any other covariates in the model, the railways variable fails to achieve standard levels of significance. We experimented with a very wide set of models, using different time specifications, excluding the second episodes of nation-states creations in Latin America and Eastern Europe from consideration, using various interaction effects and subsample analysis. The result is that railways achieve significance only in certain model specifications and thus cannot be considered robust. The main reason, according to these various analyses, is that the early nation-states in Latin America were created in a largely pre-industrial environment and that many weakly industrialized African states did achieve nation-statehood in the 1960ties, while the population of many of the highly industrialized Soviet and Yugoslav provinces had to wait for another generation to accomplish the same. This finding already points to political factors that are absent from Gellner’s industrialization account: the power of the Soviet Union to re-conquer the Baltic and Caucasian states and to keep nationalist movements

The railway variable achieves significance in two combinations of variables: First, in models that do not include general time trend and at the same time control for the specificities of the Latin American, African and Soviet territories either through Continental controls or a fixed effect model (with such a time trend, the variable is also significant in fixed effect models but the coefficient is negative). Second, the effects of Latin American and Soviet territories is weakened by eliminating the second episodes of nation-state creation (on the territories of Gran Columbia, the Centeral American Republic and Yugoslavia as well as in the Caucasian and Baltic regions). If the first episodes in South America and Yugoslavia are not considered, i.e. when only the establishment of Colombia, Venezuela, Croatia etc. are taking into account, the variable again looses significance. Subsample analysis shows that railways are significant and positive for Eastern Europe only (a result driven by the early nation-states created in highly industrialized Cisleithania), as long as we control for previous episodes of nation-state creation with a dummy variable, and the Middle East, but not for other regions (and negative significant for Africa alone). Change in the length of railway tracks over time also has no significant effect on the likelihood of nation-state creation.
elsewhere in check for generations; the failure of a Spain pre-occupied with the Napoleonic invasion and a subsequent civil war to achieve the same in its far away colonial empire.

### Table 2. Parameter Estimates for Logistic Regression Models of Nation-state Creation (robust standard errors in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of railway tracks per km²</td>
<td>.007 (.004)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Territory corresponds to state or province</td>
<td>—</td>
<td>-.34</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Percent literates among adults</td>
<td>—</td>
<td>-.006</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Central government expenditure for territory</td>
<td>—</td>
<td>—</td>
<td>-.210* (.083)</td>
<td>-.254* (.094)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Dependent territory</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.002 (.458)</td>
<td>—</td>
<td>.393 (.277)</td>
</tr>
<tr>
<td>Gov. expenditure x dependent territory</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Total No of nations-states in world</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-.031** (.007)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>No of NS created in the empire past 5 years</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.184** (.031)</td>
<td>.153** (.033)</td>
<td>.142** (.032)</td>
</tr>
<tr>
<td>No of NS created in neighborhood past 5 yrs</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.707** (.122)</td>
<td>.546** (.13)</td>
<td>.552** (.13)</td>
</tr>
<tr>
<td>Years since foundation of first nat. organization</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.016** (.003)</td>
<td>.016** (.003)</td>
</tr>
<tr>
<td>Number of wars fought in the empire</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.348** (.047)</td>
<td>.354** (.047)</td>
</tr>
<tr>
<td>Number of wars fought in the territory</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.689** (.164)</td>
<td>.576** (.174)</td>
</tr>
<tr>
<td>Center’s share of global power</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-3.879* (1.379)</td>
<td>7.273* (3.171)</td>
</tr>
<tr>
<td>Share of global power x dependent territory</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-12.256* (3.119)</td>
</tr>
<tr>
<td>Middle East</td>
<td>-2.497** (-.521)</td>
<td>-2.803** (.642)</td>
<td>-2.401** (.52)</td>
<td>-2.301** (.498)</td>
<td>-2.470** (.591)</td>
<td>-2.283** (.509)</td>
<td>-2.116** (.501)</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>-2.022** (.354)</td>
<td>-1.892** (.423)</td>
<td>-2.580** (.445)</td>
<td>-2.714** (.562)</td>
<td>-1.587** (.474)</td>
<td>-2.001** (.372)</td>
<td>-1.974** (.39)</td>
</tr>
<tr>
<td>Africa</td>
<td>-1.495** (.336)</td>
<td>-1.852* (.577)</td>
<td>-1.843** (.354)</td>
<td>-1.873** (.386)</td>
<td>-1.688** (.471)</td>
<td>-1.490** (.36)</td>
<td>-1.413** (.382)</td>
</tr>
<tr>
<td>Asia</td>
<td>-1.372** (.354)</td>
<td>-1.658* (.578)</td>
<td>-1.846** (.367)</td>
<td>-1.847* (.409)</td>
<td>-1.283* (.466)</td>
<td>-0.940* (.463)</td>
<td>-1.865* (.373)</td>
</tr>
<tr>
<td>Oceania</td>
<td>-.623 (.392)</td>
<td>-.448 (.426)</td>
<td>-.98 (.519)</td>
<td>-.946* (.481)</td>
<td>-.369 (.5)</td>
<td>.019 (.428)</td>
<td>.28 (.453)</td>
</tr>
<tr>
<td>Latin America</td>
<td>.596 (.521)</td>
<td>.504 (.607)</td>
<td>-1.256 (.753)</td>
<td>-1.428 (.784)</td>
<td>.29 (.586)</td>
<td>.387 (.413)</td>
<td>.418 (.452)</td>
</tr>
<tr>
<td>1st Cubic spline on year</td>
<td>-.009 (.005)</td>
<td>-.007 (.005)</td>
<td>.021 (.011)</td>
<td>.019 (.014)</td>
<td>-.009 (.005)</td>
<td>-.014** (.005)</td>
<td>-.014** (.005)</td>
</tr>
<tr>
<td>2nd Cubic spline on year</td>
<td>.040** (.006)</td>
<td>.043** (.006)</td>
<td>.017 (.011)</td>
<td>.02 (.014)</td>
<td>.067** (.009)</td>
<td>.042** (.006)</td>
<td>.043** (.006)</td>
</tr>
<tr>
<td>Observations (n)</td>
<td>17497</td>
<td>17519</td>
<td>9839</td>
<td>9839</td>
<td>17519</td>
<td>16485</td>
<td>16485</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>260(6)**</td>
<td>264(10)**</td>
<td>184(9)**</td>
<td>186(11)**</td>
<td>345(11)**</td>
<td>419(14)**</td>
<td>422(14)**</td>
</tr>
</tbody>
</table>

