Threats to Territorial Integrity, National Mass Schooling, and Linguistic Commonality

Keith Darden[1] and Harris Mylonas[2]

Abstract

Why are some countries more linguistically homogeneous than others? We posit that the international environment in which a state develops partially determines the extent of its linguistic commonality and national cohesion. Specifically, the presence of an external threat of territorial conquest or externally supported secession leads governing elites to have stronger incentives to pursue nation-building strategies to generate national cohesion, often leading to the cultivation of a common national language through mass schooling. Comparing cases with similar levels of initial linguistic heterogeneity, state capacity, and development, but in different international environments, we find that states that did not face external threats to their territorial integrity were more likely to outsource education and other tools for constructing identity to missionaries or other groups, or not to invest in assimilation at all, leading to higher ethnic heterogeneity. States developing in high threat environments were more likely to invest in nation-building strategies to homogenize their populations.

[1]American University, Washington, DC, USA
[2]George Washington University, DC, USA

Corresponding Author:
Harris Mylonas, George Washington University, 2115 G street, NW, suite 440, Washington, DC 20052, USA.
Email: mylonas@gwu.edu

Given that all regions of the world had high levels of linguistic and ethnic heterogeneity in the past (circa 1300), how and why did European states end up more homogeneous than states in much of Africa and parts of Asia? The question is an important one, and it is surprising how rarely it is asked. In cross-national studies, the degree of linguistic heterogeneity—typically a proxy for, or at least the central component of, measures of ethnic heterogeneity—has been associated with everything from the provision of public goods and economic development to the likelihood of civil war (Alesina, Baqir, & Easterly, 1999; Easterly & Levine, 1997; Fearon & Laitin, 2003). It has generally been assumed in such studies that the degree of linguistic heterogeneity in a country is more or less fixed, given, and exogenous. There is little effort to explain diversity, or to ascertain whether the conditions that generate linguistic diversity are not responsible for explaining diversity's purported effects.

Following the works of Hintze (1975), Tilly (1990), Tilly and Blockmans (1994), Herbst (1990), Posen (1993), Wimmer (2012), Mylonas (2012), and others, we posit that interstate competition and, more specifically, the presence of external threats to a state's territorial integrity create incentives to nation-build that result in greater linguistic commonality. When states face threats to their territorial integrity through military conquest and/or the exploitation of ethnic divisions by external actors ("fifth columns"), governing elites have a strong incentive to nation-build to ensure loyal populations, to inoculate their citizenry against separatist appeals, and to ensure resistance against the intervention and territorial encroachment of rival powers.[1]

Although nation-building—the creation of a shared national identity among subject populations—can take many forms,[2] we focus primarily on mass schooling of the population with national curricular content and in a single standardized language.[3] Nation-building through mass education can indoctrinate previously unschooled populations into a coherent, shared national identity and establish a common, durable national loyalty that supersedes previous ethnic, family, and kinship ties, inoculates the population from external agitation, and ensures resistance to alien rule.[4] States that face threats to their territory from other states should be more likely to pursue nation-building through mass schooling to counter that threat—or else, they may cease to exist, or may experience territorial loss. When successful, this process creates national identification and cohesion and, as it typically uses a common language of education, significantly increases linguistic commonality as well.[5] Ceteris paribus, states that faced external threats to their territory will be both more cohesive and more linguistically homogeneous.[6]

When states develop in the absence of interstate competition, whether as a result of international norms and agreements or because their borders are

fixed and enforced by a regional hegemon, we expect them to be less likely to pursue any homogenizing policies and thus be less likely to invest heavily in mass schooling with national content. To the extent that they pursue education as a means for economic development or religious conversion, they are more likely to rely on private educational services provided by foreign or missionary organizations. These groups—often in competition with one another for congregants—are unlikely to coordinate to provide a single, coherent, national identity or to use a common language of instruction in the curricula. On the contrary, historical cases demonstrate that they cultivate regional, ethnic, or sectarian identities that are durable, undermining subsequent efforts to build effective state institutions or constructing a common national identity. Such conditions also produce lower levels of linguistic commonality.

In sum, the incentives to nation-build differ by the international—most often, regional—context in which states are situated. The degree of international competition and the external threats to a state's territorial integrity influence the incentives that elites have to pursue nation-building policies in education, which in turn has a durable impact on national cohesion and on the degree of linguistic commonality. Higher levels of interstate competition should be associated with investment in mass schooling with uniform national content; this, in turn, leads to greater linguistic commonality when language is tied to the definition of nationhood, as is true in the vast majority of cases. Linguistic heterogeneity is neither static nor exogenous to international politics: The international environment influences the likelihood that states will pursue nation-building policies that result in linguistic commonality.

The article proceeds in three sections. Section I outlines the macro cross-regional empirical pattern linking the external territorial threats and the generation of greater linguistic homogeneity, addresses potential alternative explanations for a country's degree of linguistic homogeneity, and lays out the theoretical logic tying the international incentives to nation-building policies and ultimately to greater linguistic homogeneity. In section II we focus on nation-building through mass schooling with uniform national content, and how this leads to higher levels of linguistic homogeneity. Section III focuses on empirical cases that help us delineate the link between the external environment, the incentives of governing elites, and the choice to pursue education policies that result in greater linguistic homogeneity. To control for initially low levels of state capacity, high diversity, and post-colonial rule, we provide a focused case comparison of Indonesia and Congo/Zaire—both highly diverse post-colonial states with low initial state capacity but differing in the degree of external threat to their territory.

Indonesia, facing an immediate threat from the former colonizers as well as externally backed groups, chose uniform mass schooling as an assimilation strategy resulting in greater linguistic homogeneity. Post-colonial Congo/ Zaire emerged in an environment where external territorial threats were limited or absent, and uniform schooling was not chosen, resulting in low levels of linguistic homogeneity. Finally, we examine cases where states did not act on the structural incentives that we identify (i.e., cases where states faced external threats but chose not to pursue nation-building) and lost territorial integrity, and one case (Tanzania) where nation-building was pursued in the absence of external incentives.

External Threats and Cross-Regional Variation in Linguistic Commonality

To measure the extent to which the population has been assimilated into a single language—or *linguistic commonality*—we code the share of the population that speaks the most common national language either as a first or as a second language. In most cases, this also serves as a measure of the degree of linguistic diversity and tracks closely with the ethnolinguistic fractionalization index (ELF) or other measures of diversity.[7] Yet our measure allows for diglossia (use of different languages in different settings) as well as multilingualism. This relates to a key conceptual point: Our interest is not in the possible markers for ethnic differentiation—which as Barth (1998 [1969]) has suggested are potentially unlimited—but in the achievement of commonality. We do not wish to know whether all potentially salient linguistic differences have been eradicated, only the extent to which a common language has spread, as the degree to which a common language is shared can serve as a proxy of the success of assimilationist nation-building policies for populations that were initially diverse.

Our assumption of initial diversity—or low linguistic commonality— among the large populations and sizable territories of most contemporary states finds its source in historical linguistics, which analyzes the entropic increase in linguistic diversity over time and space that leads to the development of different languages and dialects.[8] After the initial spread of a language across space—through conquest, trade, or migration—face-to-face contact diminishes, language changes accumulate naturally, and common languages fragment into dialects and eventually into mutually incomprehensible languages. The longer the amount of time since the initial spread of the language (such as Latin, Slavic, Germanic, Melanesian, or Bantu), the greater the linguistic diversity, especially where terrain or other conditions limit face-to-face contact. Given that most languages spread over a millennium

Table 1. Average Share of a Country's Population Able to Speak the Dominant Common Language (by Region).

	Average population share	Standard Deviation	Number of countries
Middle East, North Africa	94	8	19
Europe	93	5	20
Balkans	93	11	8
East Asia	91	11	4
South America	89	9	9
Former Soviet Union	86	11	16
Central and North America and Caribbean	86	12	14
Southeast Asia	74	20	10
South Asia	50	26	7
Sub-Saharan Africa	45	26	44

Source. Authors' calculations based on data from Lewis, Simons, and Fennig (2015), and National Censuses.

prior to modern state formation, we assume that on the territory of most contemporary states there was a high degree of initial diversity prior to the introduction of censuses and surveys capable of measuring it.

In some parts of the world, this natural process of linguistic differentiation and diversity was reduced dramatically. The degree to which this initial diversity was reduced with the creation of common languages has varied considerably across countries. As shown in Table 1, these cross-national differences also cluster regionally, with countries in Europe, East Asia, and the Balkans on average having achieved far higher rates of linguistic commonality than those in Southeast Asia, South Asia, or Africa. Using the Ethnolinguistic Fractionalization Index (ELF), we see similar regional patterns. Countries in sub-Saharan Africa had an average ELF score of .65 compared with .22 for Western Europe and Japan (*Atlas Narodov Mira*, 1964; Fearon, 2003).

What might account for this variation across countries and regions? Before examining the role of the external environment, we assess three alternative explanations for contemporary variation in linguistic homogeneity: initial diversity, initial (state) capacity, and time—or the duration of rule.