*Significance level: .05
**Significance level: .01
Model 2 contains variables associated with Anderson’s and Roeder’s approaches. Territories that correspond to the boundaries of provinces or states are not more likely to transition into nation-statehood, raising doubts about the adequacy of Roeder’s institutional capacity and Anderson’s provincial horizon of identity arguments. The literacy variable is also insignificant. Analysis of an interaction effect with time (results not shown) demonstrates that increasing mass literacy is associated with a higher probability of nation-state creation up until 1920, but the effect disappears later on and turns even negative further down the road. We assume that this is because after World War I, literacy was promoted heavily by the Communist regimes all over Eastern Europe, but also by other non-national states and the colonial empires especially in Africa (on the global rise of school enrollment, see Benavot and Riddle 1988). Those regimes doing a better job at educating the masses were also more entrenched, so one could hypothesize, and better at keeping nationalists at arms length, not the least by indoctrinating the population with other, non-nationalist ideologies. In other words, the strength of nationalist movements, who came to power eventually in the African colonies and the Communist empires as well, is not systematically linked to or dependent on mass literacy. We again conclude that political variables related to the relative strength of nationalist challengers and established elites must be crucial in order to understand the rise and global spread of the nation-state.

Model 3 integrates our measurement of directness of rule, which reduces the number of observations by almost half. The variable is robustly significant, but with a negative coefficient—indicating that the more directly a territory is ruled by the political center, the less likely it is to eventually become a nation-state. In order to disentangle the effects of direct rule on autonomous states (the state-led nationalism path according to Hechter) and on imperial/colonial dependencies (the reactive nationalism path), in model 4 we introduced an interaction term between government expenditure and the status of political dependency. It shows that autonomous states are less likely to undergo the transition to the nation-state the more directly they are ruled (the opposite of Hechter’s state-led nationalism hypothesis), while there is no significant effect for direct rule on dependent territories, though the sign of the sum of the coefficients is positive. The same results are achieved through a sub-sample analysis of only dependent and only autonomous territories. It thus seems that contrary to Hechter’s

19 This result is not dependent on the number of years that we lag this variable.
20 This interaction effect remains significant even when second episodes of nation-state formation are excluded. Subsample analysis confirms that in the period up to 1920, literacy had a positive effect, mostly due to the highly literate Cisleithanian cases that became independent in 1918. In models with additional covariates, literacy is insignificant if we include only the years up to 1914. The literacy variable is also significant in models without Continents as controls and with a linear time trend as a specification of the baseline hazard rate. It is insignificant for all Continental subsamples except Asia. Change in literacy rates has no significant effect.
21 Models 3 and 4 remain almost identical for different codings of direct rule: As modern, bureaucratic rule (assigning 0 to all pre-modern states such as Dahomey or Ethiopia); as pre-modern as well as modern direct rule (coding missing values for pre-modern states where no data are available, and 0 for de-facto self-ruled, but nominally dependent territories such as Tunisia under the Ottomans). Hechter’s hypothesis can only be confirmed in a coding that could be described as “direct alien rule”, in which all pre-colonial territories, nominally dependent but de-facto autonomous territories, and all autonomous states are coded as selfruled, thus emphasizing the contrast between pre-colonial lands (which never experience a nation-state creation) and imperial/colonial dependencies, which all eventually transform into nation-states. Since it is more than doubtful that pre-colonial territories were mostly ethnically self-ruled (as the examples of the Zulu, Durrani, and Bemba empires illustrate), we think that this coding of the independent variable is not the most plausible one.
expectations, strong autonomous states that rule more directly over their territories are better capable of resisting the pressure to shift to the nation-state model and to keep nationalists in check—thus once again underlining the importance of the balance of power between ancien régime and nationalist contenders.

In order to be able to analyze the full dataset again, we exclude the government expenditure variables from the subsequent models and note here that their effects remain constant even if additional variables are added to the equation. Model 5 introduces diffusion variables. The term for global diffusion is negative and significant, due to colinearity with the time specification. However, we get strong and meaningful results for diffusion at the imperial and neighborhood level. The number of nation-states that have been founded during the past five years within an empire and within the neighborhood of a territory increase the likelihood of nation-state creation substantially, providing support for the imitation and domino mechanisms identified in the literature review section. Examples of positive cases that underlie these results are the increasing pressure on the remaining royalists in Bolivia, after the Bolivarian spirit of nationalist revolution had gained a foothold in most neighboring territories; the impetus to create a modern nation-state and to abandon the Hamidian search for imperial restoration that the Committee for Union and Progress felt when one Ottoman province in Rumelia after the other became independent states; the powerful demonstration effect that Indian independence had on many nationalist movement in the British empire, and so forth. We should note, however, that the significance of the imperial diffusion effect depends on the African British imperial cases of the 1960ies (more precisely, the Lesotho, Botwsanan, Ugandan, Kenyan, Zambian, Mauritian and Malawian nation-state creations that were inspired by Ghana’s earlier and successful struggle for independence). This is in fact the only such a strong influence by few cases that our post-estimation diagnostic tests have revealed.

Model 5 contains the variables that are meant to measure the various arguments relating to the balance of power between nationalists and pre-national elites. Our proxy for the strength of the nationalist movement—the years that have passed since the foundation of the first national organization—has an unequivocally strong effect on the likelihood of nation-state creation. We also hypothesized that nationalists will be more successful if the center is weakened by wars. Indeed, both the number of wars fought within a territory during the past year as well as the number of wars fought within an empire (excluding those fought on a territory) significantly affect the likelihood of nation-state creation. Examples for the latter include the civil war in the United States, which catapulted the country into modern nation-statehood according to our definition, the civil war in Bolivia between royalists and Bolivarists, the nationalist wars of liberation that helped to bring about the independence of the Baltic republics, or the Russian Turkish war that allowed Bulgaria to become an independent nation-state in 1879 after a Bulgarian rebellion had been crushed by Ottoman forces.

All these different versions of the independent variable of interest were also coded without standardizing its values according to the level of development of a territory. All results remain identical, except that the “direct alien rule” variable (see above) is completely insignificant in its unstandardized version.

W 22 We coded global diffusion with the total number of nation-states in the world. If we code a more dynamic variable and look at the number of nation-states created in the past five years (thus emphasizing demonstration effects that dissipate over time), the results are identical. These variables are only significant in models that include a linear time specification that does not capture the post-1945 upward trend.

W 23 We choose this coding of diffusion variables because they are reflecting the postulated mechanisms most truthfully: Rather than percentages or the absolute number of nation-states in the empire or neighborhood, imitation and domino effects are best captured by the dynamic coding of how many nation-states were created in the past five years. These two variables also result in the highest BIC value and reduce colinearity with other variables and with each other considerably. All other codings of the imperial and neighborhood diffusion variables produce similar results.
four years earlier. Among the many examples for the role of wars in other parts of the empire that weakened the imperial center enough to propel the nationalist project forwards are the First World War which debilitated the Habsburg and Soviet empire such that a wave of nation-state creations followed suit. Other cases are the Mau-Mau rebellion in Kenya and the Malaysian anti-colonial Communist insurgency, which decreased the willingness of the British empire to hold on to its imperial possessions and helped to accelerate Ghana’s independence—the first on the Continent. Similarly, the bloody struggles in Algeria and French Cameroon weakened the capacity and willingness of France to put further obstacles against the de-colonization in its West-African domains in 1960.24

The power of the center to resist nationalist challengers also depends on its international standing. The “center’s global power” variable is significantly and negatively related to the chances of a territory to achieve nation-statehood. However, the effect is different for self-rulled territories compared to imperial or colonial dependencies. An interaction term is significant and negative, while the sign of the coefficient of the non-interacted term is positive and significant, though this latter result is much less robust and becomes much weaker, for example, with a different time specification or without the variable that measure the years since the foundation of a first national organization. We can say with certainty, however, that the higher share of global power allows a government to co-opt, control or suppress nationalist movements successfully and prevent the establishment of nation-states on its dependent territories. In the robustness checks to be reported about below, we only include this interaction term for dependent territories.

How can the results of model 7 be translated into everyday language? Perhaps the best strategy is to calculate how increasing the value of each variable over its mean affects the likelihood of nation-state creation while all other covariates are set to their means. The basic probability of a nation-state being created on a territory when all variables are set to the averages is low—at 0.52%. This probability increases by 16% with every nation-state created in the empire over the past 5 years. Every nation-state established in the neighborhood over the past 5 years increases the probability by 71%. With every decade that passes since the foundation of first national organization, the probability of nation-state formation increases by 17%. With every 10% of capacity that the imperial center gains over the global average the probability of nation-state formation decreases by 35%. With an additional war in the empire, the probability of nation-state formation increases by 42%, and every war in the territory increases the probability 92%. Clearly, the variables in our model not only are statistically significant across a wide range of model specifications, but also represent substantially important effects.