The most obvious explanation for the lower levels of linguistic homogeneity of states in some regions might be that countries in those regions—because of climate, geography, country size, or historical patterns of migration and settlement—were simply "naturally" more diverse than

others (Jahan, 1972; Wallerstein, 1960). The extreme linguistic diversity of Papua New Guinea—with 836 languages currently in active use and just more than half (53%) of the population speaking the dominant national language (Tok Pisin)[9]—was likely higher than the linguistic diversity of France well before the onset of mass education or other assimilation efforts due to the density of its jungles and the resultant isolation of its communities. Laitin, Moortgat, and Robinson (2012), building on Jared Diamond's (1997) *Guns, Germs, and Steel*, have argued that country shape has a considerable effect on the degree of diversity. In particular, "linguistic diversity should be more persistent to the degree that a geographic area is oriented more north-south than east-west" because "long" countries cross multiple climate zones and thus hinder the movement of people and technological innovations whereas "wide" countries do not (Laitin et al., 2012, p. 10263).

The high initial homogeneity of the Arabic-speaking countries of the Middle East and North Africa certainly precedes intervention by governing elites and deliberate policies of modern states to cultivate a common language. Arabic spread relatively recently to these regions—primarily between the 6th and the 11th century AD—and their populations retained a high degree of interaction through trade and migration. The centrality of Arabic to religious practice and pilgrimage also checked the normal linguistic entropy. As a result, even prior to the spread of Modern Standard Arabic as a literary language through mass education and media in the region, the various dialects of Arabic were still largely comprehensible to one another (Owens, 2001). This was true within countries, and largely true across countries as well. In this sense, the high homogeneity in the Arabic-speaking countries shown in Table 1 reflects a high initial or "natural" homogeneity and is not the result of deliberate nation-building policies.

However, the initial homogeneity of Arabic-speaking countries is an exception, and although it may be true that countries have initial ethnic "endowments" with relatively higher or lower levels of diversity, this does not help to explain the current cross-regional or cross-country variation. Virtually all non-Arabic countries had high initial diversity compared with current averages, and few countries in the world achieved homogeneity without deliberate intervention. Outside the Arab world, nearly all contemporary countries of more than a million people initially lacked a single standardized national language in which even half of the population could communicate with one another.

Even European states, now relatively homogeneous with an average of 93% of the population speaking a common language, were linguistically quite diverse if we look back one or two centuries. On the territory of

Germany, there were large Slavic-speaking populations (the Wends, the Poles, Lusatian Sorbs, etc.). Even among the German speakers, the dialects were so different that the mutual comprehension of spoken German between different localities within what became the German state only began to be realized during the 19th century (Wells, 1985). As late as the 19th century, the majority of France was populated by non-French speakers, and there were distinct regional languages: Provençal, Catalan, Breton, Basque, Corsican, Occitan (Langue d'Oc), and German (Grillo, 1989). As Eugen Weber (1976) notes, "The Third Republic found a France in which French was a foreign language for half the citizens" (p. 70). This may have been a slight exaggeration, as half a century of centralized schooling had already had a considerable effect. Nonetheless, even as late as 1863, 8,381 of France's 37,510 communes, or about a quarter of the population, spoke no French (Weber, 1976), and much of the remainder spoke patois that were not mutually comprehensible—as different from one another as the many Bantu languages of the Congo. In Western Europe, or in the cases of Indonesia and Tanzania discussed below, extremely high initial levels of heterogeneity were reduced to a single common national language. What differentiates the current level of homogeneity within a country does not appear to be its past or "initial" homogeneity, but whether and how the initial linguistic diversity was reduced in some cases and not in others.

Another plausible factor that could account for the variation in linguistic diversity might be the amount of time that a territory was subject to rule by groups who spoke another language. If assimilation were a gradual, natural process, then a long period of rule might lead to greater linguistic homogenization or the adoption of the language of the dominant group. Moreover, sovereigns who expected to hold power for a long time might have greater incentives to invest in the cultural attributes of their subject populations. Yet there are ample cases of extremely long reign with no substantial linguistic effect, and it also seems that linguistic homogenization can be accomplished in a rather short period of time when it is pursued through schools. For example, the Portuguese held Macau for 410 years, but Cantonese dominated everyday life, and English was used as the language of education. The Spanish held the Philippines from 1565 to 1898, but in the 1870 Philippine Census, only 2.5% of the population could speak Spanish. The same was true for the European core. The French crown was sovereign over much of the territory of contemporary France for five centuries, but in the beginning of the 19th century, only a minority of the population spoke French—again, only a minority were literate. Duration of rule does not seem to account for the observed variation in homogeneity.

Moreover, we know from quite recent cases that linguistic homogenization can be a rapid process. Prior to 1882, Hebrew was not a conversational spoken language anywhere in the world. In 1913, as part of an effort to achieve national unity among diverse Jewish populations, Hebrew was adopted at all levels of Jewish schooling in Palestine (Hofman and Fisherman, 1971). Despite the massive influx of non-Hebrew speakers with the founding of the state of Israel, more than doubling the population, Hebrew became the dominant spoken language. Yiddish-, Arabic-, French-, Ladino-, Polish-, Russian-, and Farsi-speaking populations were rapidly assimilated through the schools and a rigid national language policy. By 1998, Hebrew was the *native* language (L1) of 81% of Israel's population, and it is currently spoken by nearly all Israeli citizens as either a first or second language. Similarly, in Tanzania, Swahili went from a language spoken by a minority to a dominant majority language spoken by more than 90% of the population within a few decades (Young, 1976).

A final possibility we consider is that differences in homogeneity may be related to prior state capacity. It is plausible that states with prior capacity might be better able to implement mass education and other nation-building policies to homogenize their populations. Yet prior state capacity in other realms (tax collection, for example) is neither necessary nor sufficient for either mass education or homogenization. As shown in the cases of Indonesia and Tanzania below, nation-building through education was pursued by initially quite weak states—the capacity emerged endogenously, and was arguably a product of nation-building (Hobsbawm, 1990). Moreover, not all elites—even those governing relatively high-capacity states—seek to reduce heterogeneity by constructing a common national language and identity or base their rule on popular sovereignty (Darden & Grzymała-Busse, 2006; Gellner, 1983; A. Smith, 1986; R. Smith, 2003; Weber, 1976). Some countries are ruled through coalitions among different ethnic groups or factions (Bueno de Mesquita, Morrow, Siverson, & Smith, 2002; Bueno de Mesquita, Smith, Siverson, & Morrow, 2003; Riker, 1962), and others still on dynastic vestiges, or have focused on alternative models of authority and rule, such as communism (Connor, 1984; Greenfeld, 1992; Kedourie, 1970). In many cases, elites have institutionalized and cultivated multi-ethnic and polyglot polities. The key is not whether governing elites had the capability to homogenize their populations—as even the weakest states have shown a capacity to pursue homogenizing education policies—but whether they chose to do so.

External Threat and Nation-Building

We hold that a high level of linguistic commonality is at least partly a product of deliberate nation-building policies, and that the decision to pursue

nation-building is partially a function of the state's international environment. In particular, we suggest that states have an incentive to pursue linguistic homogenization when they face one of two types of external threats to their territorial integrity during their initial stages of state-building and/or independence: (a) the threat of military conquest[10] and/or (b) the threat of outside powers using non-core groups in their state to promote secession and/or to annex parts of their territory.[11] In making this claim, we build on work that highlights the importance of a competitive international environment, war-making, and imitation of successful military tactics as mechanisms that account for the spread of nationalism and the nation-state system (Posen, 1993; Tilly, 1990; Tilly & Blockmans, 1994; Wimmer, 2012), and we suggest that an important motivation for leaders to adopt nation-building is the reality, or anticipation, of threats to their territorial integrity by other powers. We argue that the presence or perception of these threats created incentives for governing elites to pursue nation-building policies through mass schooling, and that such policies are partially responsible for linguistic homogenization.[12]

The presence of external threats in the environment in which states emerged and developed varies by region and over time, and in ways that would plausibly have generated the regional variation in the degree of linguistic homogeneity described above in Table 1. Given that the relevant external environment for territorial conquest was often the regional neighborhood, we would expect different regional averages in the degree of linguistic homogeneity, and the results conform to our expectations.[13]

In Europe—whose countries have among the highest levels of homogeneity—external threats to territory were consistently high. Borders were constantly changing during the centuries that modern European states developed, and all states faced external threats to their territorial integrity (Spruyt, 1994; Tilly, 1975). Military competition over territory in northern and central Europe was high in the 18th and 19th centuries. Competition spread to southern and eastern Europe in the mid-19th and 20th centuries as imperial control by the Habsburg, Ottoman, and Romanov empires receded to leave behind a moving patchwork of competing states allied with major powers. Similarly, borders in East Asia remained fluid, contested, and fought over throughout the 20th century, during decolonization, and into the present.

The threat of territorial conquest by neighboring states or former colonizers in other regions has waxed and waned, and the risk of outside interference to promote secession in rival states has not been constant for all developing states. In Southeast Asia, colonial powers largely fixed the borders of their empires in the region by treaty, but decolonization led to greater variation across the region, with some countries—such as Indonesia—facing wars for

independence and territorial threats whereas others—such as Laos—faced no external threats.