Our findings thus point towards diffusion within empires and among neighbors as well as a balance of power favoring nationalist challengers: well established nationalist organizations, a political center weakened by wars, an empire lacking the global military and economic power or an autonomous state lacking governmental capacity to co-opt, control or suppress nationalism. We found no robust evidence for the effects of industrialization, the advent of mass literacy, and increasing directness of rule. Our

---

24 The results are robust to using the average number of wars fought over the past 5 years. We experimented with a large number of war variables and present in model 5 those that have the most explanatory power. They clearly support a Skocpolian interpretation: Empires that are weakened by a high number of wars in their domains are less able to suppress nationalist movements elsewhere, and the same holds true for representatives of the ancien régime within a territory. We can distinguish these balance of power effects from potential imitation and diffusion effects related to nationalism: Nationalist wars of independence within an empire are significant on their own, but so are non-nationalist wars (results not shown); the same holds true for wars in the territory itself. In both cases, a combined variable offers the best model fit by far (results not shown). In addition, none of the variables that record wars among the neighbors of a territory are significantly related to the creation of a nation-state. Thus, we conclude that wars don’t operate through an imitation effect, but they entail a power shift in favor of nationalists that allows them to gain control of government and enact the reforms that will establish an independent nation-state.
findings thus do not support established theories of the nation-state, which link its emergence to these forces of economic, cultural or political modernization. We also found no evidence that global diffusion mechanisms are at work in the creation of nation-states.

Before we discuss the robustness of the main findings and in how far they can be generalized across time, regions, and imperial domains, we briefly mention a series of other negative findings. First, we tested global hegemonic cycles theories by including dummies for periods of hegemony (following Wallerstein 1983) or global leadership (following the definition of Modelski and Thompson 1988). We found no significant effects with almost all time specifications, contradicting the results of Strang (1990). Second, we tested various ethno-demographic variables, including ethnic or religious fractionalization and the size of the largest ethnic group in 1980. This variable can be seen as a proxy to test some of Anthony Smith’s arguments. He maintained that nationalist mobilization is easier to achieve among groups with a rich ethno-history to turn into a nationalist narrative. Nationalism is also more likely to emerge in territories with clear-cut ethnic boundaries and a demographically dominant ethnic group that can form the “ethnic core” of a future nation (most clearly Smith 1990: 14; 11). We found no effect of ethno-demographic variables on the likelihood of nation-state creation. Finally, we tested whether nation-states are more likely to emerge in a democratic environment, trying to replicate Strang’s finding. However, political regime type does not affect the likelihood of nation-state creation in any statistically significant way.25

4.2 Does history repeat?
We now proceed to an analysis of the robustness of the main findings reported in table 2, using model 7 but without the term for the center’s capacity in independent states. The first question to ask is whether the results depend on our specific definition of nation-state creation and our research design that allowed for repeated events on the same territory. Model 1 in table 3 shows that this is not the case: Our results remain identical even if we exclude second and third episodes of nation-state creation from consideration, i.e. omitting the creation of the successor states of Gran Colombia, the Central American Republic and Yugoslavia as well as the repeated nation-state creations in the Baltic and Caucasus after the fall of the Soviet Union. The same holds true if we omit the creation of Gran Colombia, the Central American Republic and Yugoslavia and thus only model nation-state creations in their “final” form (results not shown).

25 We relied on Polity IV data on the political regime of the center (coding anocracy, democracy, autocracy using the standard definitions of these variables and assigning colonies and imperial dependencies an autocracy score). Political instability of the center (coded as a 3 point change in the Polity IV score over three years) had a very significant effect, but we don’t rely on this finding because of potential endogeneity problems. The ethno-demographic data refer to the second half of the 20th century only (from Fearon and Laitin 2003), which also creates a potential endogeneity problem: The earlier a nation-state created, the more time its elites had to assimilate or ethnically cleanse their territories, leading to higher share of the largest ethnic group. However, we found that there is no significant correlation between the ethno-demographic make-up of a territory and the cumulative number of years that it was the scene of ethno-nationalist wars between 1816 and 1960.
<table>
<thead>
<tr>
<th></th>
<th>(1) First NSC only&lt;sup&gt;a&lt;/sup&gt;</th>
<th>(2) With direct rule</th>
<th>(3) Before 1915</th>
<th>(4) After 1914</th>
<th>(5) Decades&lt;sup&gt;b&lt;/sup&gt;</th>
<th>(6) Empires&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of NS created in the empire past 5 yrs</td>
<td>0.124**</td>
<td>0.146**</td>
<td>0.286**</td>
<td>0.150**</td>
<td>0.112**</td>
<td>0.132*</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.05)</td>
<td>(0.106)</td>
<td>(0.036)</td>
<td>(0.043)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>No of NS created in neighborh. past 5 yrs</td>
<td>0.704**</td>
<td>0.724**</td>
<td>0.479*</td>
<td>0.613**</td>
<td>0.465**</td>
<td>0.618**</td>
</tr>
<tr>
<td></td>
<td>(0.152)</td>
<td>(0.185)</td>
<td>(0.24)</td>
<td>(0.154)</td>
<td>(0.123)</td>
<td>(0.121)</td>
</tr>
<tr>
<td>Years since first national organization</td>
<td>0.017**</td>
<td>0.016**</td>
<td>0.025**</td>
<td>0.008*</td>
<td>0.018**</td>
<td>0.019**</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Share of global power x dependency</td>
<td>-5.935**</td>
<td>-5.614*</td>
<td>-14.608*</td>
<td>-2.508</td>
<td>-5.300**</td>
<td>-4.884</td>
</tr>
<tr>
<td></td>
<td>(2.017)</td>
<td>-2.844</td>
<td>-5.808</td>
<td>-1.723</td>
<td>-1.337</td>
<td>-2.6</td>
</tr>
<tr>
<td>Number of wars fought in the empire</td>
<td>0.353**</td>
<td>0.351**</td>
<td>0.502**</td>
<td>0.232**</td>
<td>0.279**</td>
<td>0.312**</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.055)</td>
<td>(0.105)</td>
<td>(0.049)</td>
<td>(0.046)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>No. of wars fought in the territory</td>
<td>0.740**</td>
<td>0.727**</td>
<td>0.873**</td>
<td>0.808**</td>
<td>0.930**</td>
<td>0.876**</td>
</tr>
<tr>
<td></td>
<td>(0.179)</td>
<td>(0.265)</td>
<td>(0.179)</td>
<td>(0.245)</td>
<td>(0.168)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>Central govt. exp. for territory</td>
<td>-0.243*</td>
<td>0.172</td>
<td>-0.120**</td>
<td>-0.008</td>
<td>0.018**</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td>(0.038)</td>
<td>(0.038)</td>
<td>(0.025)</td>
<td>(0.033)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Oil prod.</td>
<td>1.587</td>
<td>0.538</td>
<td>Spanish</td>
<td>(1.032)</td>
<td>0.541</td>
<td>(1.079)</td>
</tr>
<tr>
<td></td>
<td>0.621</td>
<td>-0.055</td>
<td>Habsburg</td>
<td>(1.087)</td>
<td>0.427</td>
<td>(1.057)</td>
</tr>
<tr>
<td></td>
<td>0.301</td>
<td>0.371</td>
<td>Romano</td>
<td>(1.107)</td>
<td>0.41</td>
<td>(1.049)</td>
</tr>
<tr>
<td></td>
<td>0.275</td>
<td>-0.47</td>
<td>Ottoman</td>
<td>(1.116)</td>
<td>0.517</td>
<td>(1.079)</td>
</tr>
<tr>
<td></td>
<td>-1.141</td>
<td>-0.682</td>
<td>Yugoslav</td>
<td>(1.423)</td>
<td>0.368</td>
<td>(1.079)</td>
</tr>
<tr>
<td></td>
<td>0.83</td>
<td>-0.789</td>
<td>Soviet</td>
<td>(0.83)</td>
<td>0.366</td>
<td>(1.035)</td>
</tr>
<tr>
<td></td>
<td>0.793</td>
<td>0.366</td>
<td>French</td>
<td>(1.035)</td>
<td>0.354</td>
<td>(1.035)</td>
</tr>
<tr>
<td></td>
<td>-0.203</td>
<td>0.44</td>
<td>British</td>
<td>(1.13)</td>
<td>0.463</td>
<td>(1.13)</td>
</tr>
<tr>
<td></td>
<td>1.377</td>
<td>0.148</td>
<td>Dutch</td>
<td>(1.377)</td>
<td>0.406</td>
<td>(1.377)</td>
</tr>
<tr>
<td></td>
<td>2.432*</td>
<td>-1.096</td>
<td>Portuguese</td>
<td>(1.028)</td>
<td>0.622</td>
<td>(1.028)</td>
</tr>
<tr>
<td></td>
<td>1.823</td>
<td>0.709</td>
<td>Other empires</td>
<td>(1.052)</td>
<td>0.603</td>
<td>(1.052)</td>
</tr>
<tr>
<td></td>
<td>-0.443</td>
<td>0.018</td>
<td>Indep. states</td>
<td>(1.422)</td>
<td>0.298</td>
<td>(1.422)</td>
</tr>
<tr>
<td></td>
<td>3.084**</td>
<td>0.008</td>
<td></td>
<td>(1.296)</td>
<td>0.298</td>
<td>(1.296)</td>
</tr>
<tr>
<td></td>
<td>(1.044)</td>
<td>0.275*</td>
<td></td>
<td>(1.044)</td>
<td>0.275</td>
<td>(1.044)</td>
</tr>
<tr>
<td>Observations</td>
<td>15348</td>
<td>9123</td>
<td>11113</td>
<td>5372</td>
<td>16485</td>
<td>16301</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>400(14)**</td>
<td>610(8)**</td>
<td>105(8)**</td>
<td>164(8)**</td>
<td>403(20)**</td>
<td>372(20)**</td>
</tr>
</tbody>
</table>

* p<.05; ** p<.001

Notes: Cubic splines were included in all models except model 5. Their coefficients and the constant were omitted from the table.

<sup>a</sup>Dummies of continents were included in this model, but omitted from the table. <sup>b</sup>Years 1816-1820 are the reference category. <sup>c</sup>Pre-colonial territories are the reference category.
As a further robustness check we re-introduced the government expenditure variable, our proxy for the directness of rule, into the equation, which leads to a very significant drop in the number of observations since data are not available for most pre-colonial territories and many historically less well researched independent states. As model 2 shows, the government expenditure variable is significantly and negatively associated with nation-state creation, all other variables remain unaffected.

In how far do our findings hold across time and across the different waves of nation-state creation that we observe in the modern world? Isn’t the story of the creation of Latin-American nation-states quite different from that of the break-up of the Soviet Union? Shouldn’t we expect that variables affect the outcome in different ways, depending on whether we are focusing on the early 19th or the late 20th centuries? We addressed these important theoretical (cf. Tilly 1995; Collier and Mazzuca 2006) and empirical questions in five different ways.