At certain historical junctures, borders within a region are fixed by international agreement and enforced stably by great powers and regional associations. In Africa, following the Treaty of Berlin at the end of the 19th century, the Great Powers effectively "froze" state borders, and both ruling elites and their rivals faced a reduced risk of external threat. The post-colonial African elites colluded in 1964 to preserve the existing borders.[14] Nation-building through mass schooling among sub-Saharan African states was uncommon, and on average, they exhibit low levels of linguistic commonality. However, the timing of border fixity is critical. If states have already faced external territorial threats and pursued nation-building policies, subsequent decades of stable borders and low threat will not undo the homogeneity that is already present. The Monroe Doctrine and U.S. regional hegemony largely froze boundaries in the Americas, but only after a period of earlier post-independence wars—particularly in South America—had led to nation-building efforts under conditions of external threat (Centeno, 2002). Likewise, among today's European Union members, there is little external threat to a state's territory, but their populations are already educated and homogenization has already taken place.[15]

The regional differences in linguistic commonality reflect an underlying difference in the incentives of states in different international environments. Governing elites in states that were part of an international or regional environment that guaranteed border fixity had few external incentives to pursue nation-building policies of any kind; thus, costly policies to achieve linguistic commonality became less likely. Moreover, we expect that external backing of non-core groups would be less pronounced in regions where border fixity was perceived to be high, as it would also be unlikely to produce territorial changes and would reduce the incentives of outside powers to "meddle" in this way. Governments might choose to pursue nation-building in the absence of external incentives to do so. However, to the extent that these external incentives structure government choices, we would expect countries in areas of the world with border fixity and fewer external threats to experience fewer nation-building efforts, ceteris paribus.

In contrast, in cases where states gained their independence through war against the colonial power, where there were direct military challenges to territorial integrity following statehood, or military competition in the era when nationalism and popular sovereignty came to be linked to legitimate rule, or where an external power[16] attempted to cultivate a fifth column within their territorial boundaries, we would expect states to have incentives to

nation-build, to build loyalty and cohesion to enhance military effectiveness, and to make the population of their territory resistant to both separatism and "alien" rule. In a threatening external environment—one in which military competition is intense and borders are fluid—we would expect countries, ceteris paribus, to be more likely to pursue nation-building policies. If they fail to respond to those incentives, we would expect a greater likelihood that they would lose territorial integrity or "exit" the state system altogether.

Nation-Building, Mass Schooling, and Linguistic Commonality

Although there are many potential ways to nation-build (mass killing, population exchanges, settlement policies, etc.), we focus primarily on a commonly used assimilationist policy: the indoctrination of populations using mass schooling with uniform national content.[17] Because this national content almost always includes the introduction of a common national language as well as a common "constitutive story," (R. Smith, 2001, 2003) this nation-building strategy increases linguistic commonality along with national cohesion and loyalty. Although we are careful not to conflate those two outcomes—common language and national cohesion—the former can often serve as a useful proxy for the latter because the two were typically bundled together in mass education.[18]

Our argument rests on the assumption that governing elites deliberately used mass schooling with uniform national curriculum to cultivate loyalty and linguistic commonality. The link between education and national identity is one generally recognized in the contemporary academic literature—from Gellner (1983) to Hobsbawm (1990), and others. Far more important to our argument, the link between national content in education and political loyalty was central in the minds of policy-makers in the 19th and 20th centuries when mass education was pursued in most of the world. Whether in Fichte's (1806/1968) *Addresses to the German Nation*, Stalin's extensive work on the "History of the USSR" school textbook, or international negotiations between Serbia and Austria–Hungary over the acceptable national content of Serbian textbooks, there is ample evidence that policy-makers in the competitive geopolitical environment of Europe were acutely aware of the role of schools in cultivating loyalty and homogenizing their populations.

In the cases where assimilationist nation-building policies were pursued, schools in the late 19th and early 20th century purveyed nationalist content in the basic subjects of literature, history, geography, and music, in an effort to inculcate a sense of patriotism in their students and teach them a standardized version of the national language. Where this national content was singular

and pervasive across the territory of a state through a standardized curriculum, it forged common bonds of loyalty (Darden, in press; Darden & Grzymała-Busse, 2006; R. Smith, 2001). It placed the individual in an historical group based on a fictive bond of kinship, and cultivated loyalty and willingness to sacrifice for a sovereign state and for co-nationals. It was the basis for a new and powerful form of social cohesion.

Part of this effort to generate cohesion was the promotion of national language, and promotion of language and identity were typically linked. Mass schooling, when combined with the promotion of national language, reduced the linguistic diversity that was present in virtually all unschooled populations. Such state policies affected the linguistic tipping game described by David Laitin (1998) in *Identity in Formation* favoring the equilibrium on the side of the official language. Indeed, all "national languages" are aspirational constructs, created to achieve standardization of writing and speech rather than reflections of actual linguistic communities over large territories and populations. Prior to their construction, high diversity and a rapidly diminishing ability for populations to communicate as distances increased were the norm (Nichols, 1992). The greater homogenization that we find in some states today was typically, although not exclusively, achieved by the combination of so-called "national languages" and the employment of the enormous assimilating power of mass education. Indeed, sociolinguists concerned with the preservation of languages refer to these dominant national languages as "killer languages" because of the extent to which they eliminate linguistic heterogeneity. As summarized by Mühlhäusler (1996),

> The origin of such killer languages can be traced back to the development of European public schools, in particular in France, where the idea of central government was shaped by the insistence on having one language spoken by everyone in the nation. It is this idea, which was strongly supported by the European enlightenment and the French revolution and their imitators, which has dominated the social history of the world's languages ever since. The justifications for the dominance by these languages are variably given as political (how else could a nation be held together), economic (the cost of diversity) and moral (people need languages free from superstition and local antagonisms), but can all be accommodated under the common denominator of "development." (p. 20)

The considerable (and very expensive) institutional apparatus of mass education did not simply homogenize populations through a standardized national language. It introduced new, abstract categories of peoplehood that were not based on personal experience or the lived experience of one's parents. The constitutive stories that pervaded the curriculum gave meaning, history, and

context to new group definitions and loyalties. It was recognized at the time to be important in legitimizing the rule of the state and in justifying self-sacrifice for collective national goals.[19] As Darden (in press) shows, schools were quite effective in assimilating groups into a common national identity as well as acculturating them into a common language. These subsequent identities were remarkably robust and durable. It is likely that they independently account for some of the outcomes attributed to linguistic homogeneity.

We would expect to find the following empirical relationships between international competition, mass schooling with national content, and linguistic commonality:

1. States that initially developed in an environment of external threat, typically in regions of the world with higher levels of territorial competition, would be more likely to pursue nation-building policies. An important observable implication of this process is the pursuit of mass education with national content and a single national language. The result of a successful implementation (measured by the increase in the literacy rate through this policy) is greater linguistic commonality.
2. States that developed in environments where external threats to their territory were low, typically in regions of the world with high border fixity and low interstate military competition, would be more likely to have non-unitary educational systems, with multiple languages, and lower linguistic commonality, as countries face no strong external incentive to homogenize their populations.

Empirics

Our argument is structural (Waltz, 1979/2010). We suggest that the presence (and/or perception) of external territorial threats creates incentives for states to nation-build, but governing elites may not act on them. When they do, elites may not consciously recognize or discuss these linkages, and this reasoning may not appear in direct archival or process-tracing evidence.[20] Moreover, states that fail to pursue nation-building in the presence of external threats to territory do not necessarily lose territory immediately—but our argument suggests that probabilistically they are more likely to face this outcome, particularly over the long run.[21]

Our argument is also bounded by historical time. External territorial threats have existed for millennia, but the concepts of nationalism and the use of nation-building and mass schooling as a way to achieve cohesion are relatively recent phenomena. It is only when the idea of nationalism and its utility in countering external threats spread across the world, a process that

largely began in Europe in the late 18th century, that the presence of exter-
nal threats would lead states to adopt nation-building policies pursuing
homogenization. Moreover, the lag times between the establishment of
mass educational systems with uniform national content and the linguistic
homogenization of the targeted populations are both long and variable—
measured in decades and working through intergenerational change.

Given that these conditions complicate data collection and statistical esti-
mation, it is more fruitful to explore these general patterns at work in indi-
vidual cases (Mahoney, 2000; Slater & Ziblatt, 2013).[22] The paradigmatic
cases of high territorial competition leading to mass schooling with uniform
national content and high subsequent levels of assimilation are Western
European states, such as France. In Europe, the ongoing competition between
neighboring states and the education of their populations in standard national
languages had a mutually ratcheting effect. Posen (1993) has shown how
defeat at the hands of the French mass army led Prussia to spread schooling
with a uniform national content, initially down to the level of the non-com-
missioned officer, to facilitate command, training, and political motivation.
The same logic and the conquests of Napoleon's France led other neighbor-
ing Germanic states to pursue the education and homogenization of their
populations (Harp, 1998; Schleunes, 1989). It was widely believed that the
success of the Prussian educational reforms was behind the effectiveness of
their army and their defeat of France in 1871, which led the French them-
selves to replicate the German reforms with the Ferry Laws of 1881 and for
countries as far away as Japan to bring in German bureaucrats to replicate
the German educational model (Harp, 1998; Weber, 1976). In all of the great
powers, standardized education in a common national language expanded
dramatically in the late 19th century and especially in the lead-up to the First
World War.[23] As Boli and Ramirez (1987) point out, the remarkable univer-
sality in the pursuit of mass education with national content by European
states in the 19th and early 20th centuries, despite their many differences in
levels of development and in other conditions typically considered relevant
to the expansion of education, suggests that the distinctively competitive
interstate environment best accounts for the rapid expansion and replica-
tion of this model.