First, we introduced interaction terms between all our independent variables and chronological time—to see if the explanatory power of a variable increases or decreases from 1816 to 2001—and found that none of them were significant except the interaction with literacy that we have discussed above, as well as the center’s share of global power variable, which turns from negative to positive over time (results not shown). As further analysis demonstrates, this effect arises because the break-up of the Soviet Union in the early nineties occurred all the while the share of global power of Moscow was still very high despite a significant drop since the eighties. Its share of global power is even exceptionally high compared to those territories that remain non-nation-states throughout the nineties (such as the Middle Eastern autocracies). In other words, the center’s share of global power argument does not help much in explaining the break up of the Soviet Union and its transformation into a series of nation-states, a finding to which we will return.

Second, analysis of different time periods also produces a remarkable resilience of our findings. Models 3 and 4 in table 3 relate to a subsample of observations preceding or succeeding the 1914 respectively. This is a substantially meaningful cut-off point not only because it is close to the midpoint in our time-series, but also because from a historical point of view the “long nineteenth century” ended with the First World War. One could also assume that Wilson’s 14 point plan changed the global legitimacy of the nation-state model and thus changed the likelihood of nation-state creation. All covariates remain significant and the magnitude of their effects is comparable, with the exception of the center’s share of global power argument does not help much in explaining the break up of the Soviet Union and its transformation into a series of nation-states, a finding to which we will return.

Robustness across subsamples does not mean, obviously, that additional factors are not relevant for specific time periods. As an example of such more time-bound effect we point to that of oil. Since the British navy used oil instead of coal to power its navy during the First World War, oil has been of

We also ran a fixed effect model, which controls for unmeasured, time-invariant differences across territories (results not shown). All variables behave in similar ways as in the standard model, but we loose significance on the center’s share of global power variable as well as the years since the foundation of first national organization. Both have to do with the fact that fixed effect models delete all units of observations in which the event has not occurred—precisely those territories in which nationalism arrived very late indeed.

We also analyzed periods after 1915, 1920, 1945, 1965, and 1975, as well as before 1890, 1870, and 1840. Usually one or two covariates become borderline or insignificant in these subsample analyses, with the exception of the post-1920 and again the post-1965 and -1975 models, when three core variables loose significance, including of course the center’s share of global power variable.

We record oil production on a territory (in thousand barrels per year; based on Wimmer and Min 2006). We used a natural log transformation of this variable to account for a skewed distribution due to the fact the most territories have no oil whatsoever.

---

26 We also ran a fixed effect model, which controls for unmeasured, time-invariant differences across territories (results not shown). All variables behave in similar ways as in the standard model, but we lose significance on the center’s share of global power variable as well as the years since the foundation of first national organization. Both have to do with the fact that fixed effect models delete all units of observations in which the event has not occurred—precisely those territories in which nationalism arrived very late indeed.

27 We also analyzed periods after 1915, 1920, 1945, 1965, and 1975, as well as before 1890, 1870, and 1840. Usually one or two covariates become borderline or insignificant in these subsample analyses, with the exception of the post-1920 and again the post-1965 and -1975 models, when three core variables lose significance, including of course the center’s share of global power variable.

28 We record oil production on a territory (in thousand barrels per year; based on Wimmer and Min 2006). We used a natural log transformation of this variable to account for a skewed distribution due to the fact the most territories have no oil whatsoever.
crucial importance for the dynamics of strategic big-power rivalry in the colonial world (Yergin 1993). We added an oil production per capita figure to the post-1914 model 4 and found territories with significant oil production to be less likely to become nation-states—yet another aspect of the “resource curse” that has recently received quite some attention (Ross Forthcoming). Historical examples for this effect are the broken promise of Arab independence, France’s tenacious hold onto its Algerian colony, Japan’s imperial expansion into South-East Asia (and especially Indonesia), and the Soviet re-conquest of Azerbaijan during World War I (cf. ibid). Theoretically, the effect of oil fits well into the power-shift argument that we advocate in this article: The resistance to nationalist movements depends not only on the center’s capacity, but also its strategic interest in a territory.

Other such period-specific effects could certainly be discovered. However, the aim of this project is to find a causal model of nation-state creation that is as little bound to such period-specific mechanisms as possible. Does this mean that our model describes law-like mechanisms that operate irrespective of historical circumstances and the particularities of each individual case? This would represent a rather naive interpretation of the meaning of regression results. The quantitative estimation techniques that were used above rely on a probabilistic notion of causality that does not pretend to fully explain each individual case. In many cases, other mechanisms and processes that are not captured by our data and analysis might be more important. After all, the variance explained by our models is low, and there remains ample room to underline contextual factors or the “eventful” nature of nation-state creation (Brubaker 1996).

Third, the degree to which such specifics do indeed play a role in explaining the emergence of nation-states can be further explored by a closer look at various dummy variables (for the analytical strategy of transforming context into cause, see Collier and Mazzuca 2006:483). Model 5 replaces the cubic splines as time specification with period dummies. It reveals that the only time periods significantly different from the omitted first five years in our dataset are the years 1956-1965 and 1986-1995 (as well as the last period from 1996 to 2001, which only contains 6 years, however). Further analysis shows that these two effects depend entirely on the year 1960, when many African states achieved independence, and 1991/1992, when the Soviet Union broke apart.

Our interpretation of these two decade effects points toward the lack of imperial will—as distinct from the capacity—to uphold an imperial domain. In contrast to the other transitions from empire to nation-state, the dissolution of the French and British empires in Sub-Saharan Africa were engineered in advance and in the end supported by the imperial center (on African decolonization, see XY). The Spanish, Habsburg and Ottoman empires, the Portuguese African empire or French Indochina and French Algeria provide illustrative contrasts. The Soviet Union on the other hand, quietly gave up on its dependencies and let the nationalist waves of mobilization in its former provinces proceed (Beissinger 2002). The result of such lack of willingness to uphold or defend an imperial domain explains why so many territories achieved nation-statehood in the very same year—despite a still strong center that has not been weakened by warfare and without that possible diffusion processes at work could be captured by a territory-year framework, which cannot grasp imitation and domino effects that unfold over weeks and months, rather than years. As a consequence, the concatenation and clustering of episodes of nation-state creation are not fully explained by our variables and the specificity of this constellation is picked up by period dummies. In other words, the lack of will by the metropolis to uphold control over its dependent territories has equivalent effects as do wars or a weakened political center in all other cases of imperial dissolution.

The oil variable is significant for foreign ruled territories only, as a subsample analysis indicates (results not shown here).
Fourth, we can further test whether the contextual characteristics and historical specificities of the various empires influence the likelihood of nation-state creation by including empire dummies. The results are displayed in model 6. We find that none of the imperial domains differ significantly from pre-colonial territories (the omitted category), nor do independent states (such as Switzerland, Thailand, the USA, etc.). Note that even after taking the particularities of these imperial domains into account, none of other coefficients in the model change much and remain significant, with the partial exception of the variable that measures the imperial center’s share of global power, which is now significant at the 6% level only.

Fifth, we would like to turn the reader’s attention back to table 2 and briefly comment on the effects of Continental dummies that were included in all the main models. While these dummies might capture a host of different elements—from history to culture and geography—we suggest to see them as an expression of the historical fact that nationalism originated in Western Europe and North America (the omitted categories), thus setting off regional diffusion and imitation processes earlier than in any other part of the world. This interpretation is supported by the case of Latin America, which is also not significantly different from North America and Western Europe. Latin American nationalists were inspired by the French revolution, which they came to learn about at the conference of Cadiz mentioned above, and by American independence next door. They thus were much more integrated in the political and cultural networks through which early nationalism diffused than any other part of the world. The same holds true for “Oceania”, which comprises New Zealand and Australia, which are also not significantly different from Western Europe and North America. To put it briefly, nationalism traveled more easily to societies dominated by European settlers in the New World and the Pacific than to other parts of the world that had less established connections to the regions where nationalism first originated. This network diffusion argument is consistent with the fact that the Continental dummies remain robustly significant even when other covariates that vary systematically across regions (such as the percentages of adherents to different religions) are entered into the equation.

5 Conclusion: Toward a historical institutionalist model of nation-state creation

We are now ready to go beyond detailed empirical analysis and attempt to integrate the diffusion and balance of power mechanisms into a more coherent and detailed causal narrative. We do so by relying on the comparative-historical tradition of research on state-building and political regime formation that understands the transition from one form of state-organization to another as the outcome of a political struggle between various politically organized segments of society, including state elites (Moore 1966; Tilly 1990; Skocpol 1979; Mann 1993; Breuilly 1993; Wimmer 2002). In this tradition of reasoning, shifts in power relations between these various forces determine which vision of a legitimate political order and which institutional principles will prevail. It is thus an essentially political understanding of institutional transformation and integrates both proximate reasons (historical constellations of power and the events that shape them) as well as more structural forces that change over the longue durée.

Historical institutionalism usually takes the emergence of new templates of political legitimacy as exogenously given. Conformingly, our own model does not concern the rise and global spread of nationalism, but the conditions under which nationalists are able to establish a nation-state. However,

---

30 The results change somewhat depending on the omitted category and the time specification. In some (theoretically less plausible) definitions of the omitted category, Portugal stands out as having a significantly decreased likelihood of letting its dependent territories become nation-states, in other specifications the dummy for the territories of Yugoslavia also achieves significance.
the two processes are intimately linked with each other: Nationalism has historically been associated with the rising military and political power of early nation-states such as Great Britain, the United States and France, much envied and resented by the rulers of neighboring polities (Greenfeld 1992). The success and global dominance of these three early nation-states represents the motor of the further diffusion of nationalism and the nation-state. Whether the coupling of global power and the nation-state model represents a historical accident, or whether there is some evolutionary advantage inherent to nation-states (especially the superior military mobilization capacity of a *peuple en armes*, see Spruyt 1996) need not concern us here—since the emergence and subsequent spread of nationalism across the world is not the focus of our analysis. We merely offered some hints—through the analysis of differences in the likelihood of nation-state creation across Continents—that the diffusion of nationalism might be propelled by imitation and contagion mechanisms that proceed along political and intellectual networks established within empires and neighborhoods.31

What are the conditions under which nationalist movements, once they come into existence, succeed in capturing existing states or to establish a breakaway state built on nationalist principles? At the core of our model is the constellation of power relating nationalist movements and leaders to the imperial centers, *ancien régimes* or the elites of nation-states against which the ethno-nationalist project is directed (see Figure 2).