Yet despite their many differences, European states also generally had
higher levels of prior state capacity than states in sub-Saharan Africa or Asia
where high levels of diversity were not eradicated through state-planned
assimilationist nation-building policies. Thus, to better control for prior state
capacity, longer state duration, and other qualities of European states that
might have influenced their ability to generate educational capacity and
reduce linguistic heterogeneity, we need to explore non-European cases. In

particular, the ideal cases to test our argument are countries where there was high initial linguistic heterogeneity and low initial state capacity.[24]

To highlight the structural effect of external territorial threat we compare cases of initially linguistically heterogeneous states, with relatively limited central state capacity, that either faced external territorial threats as they developed during and following independence (Indonesia) or, due to international agreements, developed in conditions where their borders were perceived as fixed and were externally guaranteed (Congo/Zaire).

Indonesia: High External Threat and Mass Schooling With Uniform National Content

When Indonesia secured independence from the Dutch in 1949, there were more than 750 languages spoken. The population of 75,500,000 was dispersed across 18,000 islands, 922 of which were permanently inhabited.[25] Per capita gross domestic product (GDP) was a mere US$803 in 1950, among the lowest in the world (Maddison, n.d.). A mere 15% to 20% of the population above 15 years of age was literate (UNESCO, 1957). The most commonly spoken language was Javanese, but it was spoken by less than a majority. Bahasa Indonesian—which was selected as the state language by the main Indonesian nationalist movement—was spoken by less than 5% of the population at the time of independence. With a large, dispersed, and diverse population and a newly independent state with limited bureaucratic capacity and low tax collection, Indonesia lacked high initial state capacity and the geographic and demographic conditions that would facilitate state-building or the construction of national cohesion.

Indonesia also faced both types of external threats discussed above at the critical early period when it achieved independence, and it initiated a nation-building campaign with mass schooling. First and foremost, Indonesia achieved independence through war. The Dutch and British, who had taken over the administration of the Dutch East Indies after the defeat of Japan, constituted a critical external threat at the formative period of independence. With the defeat of the Japanese in the Second World War, the former Dutch East Indies came under British military control, and the British stated their intention to return the sovereignty of the former colony back to the Dutch. In response, the Indonesian national movement declared the country's independence in 1945 and fighting between British and Dutch forces and the Indonesian national independence movements ensued for 4 years. Even after the withdrawal of the Dutch and their recognition of Indonesian independence in 1949 (although West Papua continued to be disputed), the Indonesian state continued to

engage in boundary disputes with former colonial powers and Malaysia over territories that the national movement considered Indonesian.[26] Borders remained contested, with some disputes resolved by aggressive military action and annexation.

Second, much of Indonesia's late colonial and early independence history involved the mobilization of different ethnic groups by outside powers for the purpose of undermining an independent Indonesian state or limiting its territory. External meddling was rife.

The leadership of Indonesia's independence movement was forged as a collaborationist government under Japanese rule during the Second World War. The Japanese, in an effort to build solidarity against the British and the Dutch, constructed nationalist forces drawing primarily on the Javanese—the largest linguistic group. This was, in part, a reaction to the British and the Dutch, whose colonial administrations had sought to mobilize non-core groups in a divide-and-rule strategy to balance the nationalism of the core (Javanese), and also to mobilize the population against Japanese occupation and, subsequently, Indonesian independence. The Dutch pushed for the proliferation of small ethnic states out of their former colonial holdings by supporting a "Pasundan" state, Balinese nationalism, and the Ambonese (Moluccan islands), as well as other territorial challenges to the Indonesian state. As Horowitz (1985) notes, "the existence of a strong anti-colonial movement sharpened divisions between those groups most active in the movement and those most closely associated with the colonial regime" (p. 517). We posit that it also increased the incentives to pursue nation-building policies to transcend ethnic lines, homogenize the population linguistically, and generate national cohesion—which is precisely what the Indonesian government did.

Even prior to the War, and in response to Dutch efforts to promote ethnic divisions, the nationalist movement—which developed in Dutch-controlled Indonesia in 1928—selected a Malay language, Indonesian, as the national language to foster their efforts against the Dutch. During the war, the nationalists collaborated with the Japanese. The latter happily promoted Indonesian nationalism and Indonesian language in an attempt to undermine the Dutch and British. Under Japanese occupation, the Dutch language was banned, and Bahasa Indonesian (hereafter, "Indonesian") was made the language of administration and was the only language taught in schools.

After the Cold War, rivalry with neighboring Malaysia (one largely of Indonesia's own making) and post-colonial meddling by former imperial powers continued to raise the fears of externally promoted secession and insurgency that manifested during the independence period.[27] A persistent fear of links between ethnic Chinese and external powers—both Taiwan

and the People's Republic of China—also spurred incentives for nation-building.

In this context of external threat, and beginning in the 1950s, Indonesia pursued a very active strategy of national language promotion through schools, with the stated aim of forging national unity and inoculating the country from the threat of foreign-promoted agitation and secession.[28] Beginning in the early 1950s, the Indonesian government conducted a mass schooling and literacy campaign, complete with adult education classes, the creation of "people's libraries," and the opening of thousands of schools. The campaign did not bring literacy levels above 50% in any area except for the municipality of Jakarta, but it established mass educational institutions with Indonesian as the sole national language and language of instruction (beyond Grade 3) throughout the country.[29] In Indonesia as a whole in 1952-1953, there were 12,307,026 school-aged children, and 6,391,101 (or 52%) were enrolled and attending school.[30] Enrollments and literacy rates continued to increase throughout the 1950s and 1960s, with literacy increasing from 42.9% in 1961, to 59.6% in 1971, to 67.3% in 1980, to 81.5% in 1990, and above 90% in 2004.[31]

By 2010, the population had reached 237,642,000, and 210 million (88%) were Indonesian speakers (Indonesia Population Census, 2010). Although there is continued bilingualism, according to Mühlhäusler (1996), "The diglossia or bilingualism appears to be of a transitional kind, leading within the next two generations towards monolingualism in Indonesian" (p. 205). Javanese and other languages spoken by millions of Indonesians have not disappeared, but Indonesian is the only language that is sustained institutionally in government, education, and public life.[32] Linguists concerned with preservation of diversity write of the "epidemic language shift" in Indonesia, or the role of Indonesian as a "killer language" as a result of its promotion through mass education. Normative issues about diversity-preservation aside, it is clear that mass homogenization has been accomplished in the Indonesian case as part of a deliberate nation-building policy forged at the moment of independence—side by side with more exclusionary policies toward certain non-core groups—in an environment of clear, present, and perceived threats to territorial integrity.

Low Geopolitical Competition and Mass Schooling With Heterogeneous Content: Sub-Saharan Africa

As noted above, in sub-Saharan Africa, there has been a near absence of interstate war, boundary change, and external threat since the late 19th century. Most countries achieved their independence following the relatively

peaceful withdrawal of each respective colonial power, and for most of the last century, even where there was external involvement in African civil wars, there was no effort to seize territory and alter borders (exceptions include the Eritrean–Ethiopian war, the Ethiopian–Somali War, and the Uganda–Tanzania war). The majority of the conflicts in sub-Saharan Africa were civil wars. All in all, in the 20th century, the incentives to nation-build were simply not as strong as in parts of Asia or in 19th century Europe.

How has this international environment shaped the internal development of nation-building policies and ultimately the degree of linguistic commonality? Sub-Saharan African countries are among the most linguistically diverse. Virtually all sub-Saharan African states have preserved a high degree of "cultural pluralism" (Young, 1976). We suggest that this diversity is the result of a failure of both colonial powers and independent African states to adopt nation-building policies, mainly because they lacked the incentives to do so. In these cases, initial linguistic heterogeneity was often reinforced rather than reduced. We suggest that in a different international environment, things would have been otherwise.

Take the case of Congo/Zaire. Like Indonesia, Congo began with a high degree of linguistic heterogeneity, with 110 distinct languages reported by *Ethnologue* in 1969 and the largest being Luba, with 3 million speakers out of a total population of 17.5 million, or a mere 17%. Literacy rates at independence were similarly low, with 33% being able to read and write in any language as recorded in 1970—and even at this low level, there was literacy in multiple languages.[33] Moreover, Congo's state capacity at decolonization was not much better than Indonesia's. Its GDP was $748 per capita in 1960.

As Young (1985) points out regarding Congo/Zaire, there was high potential for consolidation, homogenization, and standardization among the related Congo–Niger languages, and particularly the Bantu languages that were spoken by a vast majority of the population. In particular, Young describes a general "Kongo linguistic zone" in which a standard written (national) language could have been easily established, much as Bahasa Indonesian was established as a national language to integrate the Malay dialects. Yet education within this "Kongo linguistic zone" was fragmented among many different evangelical agencies, each of which pursued different languages and identities in the primary education during the colonial period.[34] The primary unit both for language and loyalty was not the nation but the congregation, and education reinforced and differentiated those units (Young, 1985). Even after independence, education remained primarily in the hands of the churches (MacCaffrey, 1982). Despite a short-lived and incomplete effort by the state to take control in 1974 and pursue "Zairianization," the churches were

granted full authority again in 1976. Literacy increased to 50% by 1980, but there was neither a single centralized government curriculum with a uniform national content nor a national language cultivated, there were multiple. In the case of Congo/Zaire, the cultivation of linguistic heterogeneity stemmed from both colonial and post-independence decisions to delegate schooling to missionary organizations. In an environment of fixed and guaranteed borders, they had no incentive to nation-build.

The result has been continued linguistic heterogeneity. Although French is the official national language, only 9% of the population speaks French proficiently, with an additional 30% having some French language ability. Lingala, which was selected as a lingua franca in the colonial period, has approximately 2,141,300 native speakers and an additional 7 million who speak it as a secondary language for a total of only 13% of the population. A few languages cultivated in schools are in official use in different regions or branches of government (Kikongo, Lingala, Tshiluba, and Swahili), but there is no dominant language. Lingala, for example, is the official language of the army, a telling artifact of the country's history of mobilizing ethnic factions against internal rivals in the absence of a need for collective mobilization against external threats.

Other sub-Saharan African cases provide us with a similar story. Zambia began with high levels of linguistic heterogeneity, but also with some core linguistic commonalities that might have allowed for the development of common national languages. Posner (2003) focuses on the case of Zambia and explains "how colonialism affected not just the formation of ethnic groups, but also their numbers, relative sizes, and spatial distributions" (p. 127). Posner shows how colonial policies as well as missionary and mining company actions led to the consolidation of the language map in Zambia from more than 50 to just 4, although literacy rates at the end of the colonial period were only 41.3%.

In the independence period, Zambia faced no external territorial threats. By 1990, 78.8% of Zambians used Bemba, Nyanja, Tonga, or Lozi as their first or second languages of communication (Posner, 2003). Despite this consolidation process from many to few languages—there was no concerted nation-building effort and thus linguistic commonality remains low, with Bemba being the largest spoken language but covering only 26% of the population. Moreover, even this limited consolidation has preserved a high degree of diglossia. As Posner (2003) put it, "[P]eople did not trade in their 'old' language for a 'new' one but developed language repertoires that included both" (p. 142).

Zambia and Congo were able to pursue nation-building, yet they lacked the external incentives to do so. Their states were no weaker and societies no

more diverse than Indonesia's. The exceptional case of Tanzania demonstrates that the opportunity to nation-build through assimilation and schooling was available to similar sub-Saharan African states, even if the geopolitical environment provided few incentives to pursue it. Like Zambia and Congo (or Indonesia), Tanzania started out with a very high level of linguistic heterogeneity and low state capacity. Upon independence, Tanzania had approximately 181 different languages, with the largest group being the Sukuma with approximately 15 million native speakers, or a mere 12% of the population at the time (Kaplan, 1968, Appendix A).[35] Taxes as a share of GDP were a mere 15% in 1961 and GDP per capita was only $459. Tanzania also had inherited missionary schools that, although they reached only a small percentage of the population with literacy rates less than 20%, had initially cultivated additional heterogeneity.

Shortly after independence, however, the state seized control of all educational institutions and pursued a standard education in Swahili with strong efforts to cultivate a common national identity out of the many ethno-tribal distinctions in place following independence (Hydén, 1980).[36] Spending on education increased significantly as part of a national education campaign, and by 1981 the literacy rate exceeded 50% and Swahili had become the dominant language. To be sure, Tanzania's choice was not a function of its external environment, but of a communist-style leadership that pursued nation-building for other reasons. Yet despite the different reasons for selecting assimilation through schools as a strategy, the effects were quite similar. And here, as in the Indonesian case, a high level of linguistic commonality was achieved as a result of nation-building policies.[37] What this exceptional case reveals is that nation-building through mass schooling with national content was an available and feasible option. Yet for most of sub-Saharan African states, the external incentive to nation-build was absent and thus such policies were rarely pursued.

The Bite of Structural Reality: Failure to Respond to External Incentives and Loss of Territorial Integrity

The case of Tanzania serves as a reminder that governing elites do not always make nation-building choices based on the external threat environment. Where states with no external incentive to nation-build through schooling pursue these policies nonetheless, the result is typically linguistic commonality and national cohesion that may serve other developmental goals (Miguel, 2004). Yet when governing elites face external threats and either do not respond to these incentives with nation-building policies or govern educated populations with distinct and formed national identities,[38] then this inability

to produce national cohesion will come with a cost. When the structural incentives we posit are present, but nation-building is not pursued or is unsuccessful, we would expect the loss of territorial integrity.

A common cause for "failure" of nation-building policies is that prior efforts at identity construction have already been successful, that is, the population was already schooled in another language and identity. Such situations resulted when countries were formed out of or incorporated into territories that had already been schooled with a uniform national content. In Yugoslavia, for example, Slovene and Croat populations had already been schooled in the Habsburg Empire, and literacy rates were approximately 85% and 60%, respectively, when these areas were incorporated into the Kingdom of Serbs, Croats, and Slovenes in 1918.[39] Moreover, neither the interwar Yugoslav regime nor post–World War II Communist Yugoslavia ever centralized their school curriculum. Fragmentation served Tito's goal to keep the Serbs in check and prevent the subnational assimilation of smaller Slavic groups by the Serb nation following World War II. Separate identities, not a common Yugoslav identity, were cultivated in each of the curricula of the Yugoslav republics.

As a result, literacy levels in Yugoslavia were well above 90% by the 1980s, indicating that mass schooling had been fully completed, but the population had been schooled in different national identities in the curricula generating little cohesion and loyalty to the Yugoslav state relative to the sub-state national identities. Yugoslavia repeatedly struggled with periods of loss of territorial integrity culminating in its violent dissolution in the 1990s. We find an analogous case in Lebanon, where approximately 50% of the population was literate as early as 1950, but the educational system was divided along regional sectarian lines without a common national content (UNESCO, 1957). Here, too, territorial integrity has barely been retained and sometimes even lost as a result of externally backed groups engaging in periodic warfare.

Cyprus, one of the most durable and intractable civil conflicts, exemplifies the cases of high literacy, high heterogeneity, and the loss of territorial integrity due to external threat and incursion. Ethnic diversity was cultivated and politicized through education. Communal education and the cultivation of conflicting constitutive stories about the island were a function of early British colonial stewardship as part of their "divide and rule" policy. In the 19th and early 20th centuries, the British, with no incentive to build a Cypriot national identity did not interfere in the education of the island treating the Turkish and Greek communities as distinct religious communities with the right to control their own schools—an Ottoman legacy. From 1878 to 1895, the schools were entirely organized on a voluntary communal basis, with

limited funding from the British administration. The Education Law of 1895 formalized the communal division of the schools by placing the Turkish and Greek schools under the authority of separate boards. It was only with the Education Laws of 1929 and 1933, as Cyprus became a Crown Colony of the Empire, that the schools were placed under the direct authority of the British colonial administration (Persianis, 1978, 1996). Even at this point, the communal structure was largely preserved.

Greek Cypriot elites mobilized first, and the development of a Turkish identity took place at least partially in response, and with outside assistance from the Turkish Republic, so as to counterweight Greek claims to the whole island. Greek nationalism spread among Christian Orthodox Cypriot elites in the late 19th and early 20th centuries, largely with the encouragement of the Orthodox Church and with the support of the Greek government. As a result, most Greek Cypriot schools used standard Greek textbooks and curricula, which stressed the gathering of all historical Greek lands and Greek peoples under the sovereign rule of the Greek state.[40] The education of the Greek population progressed rather rapidly and extended from elites toward nearly full enrollment by the 1930s. When the British took over the school system of the island in 1931, they pushed for universal primary education, but as with British imperial practices elsewhere, they retained the highly decentralized control of curricula (Persianis, 1996).

As a result of communal education, Cyprus was left with two communities with sharply different national loyalties. The Christian Orthodox community of the island had been schooled to identify strongly as Greek and to desire *enosis* with the Greek nation-state. The Muslim Cypriot community, although less educated, had been schooled to identify with the Turkish state.[41] Schooled to view the Turkish or Greek states as a mortal enemy, neither community could be expected to allow Cyprus to fall under the sovereignty of either state. It is not surprising that Greek efforts at *enosis* sparked militant opposition among the Muslim community, nor that the Turkish invasion of the island in 1974 led to the loss of territorial integrity—an event welcomed by the Muslim community but that appalled and traumatized the Christian Orthodox community.

Finally, some states developed within other states or empires that were not motivated by a national homogenization imperative and thus did not attempt to build cohesion based on common identity and language. For instance, many of the Soviet republics that became independent in 1991 had embedded groups with non-national ethnic identities as a result of Soviet nationalities policies in education (Martin, 2001; Suny & Martin, 2001). Following the collapse of the Soviet Union and consequent independence, they have faced both threats to and actual loss of territorial integrity as the result of externally

backed separatist movements. South Ossetians, Abkhaz, and Adjars each had autonomous republics with educational curricula that set them apart from Georgians. Chechens were not taught to be Russian. Pamiris in the Gorno-Badakhshan region of Tajikistan had an autonomous province and education, and Tajik identity was not cultivated. Slavs and Moldovans in Transdniester had been educated in the Soviet system, whereas Romanian speakers in the majority of Moldova had been schooled to identify as Romanians in interwar Romania. Armenians in the Karabakh region of Azerbaijan were educated to identify as Armenians, with a history curriculum that was antagonistic to the Turks. Russia, Georgia, Moldova, Tajikistan, and Azerbaijan all had high literacy rates in 1991, but the heterogeneity of the content of their national education meant that these countries did not have cohesive national loyalties and no common language within their populations other than Russian. All have experienced temporary or permanent loss of territorial integrity.

Conclusion

Every corner of the world was once linguistically and ethnically heterogeneous. Patterns of migration, trade, and settlement, as well as natural entropy with time and geographic isolation, generated linguistic and ethnic heterogeneity in virtually all of the territories that eventually became sovereign states. In some countries, regionally concentrated in Europe and East Asia, linguistic heterogeneity was reduced through the propagation of national languages through mass education as well as through exclusionary policies. In others, with a regional concentration in sub-Saharan Africa, heterogeneity was consolidated and sometimes even reinforced. Nation-building policies were never pursued.

In this article, we provided part of an explanation why. Put simply, we suggested that there is a link between external threats—both direct confrontation and indirect efforts to promote ethnic divisions to undermine and divide rival states—and the decision to nation-build, and that this connection can account for some of the broad, basic regional patterns in the degree of linguistic commonality and national cohesion that we currently find across the world. State elites facing external threats in an era when ethnic differences could be politicized had incentives to homogenize their populations. In cases where external threats were largely absent, heterogeneity was often preserved or unintentionally reinforced.

These links between territorial competition, nation-building, and the degree of linguistic homogeneity had hitherto been insufficiently explored.[42] A longstanding literature has explored interstate military competition and statebuilding—a tradition that runs from Otto Hintze through Charles Tilly, Jeffrey

Herbst, and an array of contemporary scholars from Atzili to Thies. The links between interstate military competition and the state's ability to tax and spend have been long established, but the links between external threat and linguistic commonality had not. Given the importance attributed to ethnolinguistic diversity (ELF) in explaining development, state capacity, and internal war, the causes of homogeneity and national loyalty must be identified. We suggest that many of those causes are found in the external, interstate environment.

One thing is clear; linguistic heterogeneity is often endogenous to many of the political and developmental processes it is often drawn on to explain. As Posner (2003) noted in his study of linguistic divisions in Zambia, a country's ethnolinguistic diversity is not "just a social fact but a historical product" (p. 142). In this article, we have demonstrated that a systematic pattern underlies the historical production of diversity and commonality. This "product" is oftentimes the result of ruling elites' conscious efforts to nation-build using mass schooling with national content in reaction to external threats to the territorial integrity of their states.

What we offer is a partial explanation for why governments would pursue nation-building and, in particular, mass education with uniform national content based on external incentives. To be sure, mass schooling was not the sole means through which states achieved national cohesion and linguistic commonality. States have also used religious institutions, subsidies, and tools other than education to assimilate subjects and to secure their loyalty.[43] Moreover, rather than assimilating populations, state elites have also exchanged, expelled, or killed populations believed to belong to potentially disloyal groups. According to Mylonas (2012), which groups are targeted with assimilation rather than extermination largely depends on the content of the state's constitutive story as well as the interstate relations between a group's external patrons and the host state.

Regardless of the path to national cohesion and linguistic homogenization, the long-term consequences of these processes may be both durable and significant. The broad patterns we have identified suggest that the decision to nation-build through schools may subsequently determine whether states are likely to experience internal war, particularly ethnic civil war, and perhaps enjoy the benefits of national loyalty in tax collection, conscription, and the provision of public services. Heterogeneity, or ethnolinguistic fractionalization, may be better understood not as an obstacle to state-building or an explanation for state weakness, but as the *product* of a decision not to engage in nation-building because of the absence of external threats. Linguistic heterogeneity, in this sense, is more of an attribute of weak states or an outcome of prior decisions—a condition shaped by international causes—rather than an exogenous or underlying cause of their weakness or instability.

Authors' Note

Acknowledgments

Declaration of Conflicting Interests

Funding

Notes

1. We assume that domestic elite consensus will develop under such conditions.
2. For more on nation-building policies, see Mylonas (2012).
3. For more on violent policies toward non-core groups, see Bulutgil (2009), Downes (2006, 2008), Harff (1987, 2003), Mann (2005), Mylonas (2010), Naimark (2001), Rae (2002), Snyder (2000), Straus (2006), Valentino (2004), and Wimmer (2002).

4. For the full development of this argument, see Darden (in press).
5. To be sure, (a) this is not the only path that produces linguistic commonality, and (b) not all states pursue this policy by enforcing a standardized language, for example, Switzerland. We do not conflate language and identity—common language is a usual by-product of the same educational process by which national cohesion is generated, but linguistic commonality and identity/cohesion are distinct phenomena. It should be noted, however, that the binding of language and national identity is prevalent in all of the ethnically defined nationalisms and most of the civic versions as well. Notable exceptions include Switzerland, and France in the early part of the 19th century.
6. Although it is a topic for further research, we suspect that many of the supposed effects of linguistic homogeneity in the political economy literature are actually driven by higher levels of national cohesion, which co-varies with linguistic commonality but is far more difficult to measure and observe.
7. The ethnolinguistic fractionalization (ELF) measure is capturing the likelihood that two randomly selected people from a country will be members of different ethnic or linguistic groups. It is based on data collected by Soviet social scientists in the early 1960s. For more, see *Atlas Narodov Mira* (1964); for a discussion of the problems with this measure, see Posner (2004).
8. For a review, see Nichols (1992).
9. Authors' calculations based on Lewis et al. (2015).
10. Charles Tilly has argued that external military threats operated as incentives for governing elites to state-build. For more, see Tilly (1990). Barry Posen (1993) has argued that imitation of the Mass Army model utilized by the French in the early 19th century was the mechanism for the spread of nationalism in the case of Prussia and beyond.
11. For a more elaborate version of this argument, see Mylonas (2012).
12. In virtually all cases—both for civic and ethnic nationalisms—nation-building was tied to the creation of a common national language.
13. This sub-section is drawing from ideas developed in Mylonas (2013). Many scholars have demonstrated that the threat to territory has varied both over time and across different regions of the globe; see Atzili (2012); Gross (1948); Thies (2005); and Zacher (2001).
14. The borders were fixed by a decision of the Organization of African Unity (OAU). Four disputes contributed to the OAU decision to treat existing borders in Africa as relatively fixed (Morocco–Algeria, Somalia–Kenya, Somalia–Ethiopia, Ghana–Upper Volta; Förster, Mommsen, & Robinson, 1989; Herbst, 1990; Jackson & Rosberg, 1982; Touval, 1967).
15. This will be a stable outcome, although large-scale migration waves may upset this equilibrium.
16. The term external power, includes other states (great powers, regional powers, neighboring states) as well as diaspora groups.
17. For a full discussion of the link between mass education and the formation of national identity, see Darden (in press). For a richer treatment of the selection

of alternative strategies of nation-building, and on the different strategies that states pursue toward their non-core groups, see Mylonas (2012). For more on the management of cultural diversity, see Hechter (1975) and McGarry and O'Leary (1994).

18. In those (relatively rare) cases where national identity is not bundled with a common language or identity in the constitutive story—such as Switzerland—we would nonetheless expect that territorial competition and external threat should lead to incentives to build national cohesion, even if linguistic heterogeneity is preserved.

19. This was a central point of the seminal Prussian writing on education in the 18th and 19th centuries. For more, see Schleunes (1989) and Harp (1998). On the general pattern, see Boli and Ramirez (1987).

20. Having said that, Mylonas (2012) has provided ample evidence that governing elites are consciously pursuing such policies in response to external threat.

21. For example, elites may express the priority of creating national cohesion, but they may take for granted that the need for national cohesion is generated by their particular external environment, and thus would not mention it. In environments where the external environment does not provide incentives, elites may simply not recognize national cohesion as a goal. Whether or not national cohesion (and linguistic homogenization) is a goal may be determined by the environment, but without conscious or direct mention by those who formulate education policy. It will be only identified in the broader empirical pattern.

22. There are impressive efforts to measure external territorial threat of states on a regional level for Latin America; see Thies (2005).

23. Speaking to the proposed link between education, loyalty, and military effectiveness, the primary statistics on literacy from this period are often taken from surveys of incoming conscripts (Cipolla, 1969).

24. Israel is probably the case with the highest initial linguistic diversity and the clearest external threat—and one of the most rapid recorded cases of linguistic homogenization—but it is exceptional in other ways, and many of its inhabitants were migrants from European countries.

25. Population in 1950: UNESCO (1957, p. 39).

26. For more on the logic of territorial claims by nationalist movements, see Mylonas and Shelef (2014).

27. In principle, the maritime border between Indonesia and Malaysia was set by the 1824 Anglo-Dutch treaty, but the land border on Borneo had been revised multiple times (1891, 1915, and 1928). Both Indonesia and Malaysia, as the successor states to the British and the Dutch, were bound by the earlier treaties, but border clashes and boundary disputes have marked the history of the relations between the two countries—the most serious being the *Konfrontasi* between 1963 and 1966, in which Indonesia refused to recognize the newly independent Malaysia and challenged the borders of Malaysia using force on the island of Borneo.

28. The strategy of linking national identity to Bahasa Indonesian and the importance of a common national language were articulated by nationalist elites during

the struggle for independence from the Dutch and British. Its origins lie in the period in which external threat was most present and acute (Alisjahbana, 1962, 1974, 1976; Anderbeck, 2015; Mühlhäusler, 1996).

29. *Development of Education in Indonesia*, Table 9 (June, 1954), p. 18-19.

30. *Development of Education in Indonesia*, Table 4 (June, 1954), p. 12.

31. 1961 and 1971 data from the Indonesian National Census. Series after 1980 from World Development Indicators.

32. According to coding in Lewis et al. (2015), there are 719 languages in Indonesia, 706 living and 13 extinct. Nineteen are institutional, 86 are developing, 261 are vigorous, 265 are in trouble, and 75 are dying.

33. Subsequent data define literacy as the ability to read and write in Lingala, Tshiluba (Luba), Kingwana, or French.

34. Describing the heterogeneity of schools in the non-Muslim areas of Zaire, Crawford Young (1985) lists "Walloon Scheutists in Mayombe, Redemptorists in the Matadi area, Flemish Jesuits centered in Kisantu, Protestant Swedes north of the river, American Baptists towards the West, and the English Baptist Mission Society in the Eastern areas. Each of these mission societies developed its own standard version [of the language] . . . The resulting institutionalized fragmentation of the language has been a major barrier to its diffusion as a lingua franca" (p. 79).

35. Young (1976) reports that the difference between the Sukuma and the Nyamwezi was an artificial colonial construction, and the two groups combined would constitute 17% of the population. Although there were only 7,721 native Swahili speakers in 1966 (Kaplan, 1968), Young cites estimates that approximately half of the population could speak Swahili (as a second language) in 1942.

36. For a review of educational nation-building strategies in Tanzania and Kenya, see Miguel (2004). Miguel's findings are very much consistent with those predicted by our theory.

37. And state capacity also emerged endogenously, with taxes as a percentage of gross domestic product (GDP) rising to 35% by 1974.

38. For the detailed argument linking the initial mass education of the population with the establishment of durable national identities, see Darden (in press).

39. Based on the Habsburg census of 1910 (cited in Jelavich, 1990).

40. The classic statement of the Megali Idea, as articulated by Kolettis before the constitutional assembly in January 1844 reads, "The kingdom of Greece is not Greece; it is only the smallest and poorest part of Greece. Greece includes not only the kingdom, but also Janina and Thessaloniki, and Serres, and Adrianople, and Constantinople, and Trebizonde (sic), and Crete, and Samos, and any other country where Greek history or the Greek race was present" (cited in Jelavich, 1990, p. 78).

41. Although it is likely, given the later development of Turkish Cypriot schooling, that this minority population could have been more influenced by later British efforts to develop a common Cypriot national identity, this is a matter that requires further research.

42. An important exception is Posen (1993), who argues that national loyalty has been cultivated for the battlefield effectiveness of the mass army, but not as a general strategy for achieving a loyal population that will resist external annexation, secession, or external provocation.
43. Sambanis, Skaperdas, and Wohlforth (2015) suggest that victory in interstate wars can significantly increase a state's international status, and thus making it easier for leaders to induce individuals to identify nationally, thus reducing internal conflict by increasing investments in state capacity.

References

Alesina, A., Baqir, R., & Easterly, W. (1999). Public goods and ethnic divisions. *Quarterly Journal of Economics*, *114*, 1243-1284.

Alisjahbana, T. S. (1962). *Indonesian language and literature: Two essays (The Modernization of the Indonesian Language in Practice and The Development of the Indonesian Language and Literature)*. New Haven, CT: Yale University.

Alisjahbana, T. S. (1974). Language policy, language engineering and literacy in Indonesia and Malaysia. In J. A. Fishman (Ed.), *Advances in language planning* (pp. 391-416). The Hague, The Netherlands: Mouton.

Alisjahbana, T. S. (1976). *Language planning and modernization: The case of Indonesian and Malaysian*. The Hague, The Netherlands: Mouton.

Anderbeck, K. (2015). Portraits of language vitality in the languages of Indonesia. In I. W. Arka, N. L. N. Seri Malini, & I. A. M. Puspani (Eds.), *Language documentation and cultural practices in the Austronesian world: Papers from 12-ICAL* (Vol. 4, Asia-Pacific Linguistics Open Access Monographs, A-PL 019 / SAL 005, pp. 19-47). Canberra, Australia: Asia-Pacific Linguistics. Retrieved from http://hdl.handle.net/1885/13514

Atlas Narodov Mira. (1964). Moscow: Miklukho-Maklai Ethnological Institute at the Department of Geodesy and Cartography of the State Geological Committee of the Soviet Union.

Atzili, B. (2012). *Good fences, bad neighbors: Border fixity and international conflict*. Chicago, IL: University of Chicago Press.

Barth, F. (1998 [1969]). *Ethnic groups and boundaries: The social organization of culture difference*. Long Grove, IL: Waveland Press.

Boli, J., & Ramirez, F. (1987). The political construction of mass schooling: European origins and worldwide institutionalization. *Sociology of Education*, *60*, 2-17.

Bueno de, Mesquita, B., Morrow, J. D., Siverson, R. M., & Smith, A. (2002). Political institutions, policy choice and the survival of leaders. *British Journal of Political Science*, *32*, 559-590.

Bueno de, Mesquita, B., Smith, A., Siverson, R. M., & Morrow, J. D. (2003). *The logic of political survival*. Cambridge, MA: The MIT Press.

Bulutgil, Z. (2009). *Territorial conflict and ethnic cleansing* (Doctoral dissertation). University of Chicago, IL.

Centeno, M. A. (2002). *Blood and debt: War and the nation-state in Latin America*. University Park: Penn State University Press.

Cipolla, C. M. (1969). *Literacy and development in the West*. New York, NY: Penguin.

Connor, W. (1984). *The national question in Marxist-Leninist theory and strategy*. Princeton, NJ: Princeton University Press.

Darden, K. (in press). *Resisting occupation in Eurasia: Mass schooling and the formation of durable national loyalties*. New York, NY: Cambridge University Press.

Darden, K., & Grzymała-Busse, A. M. (2006). The great divide: Literacy, nationalism, and the communist collapse. *World Politics*, *59*, 83-115.

Diamond, J. (1997). *Guns, germs, and steel*. New York, NY: W.W. Norton.

Downes, A. B. (2006). Desperate times, desperate measures: The causes of civilian victimization in war. *International Security*, *30*(4), 152-195.

Downes, A. B. (2008). *Targeting civilians in war*. Ithaca, NY: Cornell University Press.

Easterly, W., & Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics*, *112*, 1203-1250.

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

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review*, *97*, 75-90.

Fichte, J. G. (1968). *Thirteenth address, addresses to the German Nation* (G. A. Kelly, Ed.). New York, NY: Harper Torchbooks. (Original work published 1806)

Förster, S., Mommsen, W. J., & Robinson, R. E. (1989). *Bismarck, Europe, and Africa: The Berlin Africa Conference 1884–1885 and the onset of partition*. Oxford, UK: Oxford University Press.

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

Greenfeld, L. (1992). *Nationalism: Five roads to modernity*. Cambridge, MA: Harvard University Press.

Grillo, R. D. (1989). *Dominant languages: Language and hierarchy in Britain and France*. Cambridge, UK: Cambridge University Press.

Gross, L. (1948). The peace of Westphalia, 1648-1948. *The American Journal of International Law*, *42*, 20-41.

Harff, B. (1987). The etiology of genocides. In I. Walliman & M. N. Dobkowski (Eds.), *Genocide and the modern age: Etiology and case studies of mass death* (pp. 41-59). New York, NY: Greenwood Press.

Harff, B. (2003). No lessons learned from the Holocaust? Assessing risks of genocide and political mass murder since 1955. *American Political Science Review*, *97*, 57-73.

Harp, S. L. (1998). *Learning to be loyal: Primary schooling as nation building in Alsace and Lorraine, 1850-1940*. DeKalb: Northern Illinois University Press.

Hechter, M. (1975). *Internal colonialism: The Celtic fringe in British national development*. Berkeley: University of California Press.

Herbst, J. (1990). War and the state in Africa. *International Security*, *14*(4), 117-139.

Hintze, O. (1975). Military organization and the organization of the state. In F. Gilbert (Ed.), *The historical essays of Otto Hintze* (pp. 159-215). New York, NY: Oxford University Press.

Hobsbawm, E. (1990). *Nations and nationalism since 1780*. Cambridge, UK: Cambridge University Press.

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

Hydén, G. (1980). *Beyond Ujamaa in Tanzania: Underdevelopment and an uncaptured peasantry*. Berkeley: University of California Press.

Indonesia Population Census. (2010). *Penduduk Indonesia Hasil Sensus Penduduk 2010* [Result of Indonesia Population Census 2010]. Retrieved from http://bps.go.id/website/pdf_publikasi/watermark%20_Penduduk%20Indonesia%20Hasil%20SP%202010.pdf

Jackson, R., & Rosberg, C. (1982). Why Africa's weak states persist: The empirical and the juridical in statehood. *World Politics*, *35*, 1-24.

Jahan, R. (1972). *Pakistan: Failure in national integration*. New York, NY: Columbia University Press.

Jelavich, C. (1990). *South Slav nationalisms: Textbooks and Yugoslav Union before 1914*. Columbus: Ohio State University Press.

Kaplan, I. (Ed.). (1968). *Tanzania: A country study*. Washington, DC: American University Press.

Kedourie, E. (Ed.). (1970). *Nationalism in Asia and Africa*. New York, NY: Frank Cass.

Laitin, D. D. (1998). *Identity in formation: The Russian-speaking populations in the near abroad*. Ithaca, NY: Cornell University Press.

Laitin, D. D., Moortgat, J., & Robinson, A. (2012). Geographic axes and the persistence of cultural diversity. *Proceedings of the National Academy of Sciences*, *109*, 10263-10268.

Lewis, P. M., Simons, G. F., & Fennig, C. D. (Eds.). (2015). *Ethnologue: Languages of the world, eighteenth edition*. Dallas, TX: SIL International. Available from http://www.ethnologue.com

MacCaffrey, W. (1982). Education, religion, and social structure in Zaire. *Anthropology & Education Quarterly*, *13*, 238-250.

Maddison, A. (n.d.). *Countries Compared by Economy > GDP per capita in 1950—International Statistics at NationMaster.com*. NationMaster. Retrieved from http://www.nationmaster.com/country-info/stats/Economy/GDP-per-capita-in-1950

Mahoney, J. (2000). Strategies of causal inference in small-N analysis. *Sociological Methods & Research*, *28*, 387-424.

Mann, M. (2005). *The dark side of democracy: Explaining ethnic cleansing*. Cambridge, UK: Cambridge University Press.

Martin, T. D. (2001). *The affirmative action empire: Nations and nationalism in the Soviet Union, 1923-1939*. Ithaca, NY: Cornell University Press.

McGarry, J., & O'Leary, B. (1994). The political regulation of national and ethnic conflict. *Parliamentary Affairs*, *47*, 94-115.

Miguel, E. (2004). Tribe or nation? Nation building and public goods in Kenya versus Tanzania. *World Politics*, *56*, 327-362.

Mühlhäusler, P. (1996). *Linguistic ecology: Language change and linguistic imperialism in the Pacific region*. London, England: Routledge.

Mylonas, H. (2010). Assimilation and its alternatives: Caveats in the study of nation-building policies. In A. Lawrence & E. Chenoweth (Eds.), *Rethinking violence: States and non-state actors in conflict* (pp. 83-116). Cambridge, MA: BCSIA Studies in International Security, The MIT Press.

Mylonas, H. (2012). *The politics of nation-building: Making co-nationals, refugees, and minorities*. New York, NY: Cambridge University Press.

Mylonas, H. (2013). *Whither nation-building?* e-International Relations. Retrieved from http://www.e-ir.info/2013/05/08/wither-nation-building/

Mylonas, H., & Shelef, N. (2014). Which land is our land? Domestic politics and change in the territorial claims of stateless nationalist movements. *Security Studies*, *23*, 754-786.

Naimark, N. (2001). *Fires of hatred: Ethnic cleansing in twentieth century Europe*. Cambridge, MA: Harvard University Press.

Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago, IL: University of Chicago Press.

Owens, J. (2001). Arabic sociolinguistics. *Arabica, T. 48, Fasc. 4*, 419-469.

Persianis, P. (1978). *Church and state in Cyprus education*. Nicosia, Cyprus: Violaris Press.

Persianis, P. (1996). The British colonial education "lending" policy in Cyprus (1878-1960): An intriguing example of an elusive "adaptive education" Policy. *Comparative Education*, *32*, 45-68.

Posen, B. (1993). Nationalism, the mass army and military power. *International Security*, *18*(2), 80-124.

Posner, D. N. (2003). The colonial origins of ethnic cleavages: The case of linguistic divisions in Zambia. *Comparative Politics*, *35*, 127-146.

Posner, D. N. (2004). Measuring ethnic fractionalization in Africa. *American Journal of Political Science*, *48*, 849-863.

Rae, H. (2002). *State identities and the homogenisation of peoples*. Cambridge, UK: Cambridge University Press.

Riker, W. H. (1962). *The theory of political coalitions*. New Haven, CT: Yale University Press.

Sambanis, N., Skaperdas, S., & Wohlforth, W. C. (2015). Nation-building through war. *American Political Science Review*, *109*, 279-296.

Schleunes, K. A. (1989). *Schooling and society: The politics of education in Prussia and Bavaria 1750-1900*. Oxford, UK: Berg Publishers.

Slater, D., & Ziblatt, D. (2013). The enduring indispensability of the controlled comparison. *Comparative Political Studies*, *46*, 1301-1327.

Smith, A. (1986). State-making and nation-building. In J. Hall (Ed.), *States in history* (pp. 228-263). Oxford, UK: Basil Blackwell.

Smith, R. (2001). Citizenship and the politics of people-building. *Citizenship Studies*, *5*, 73-96.

Smith, R. (2003). *Stories of peoplehood: The politics and morals of political member-ships*. Cambridge, UK: Cambridge University Press.

Snyder, J. L. (2000). *From voting to violence: Democratization and nationalist con-flict*. New York, NY: W.W. Norton.

Spruyt, H. (1994). *The sovereign state and its competitors*. Princeton, NJ: Princeton University Press.

Straus, S. (2006). *The order of genocide: Race, power, and war in Rwanda*. Ithaca, NY: Cornell University Press.

Suny, R. G., & Martin, T. (Eds.). (2001). *A state of nations: Empire and nation-making in the age of Lenin and Stalin*. Oxford, UK: Oxford University Press.

Thies, C. (2005). War, rivalry, and state building in Latin America. *American Journal of Political Science*, *49*, 451-465.

Tilly, C. (Ed.). (1975). *The formation of national states in Western Europe*. Princeton, NJ: Princeton University Press.

Tilly, C. (1990). *Coercion, capital and European states: AD 990-1990*. Cambridge, MA: Basil Blackwell.

Tilly, C., & Blockmans, W. P. (Eds.). (1994). *Cities and the rise of states in Europe, AD 1000 to 1800*. Boulder, CO: Westview Press.

Touval, S. (1967). The organization of African unity and African borders. *International Organization*, *21*, 102-127.

UNESCO. (1957). *World illiteracy at mid-century: A statistical study*. Paris, France: Author.

Valentino, B. (2004). *Final solutions: Mass killing and genocide in the twentieth cen-tury*. Ithaca, NY: Cornell University Press.

Wallerstein, I. (1960). Ethnicity and national integration in West Africa. *Cahiers d'études africaines*, *1*(3), 129-139.

Waltz, K. N. (2010). *Theory of international politics*. Long Grove, IL: Waveland Press. (Original work published 1979)

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

Wells, C. J. (1985). *German: A linguistic history to 1945*. Oxford, UK: Oxford University Press.

Wimmer, A. (2002). *Nationalist exclusion and ethnic conflict: Shadows of modernity*. Cambridge, UK: Cambridge University Press.

Wimmer, A. (2012). *Waves of war: Nationalism, state formation and ethnic exclusion in the modern world*. New York, NY: Cambridge University Press.

Young, C. (1976). *The politics of cultural pluralism*. Madison: University of Wisconsin Press.

Young, C. (1985). Ethnicity and the colonial and post-colonial state in Africa. In P. Brass (Ed.), *Ethnic groups and the state* (pp. 57-93). London, England: Croom Helm.

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

Author Biographies

Keith Darden is an associate professor in the School of International Service at American University specializing in comparative politics, international relations, and the politics of Eurasia. His first book, *Economic Liberalism and Its Rivals: The Formation of International Institutions Among the Post-Soviet States*, was published in 2009 by Cambridge University Press and won the 2010 Hewett award from Association for Slavic, East European, and Eurasian Studies (ASEEES) and an honorable mention for the Shulman prize. His second book, *Resisting Occupation: Mass Literacy and the Creation of Durable National Loyalties*, to be published by Cambridge University Press in 2014, explores how the national identities initially introduced to a community through schools account for subsequent patterns of voting, secession, and armed resistance to foreign occupation. His work has been published in *World Politics*, *Ethnopolitics*, *Politics and Society*, and *Foreign Affairs*.

Harris Mylonas is an assistant professor of political science and International Affairs at George Washington University. His research focuses on the processes of nation- and state-building, the politicization of cultural differences, and diaspora management policies. He is the author of *The Politics of Nation-Building: Making Co-Nationals, Refugees, and Minorities* (Cambridge University Press, 2012), for which he won the 2014 European Studies Book Award by the Council for European Studies, which honors the best first book on any subject in European Studies published within a 2-year period as well as The Peter Katzenstein Book Prize for the best first book on international relations, comparative politics, or political economy in 2013. His second book project, *The Strategic Logic of Diaspora Management*, analyzes why nation states develop policies to cultivate links with and/or to attract back certain diasporic communities but not others. His work has been published in *Perspectives on Politics*, *Security Studies*, *Comparative Political Studies*, *Ethnopolitics*, *European Journal of Political Research*, *Social Science Quarterly*, and various edited volumes.