**Figure 2**
A diffusionist model of nationalists’ empowerment

What determines whether a power shift in favor of nationalist reformers occurs? Several factors have turned out to be crucial, and we discuss them subsequently: (1) A power shift in favor of nationalism is more likely the stronger the nationalist movement, i.e. the more it has been able to mobilize larger segments of the population beyond the confines of the intellectual circles, army factions, segments of the lower clergy or lower level colonial bureaucrats that form the nucleus of nationalist movements (cf. Hroch 2000 [1969]). We have treated this as a monotonous process, assuming that the mobilizational

---

31 We conducted a preliminary quantitative analysis with the foundation of a first national organization as the outcome. This relates only to successful nationalisms, since the large number of nationalist movements that never managed to reach their goal is all excluded from consideration. We again find that modernization variables (railways track length, adult literacy rates, government expenditure for a territory) are not helping to explain nationalism, but diffusion mechanisms on the level of empires and neighborhoods are. Results are available upon request.
capacity increases over time. Other factors, such as the ethno-demographic homogeneity of a population, do not seem to be systematically related to this process, as far as our data allow us to tell.

(2) A shift of power also depends on how the strength of the established regime changes over time—its capacity to resist nationalist forces and to avoid institutional reform, abdication, or the loss of territory to secessionists. Wars fought either on the territory in question or elsewhere in the empire represent the most proximate cause influencing the staying power of the established elites.

Two additional factors influence the capacity to defend the status quo. They apply either to autonomous territories or to imperial dependencies only. The story therefore bifurcates and we need to distinguish between these two causal trajectories. The larger the size of government of autonomous states, the more they are able to control civil society forces, satisfy demands through reform and through provision of public goods, and thus hold nationalist forces in check. State strength cannot be translated into control and legitimacy in imperial states or in the internal colonies of nation-states, however, given that an expanded imperial government continues to contradict the like-over-like principle that nationalism establishes. In such states, the capacity to resist nationalist forces depends on the share of global economic and military power that the imperial center commands.

Additional, more context specific factors were discovered that relate to the willingness, rather than the capacity, of the political center to give in to nationalist demands. The first is the presence of natural resources such as oil, which increases the strategic utility of a dependency for the center. We also found that the French and British retreat from their African possessions, once these had lost their strategic and moral value for these empires, produced an unusual clustering of nation-state creations not fully explained by the more general mechanisms.

(3) Finally, we found two diffusion effects that produce a power shift in favor of nationalists. Already established nation-states within the same imperial domain demonstrate that independence is feasible and that the imperial or colonial center is no longer willing or able to uphold the status quo. Such imitation and demonstration effects empower nationalist movements and de-legitimize other, competing political forces, including those allied with the pre-national regime. Within neighborhoods, a nation-state creation may propel its own replication in adjacent territories through the domino effects produced by the lack of congruence between ethno-national and political boundaries: The fear (or reality) of second class citizenship in a newly nationalizing state empowers leaders of minority nationalists who seek a break-away state under their own control (see the break-up of British India). The possibility (or reality) of loosing territory to a neighboring nationalizing state empowers nationalist forces that are willing and able to defend the territory against irredentist challengers (see the rise of Committee for Union and Progress in Ottoman Turkey). More standard learning and imitation effects among the elites of neighboring territories could also responsible for the neighborhood diffusion effect. Such diffusion effects among neighbors or within empires not only empower nationalists in the remaining territories, but also weaken their political elites, who see themselves more and more surrounded by territories governed as modern nation-states, thus loosing resources to support their rule (in the case of empires) or alliance partners to keep nationalists under control (in the case of autonomous states).

We thus integrate balance-of-power and diffusion mechanisms into a simple model of power-shift that can thus subsume various types of factors of both causally distant and proximate nature. In contrast to a variety of political modernization arguments, our model emphasizes more contingent political factors such as war and the strategic considerations of imperial elites; it even turns some of the expectations of political modernization theory on their head: more expansive (and thus intrusive) governments do fend
off, rather than stimulate nationalist revolutionaries. In contrast to the classic theories of Gellner and Anderson, we show that economic and cultural modernization dynamics don’t play a crucial role in the spread of the nation-state across the world. In contrast to world-polity theory, we emphasize diffusion effects operating within neighborhoods and imperial domains, while the growing global legitimacy of the nation-state is not a better predictor of nation-state creation than the simple progression of time. In contrast to Roeder’s institutional capacity argument, we point to different factors that effectively produce a power shift in favor of nationalists, while at the same time supporting and building upon his general theoretical program to conceive nation-state creation as the outcome of a power struggle between contending political forces.

There is ample room for future research to improve upon what we have been able to achieve in this paper. First, obvious measurement and data limitations have been highlighted along the way and need to be addressed in the future. Given the scarcity and poor quality of data on units that have not already achieved independent nation-statehood, however, there are limits to what can be expected from such endeavors. Second and equally important are attempts to build more complex theoretical models of how the rise of nationalist doctrine interacts with inter-state and imperial rivalries, the dynamics of political contention and conflict between nationalist (and other) social movements and state authorities, cross-border linkages between such movements and states, and the co-evolution of political boundaries between provinces, states, and empires. Ideally, such a model would be of a fully processual nature and not shy away from treating variables as cause or outcome depending on which stage in the unfolding of a historical process we focus upon. This would allow to endogenously explain the emergence of new politically relevant identities such as nationalism, the shifting political geography of an ever more connected modern globe, as well as the various conflicts and wars that this realignment of identities and polities entails and through which they are in turn produced.
Bibliography


Woronoff, Jon. Various Years. *Historical Dictionaries (various volumes)*. Lanham: Scarecrow Press: