Nation-building*

Alberto Alesina
Harvard and IGIER Bocconi

Bryony Reich
University College London

First Draft: October 2012
Latest Revision: August 2013

Abstract

Nations stay together when citizens share enough values and preferences and can communicate with each other. Homogeneity amongst people can be built with education, teaching a common language to facilitate communication, infrastructure for easier travel, but also by brute force such as prohibiting local cultures. Democracies and non-democracies have different incentives when it comes to choosing how much and by what means to homogenize the population. We study and we compare both regimes in a model where the size of countries and the degree of active homogenization is endogenous. We also offer some historical discussions of cases which illustrate our theoretical results.

1 Introduction

> *"There cannot be a firmly established political state unless there is a teaching body with definitely recognized principles. If the child is not taught from infancy that he ought to be a republican or a monarchist, a Catholic or a free-thinker, the state will not constitute a nation; it will rest on uncertain and shifting foundations; and it will be constantly exposed to disorder and change."* Napoleon I, 1805[1]

In 1860 French was still a foreign language to half of all French children.[2] Outside major cities, France was a country of dialects[3] and diverse currencies.[4] Travel far outside one's own

*We thank Jeffrey Frieden, Oded Galor, Alessandro Riboni, Enrico Spolaore and participants in seminars at UCL, Warwick, a CEPR meeting, an ISNIE meeting, and the NBER Summer Institute for useful comments. Giulia Giupponi and Andrea Passalacqua provided excellent research assistance.

[1] Quote by Napoleon I, Ramirez and Boli (1987).

[2] Estimate Weber (1979) p67. Hobsbawm (1990) p60 gives a figure of $12-13\%$ of the population who spoke French 'correctly' at the French Revolution.

[3] Weber (1979) in just a few case studies mentions Basque, Béarnais, Catalan, Flemish, Germanic dialects, Artesian, Picard, dialects of Boulongne, Artois, Picardy, and so on.

[4] Weber (1979), $p30-40$.

village was rare, and indifference or hostility to the French state common.[5] Napoleon had considered the making of 'Frenchmen' a prerequisite to the future of a stable France. Indeed from the French Revolution and throughout the 19th Century the French rulers expressed the imperative 'to form French citizens'.[6] Following the unification of Italy (1860), Massimo d'Azeglio (one of the founders of unified Italy) famously remarked: 'Italy has been made; now it remains to make Italians.' In 1860 at most 10% of the Italian population spoke what would become the Italian language, there was only one railway line which crossed any of the pre-unification states, and many were openly hostile to the new nation.[7] During the 19th Century those who governed France and Italy implemented a range of policies with the aim of building commonality among the population and 'forming' what they determined to be 'Frenchmen' and 'Italians.' For instance they introduced state controlled education, including compulsory elementary schooling; banned languages other than the 'national language' in schools, religious services and administration; introduced systematic military service often with the explicit aim of integrating individuals from different parts of the country; and extended road and rail links.

These are just two examples. History has witnessed a multitude of efforts to "nation build", that is policies enacted by rulers to build homogeneity amongst their populations. Tilly (1975) observes that 'almost all European governments took steps which homogenized their populations: the adoption of state religion, expulsion of minorities, institution of a national language, eventually the organization of mass public instruction.' Hobsbawn (1990) observes that 'states would use the increasingly powerful machinery for communicating with their inhabitants, above all primary schools, to spread the image and heritage of the 'nation", while Anderson (1979) notes that 'the official or culture-language of rulers and elites usually came to be the official language of modern states via public education and other administrative mechanisms.' In contrast, European elites did not enact such policies in their colonies (Christopher, 1988; Michalopoulos and Papaioannou, 2012); yet once these colonies gained independence in the 1950's and after, many introduced policies to create a national language and national identity, similar to the policies of 19th Century Europe. As well as homogenization in newly independent countries, the 20th Century also saw dictators and political elites who prohibited local cultures and attempted to impose their ideologies, often through odious means (for example the Soviet Union, Nazi Germany, China, or Franco's Spain). Franco's policies, as he put it, were aimed to create 'a single language, Castilian, and a single personality, the Spanish one.'[8]

Why did 19th Century elites see homogenization as imperative? Why not in their colonies? Why did those colonies undertake nation-building after independence? Why did the Soviet Union and other modern dictatorships undertake such harsh homogenization? Do these experiences have implications for the long-run stability of a country?

[5] Weber (1979), p95 − 114; 485 − 496. It is also argued that knowledge of the nation of France itself was not always guaranteed. In 1864 a school inspector in Lozère noted that not a single child could answer questions such as 'Are you English or Russian?', p110. On travel, p195 − 220. Note that fifty percent of France's population was estimated to be Farmers or peasants in 1870, Weber (1979) p8.

[6] Quote from Félix Pécault in 1871 who conducted a general inspection of public education for the French government. See Weber (1979) for many more examples.

[7] Duggan (2007); the railway line was the Piacenza-Bologna line, Schram (1997).

[8] Jones (1976).

The goal of this paper is to analyze policies of nation building in its more or less benevolent forms, across political regimes and in times of transition from one regime to another. We define "nation building" as a process which leads to the formation of countries in which the citizens feel a sufficient amount of commonality of interests, goals and preferences so that they do not wish to separate from each other.[9] The equilibrium size of a country emerges from a trade-off between economies of scale in the production of public goods and services or the size of the market and the heterogeneity of the population, which may have different priorities and preferences for shared public goods, languages or institutions. In this respect we follow Alesina and Spolaore (1997).[10] We depart from these authors, however, in an important way since we assume that the degree of divergence of preferences amongst the population is endogenous: we explicitly model the choice of the central government regarding how much to homogenize the population.[11]

When and why would a particular regime undertake such homogenization? First consider a democracy. Within a country the population only has access to one "government", a catch-all term for what a public sector does. However, people disagree on which "government" they prefer, i.e. the "location" of the government which can be interpreted either geographically or in terms of preferences. The majority benefit from a certain degree of homogenization. For example, better roads or railways to the capital city improve the individuals' access to resources located there and may avoid distant minorities becoming isolated and disenfranchised;[12] schooling in a common language enables better participation in the democratic process; "indoctrination" in common values reduces heterogeneity of preferences so that policies and public goods are a better fit. However, since homogenization (schooling, roads, etc.) is costly, the majority chooses to homogenize up to the point in which marginal benefits equal marginal costs. In some cases the median voter might choose a level of homogenization which avoids an otherwise sure split of the country.[13]

By comparison, a non-democratic regime which is in full control of the population and faces little probability of being overthrown has different incentives. A non-democratic regime, like a dictator or ruling elite, will choose a "government" which matches their own preferences and will tax the maximum possible number of people in order to pay for the public goods chosen (or to simply extract rents). Such a regime has no concern for the welfare of the population and, since policies and public goods already match the preferences of those in charge, homogenization of the wider population is useless.

The incentives of a non-democratic regime which faces a more substantial probability of

[9] Recently, state building and nation building have sometimes been used interchangeably; however state building generally refers to the construction of infrastructure for a functioning state, while nation building the construction of a national identity, also for a functioning state.

[10] See Alesina and Spolaore (2003) for a review of the economic literature on country size.

[11] Alesina and Spolaore (2003) in their discussion mention this avenue of possible research but they do not develop it.

[12] For instance Michalopoulos and Papaioannou (2012) provide evidence of how national rule, institutions and policies in African countries do not reach isolated ethnicities far from the capital. These ethnicities revert to ethnic based rules, making the country unstable.

[13] One could also think of "private" forms of homogenization. For instance a linguistic minority setting up its own private schools to learn the dominant language, or isolated communities building private roads to be more connected to the rest of the country. We leave this point for future research.

overthrow (and the establishment of a democracy) are different again. As above, the ruler or rulers choose public goods to perfectly match their own preferences and tax the maximum number of people. However, this regime faces a significant possibility of being overthrown by a democratic movement and therefore the prospect of living under a democracy in the future. This democratic government may choose public goods that differ from the preferences of the ruler or rulers; in addition a democratic vote may break up the population into more than one country. In general democratic rule will not produce the most preferred policy of the ruling group. Homogenization, sometimes by brutal means, allows those in charge to better preserve the status quo (their preferred policies and a larger country). It follows that a higher threat of democracy induces more homogenization. In more colorful terms: the dictators will indoctrinate people in order to teach them to "enjoy" the current regime defined by the type of government (and potentially also ensure a larger country). We show that the most extreme episodes of homogenization will be undertaken by non-democratic regimes under threat. We discuss two forms of homogenization. A "benevolent" form in which the costs are equally distributed amongst the population and, what we term, "odious" homogenization, which concentrates the costs on "distant" minorities, i.e. on individuals very far from the government's preferences/location. We show how dictators will always chose odious means of homogenization. In the most extreme cases these forms of odious homogenization may simply imply extermination of minorities.

We then discuss cases in which the choice of homogenization of the ruler directly affects the probability of success of a revolution resulting in a democratic regime. In one case more homogenization, if it reduces distaste towards the government, may reduce the incentive to overthrow. In this case the ruler has an additional incentive to homogenize even more than in the case in which the probability of insurrection is exogenous. There is however another interesting case in which more homogenization may actually increase the probability of insurrection. A less heterogeneous population may communicate better, develop common goals and this may increase the likelihood of coordination in an insurrection attempt. In this case this effect works against the other incentives of the ruler to homogenize. It is a sort of "divide and rule" effect.

These results imply non obvious and "non linear" comparisons between public policies in democracies and non-democracies, an insight broadly consistent with Aghion, Persson and Rouzet (2012) and Mulligan, Gil and Sala-i-Martin (2004) as we show below. Safe dictators homogenize less than democracies, unsafe dictators more than democracies. In the final part, we discuss the theoretical results of the paper in light of specific cases and review some evidence regarding more or less successful policies of homogenization in democracies and non-democracies in recent years and in earlier centuries.

We are not aware of any formal model directly related to endogenous homogenization but our paper relates to the literature on the need for education in the better functioning of institutions, as in Glaeser, Ponzetto and Shleifer (2007) or Bourgignon and Verdier (2000). It also relates to the literature on "state capacity" as in Besley and Persson (2010, 2011) in the sense that for a state to be "capable" it needs a minimum level of homogeneity of its population. There is a non-formal literature on nationalism and nation building but,

to the best of our knowledge, there is no formal or informal theory of incentives to nation build across regimes, although the concurrence of franchise extension and nation building is frequently highlighted.[14]

This paper is organized as follows. Section 2 introduces the model. Section 3 examines the case of a democracy and Section 4 the case of non-democratic regimes facing various probabilities of overthrow. Section 5 compares different homogenization technologies and different types of regimes. Section 6 examines an endogenous probability of democracy. Section 7 examines the case of rulers who may exit the country should democracy prevail (e.g. colonizers). Section 8 discusses historical examples and the last section concludes.

2 The Model

Consider a population composed of a continuum of individuals of mass 1 distributed uniformly on the segment $[0, 1]$. This population forms a single country or splits into two equal-sized countries, A and B, comprising the intervals $[0, 0.5]$ and $(0.5, 1]$ respectively. We adopt the restriction of having at most two countries to keep the analysis simple while still allowing for endogenous country size. For simplicity we assume that the countries have equal size, but this implication can be easily derived using the same "stability" condition of indifference at the border, which delivers countries of equal size.[15][16]

Each country has a government which is located at some point $j \in [0, 1]$ inside that country. By "government" we mean a set of public goods and policies provided by an authority. Denote by $d_{ij} = |i - j|$ the distance of an individual $i \in [0, 1]$ from government j in his country. The parameter a measures the cost of this distance. We think of "distance" in terms of geography or preferences. If one wants to keep both interpretations one needs to assume that geographic location and preferences are perfectly correlated, as in Alesina and Spolaore (1997, 2003). More on this below.

The cost of public good(s) in a given country, funded by taxes, is k.[17] Since the costs k can be divided amongst all citizens in the country this captures the benefits of forming a single country rather than breaking into two.[18] However, when a population splits into two countries, A and B, the separate populations are more homogeneous and so the policies/public goods provided in those countries can be closer to the preferences of the median individual. Thus while a larger country reduces the costs of public goods, the more diverse a country is the harder it is for public choices to satisfy everyone; more precisely in larger country the average and median distance of a citizen from his government is higher. This setup captures a trade-off between size and heterogeneity.

[14] See Smith (1998) for a detailed description of and key references in the development of the study of nationalism and Laitin (2007) for a discussion of nationalism, homogenization and state formation.

[15] See Alesina and Spolaore (1997).

[16] We do not allow for unilateral secessions, namely a situation in which without any majority vote a group of citizens form a third country. See Alesina and Spolaore (2003) for a discussion of this case in a model without endogenous homogenization.

[17] Obviously the assumption of a fixed cost is extreme and adopted for simplicity of notation. It could be easily generalized to the case of $k = \alpha + s$ where s is the size of the country and α a fixed cost.

[18] Alesina, Spolaore and Wacziarg (2000) and Alesina and Spolaore (2003) investigate sources of benefits of size, like the dimension of the market and diversity of inputs in productivity.

We now depart from Alesina and Spolaore (1997) since we assume that the degree of preference heterogeneity and/or the heterogeneity costs, measured by the parameter a, are endogenous. We model homogenization as a technology that uses government apparatus to reduce the degree and/or cost of distance from the government. Specifically, diversity within a given country can be reduced by fraction λ, where $0 \leq \lambda \leq 1$, to $(1-\lambda)a$. So that for a country with government j and for any individual in that country, i, the difference between the policies of government j and i's ideal policies is reduced by fraction λ, from ad_{ij} to $(1-\lambda)ad_{ij}$. We refer to this as homogenization of the population. Homogenizing a population of size s by λ costs $sC(\lambda)$. We restrict the options of homogenizing such that any degree of homogenization, λ, must be applied to the whole population within a given country.

Assumption: *The cost of homogenizing, $C(\cdot)$, is strictly increasing, strictly convex and continuously differentiable as λ increases from 0 to 1. Further $C(0) = 0$, $\lim_{\lambda \to 0} C'(\lambda) = 0$ and $\lim_{\lambda \to 1} C'(\lambda) = \infty$.*

Individual i's utility, in a country of size $s \in \{1/2, 1\}$ with the government located at j, is:

$$U_i = g(1 - (1-\lambda)ad_{ij}) + y - t. \tag{1}$$

The first term measures the value of the government for individual i. We let g denote the maximum value of the government/public goods when the distance from it is zero. The term $1-(1-\lambda)ad_{ij}$ measures the benefit that individual i receives from the public good. The second term is income (consumption in a static model) y, and the third term is taxes t which are split equally amongst the population of the country.[19] The budget constraint is $k + sC(\lambda) \leq st$. Thus "homogenization", λ, increases welfare for the individual, although at a cost. One could justify the assumption of an equal splitting of cost with a "no secession constraint." Imagine that if more distant individuals were charged more for homogenization they might choose to unilaterally break away from the country. We relax the assumption on equal costs in Section 5; in particular we will allow for forms of homogenization in which the costs are unequally split and fall more heavily on those further away from the decision maker.

Let's be clear about what we mean by homogenization. The simplest way to think about it is building roads (or railroads or airports) to reduce travel costs, this improves access to resources of the government/capital. A second interpretation is to reduce the cost of communication in terms of language, written or spoken. Imagine that the further an individual is from the government the more different is his/her language. Reducing the distance in this case can be interpreted as teaching a common language (literally, reducing the distance between languages) so that individuals can better communicate with the government/capital. Neither of these two interpretations of homogenization imply a change in individuals' preferences. A third interpretation implies changing individual preferences by indoctrination (by more or less "kind" means). That is, it implies convincing individuals far from the type of government chosen that they do not dislike it that much. For instance one may argue that in schools, say

[19] See Bolton and Roland (1997) for a discussion about separatist movements due to income differences.

in France or Scandinavia, the benefits of regulation and social welfare are emphasized while in the US and the UK the merits of individualism are stressed more.[20] In communist countries indoctrination in schools of Marxist-Leninist ideas was common.[21] Changing preferences can also involve severe repression or elimination of groups with particular preferences (political or otherwise). One can choose the preferred interpretation of homogenization. In order to maintain all three together one needs to make the same assumption that geographic location, language and preferences are perfectly correlated. From now on with the term "distance" we summarize either one of the three interpretations above (or a combination of the three) and with the term "homogenization", a reduction in such distance.

In our model income is exogenous. However at least up to a point diversity of skills, education, backgrounds, and culture may increase productivity.[22] In this case a reduction in diversity would have costs and benefits. The latter are already modeled. The former would include not only the costs modeled above but also a reduction in productivity, therefore of income. Given that income/consumption enters linearly in the utility function and taxes are lump sum this reinterpretation of the costs and benefits of diversity would be immediate.

Decision-making proceeds in this order: 1) whether to form a single country or split into two; 2) where to locate the government; 3) to what extent to homogenize. What differentiates regimes is that these decisions are made by different agents, be it a dictator, elite group, or the population as a whole. This order is realistic since a "government" cannot be chosen before borders are set, and only an established government can choose public policies regarding homogenization. We study the case of a democracy first and then the case of a non-democratic regime.

Note that instead of homogenizing to reduce diversity, diverse countries could be kept together by transferring resources to the citizens further away in geography and utility from the government. We do not explore this issue here but note that once homogenization occurs it may last forever (say having a common language) while transfers may need to be paid every period and in the long run they may be more expensive for the center (i.e. the median voter in a democracy or a dictator in a non democratic state).

3 Democracy

In a democracy decisions are made by majority rule with the timing of votes described above. First let us examine the optimal level of homogenization for person i in a fixed country with a fixed government. Denote by λ_{ij} the optimal level of homogenization for individual i at distance d_{ij} from the government j in a country of size $s \in \{1/2, 1\}$:

$$\lambda_{ij} = \arg\max_{\lambda \in [0,1]} (g - (1-\lambda) g a d_{ij} + y - k/s - C(\lambda))$$

[20] See Alesina and Glaeser (2004) for a discussion of these cultural differences. See also Aspachs-Bracons et al. (2008) for a study of the effect of Catalan compulsory language education on identity.

[21] For instance, Alesina and Fuchs-Schündeln (2007) present evidence of a large amount of indoctrination in East Germany.

[22] On this point see Alesina, Spolaore and Wacziarg (2000) and Alesina, Harnoss and Rappoport (2013).

The first order condition is:

$$gad_{ij} = C'(\lambda_{ij})$$

which implies that the marginal cost of homogenization has to equal the marginal benefit. The latter depends on the distance of individual i from the government and the former on the cost of the homogenization technology. Individuals who are further from the government prefer more homogenization (higher λ). Note that we have assumed a technology that benefits those furthest out the most, while sharing the cost equally among the population (for example building roads to the capital where the cost is shared equally benefits those who live farther from the capital).

The intuition for this first order condition is immediate if we interpret homogenization in terms of roads or public schools teaching a common language. The "preference" interpretation of homogenization requires some thought. Literally speaking it implies that an individual "chooses" a policy that changes his preferences, knowing that after such change he/she would feel better in the country in which he lives, he would "fit" better. This interpretation becomes more plausible if we think of a dynamic extension in which parents transmit values and educate their children in such a way which makes them fit better in the country they live in by adopting certain social norms and types of behavior.[23] Strong attachment to cultural values can be interpreted as very high costs of homogenization.

Preferences over homogenization are single peaked within any given country. Thus the level of homogenization chosen by majority rule in a country with the government at j is the median λ_{ij} within the country. Further, for any given borders and degree of homogenization the location of the government that cannot be beaten by majority rule in pairwise voting is in the middle of the country, namely the most preferred location of the median voter. See Lemma 1 in the appendix for a proof.[24] Thus if the population forms a single country the government is located at $i = 0.5$ and the level of homogenization chosen is the optimal one for the individual at median distance from the central government, at distance 0.25.[25] Denote this level λ_1^m. If the population splits, the two governments would be located at 0.25 in country A and 0.75 in country B and the level of homogenization chosen when the population splits into two countries is the optimal level for the individual at median distance from a government, at distance 0.125.[26] Denote this level by λ_2^m. Note that $\lambda_1^m > \lambda_2^m$ since the marginal benefit of homogenization is higher the further an individual's ideal point is from a government and in a single country the median voter is further away from the center. In other words, more heterogeneous populations choose a higher level of λ.

Each individual evaluates whether he would be better off under a single country with a government in the center and homogenization level λ_1^m or under two countries, A or B, with

[23] For models related to parents "choosing" values for children see Alesina et al. (2013) and Bisin and Verdier (2000). Algan et al. (2012) discuss the cost of lack of assimilation of Arabs in France and their effort to do so. They document a substantial increase in salaries for children of families which signal assimilation by choosing French rather than Arab first names.

[24] The location of the government affects the location of the median voter when the amount of homogenization is chosen thus preferences are not necessarily single peaked nor necessarily satisfy single-crossing.

[25] There are two median voters, at $i = 0.25$ and $i = 0.75$.

[26] The median voters are at $i = 0.125$, $i = 0.375$, in A and $i = 0.625$, $i = 0.875$ in B.

a government at the center of both countries and each country homogenizing to the level λ_2^m. Denote by l_i the distance of individual i from the center of the population. The value of forming a single country relative to splitting into two countries for individual i is then:[27]

$$[g-(1-\lambda_1^m)gal_i+y-k-C(\lambda_1^m)]-[g-(1-\lambda_2^m)ga|0.25-l_i|+y-2k-C(\lambda_2^m)]$$

which can be rewritten as

$$(1-\lambda_2^m)ga|0.25-l_i|-(1-\ \lambda_1^m)gal_i+k-[C(\lambda_1^m)-C(\lambda_2^m)]\,. \tag{2}$$

Proposition 1 *The median voter is individual i at distance*

$$l_i=0.25\frac{(1-\lambda_1^m)+(1-\lambda_2^m)}{2(1-\lambda_2^m)} \tag{3}$$

from the center of the population. A democracy chooses to organize itself as one country, locate the government at the center, and homogenize to degree λ_1^m when expression (2) is positive as evaluated for the median voter. A democracy chooses to split into two countries, locate the governments at the centers, and homogenize to degree λ_2^m when expression (2) is negative as evaluated for the median voter.

The proof of this proposition is in the appendix. To break the tie, we assume that when indifferent between one country or two a democracy forms a single country.

Note that we have three decisions in this model with three different median voters. The decision of how much to homogenize, with median voters λ_1^m or λ_2^m; the decision of where to locate the government, with the median voter in the middle of the country or countries; and the decision of whether to have one or two countries, with the median voter given by Proposition 1.[28] Expression (2) encompasses three considerations for the median voter: how much closer he is to the government when the population splits into two countries versus when the population forms a single country (taking into account the homogenization in either case); the extra cost of running two countries; and the difference in costs of homogenization as a single country and homogenization when the population splits into two. *Ceteris paribus* the higher is k the cost of "government", the lower is a the cost of distance/diversity and more generally the benefit of being closer to the government, and the smaller the difference between $C(\lambda_{1,0.25})$ and $C(\lambda_{2,0.125})$, the extra cost of homogenization in a single versus two countries,

[27] This is symmetric for i either side of the center of the population.

[28] The median voter given in Proposition 1 is an individual $i \in (0.25, 0.375)$. In fact there are four individuals who have the same valuation of forming a single country or splitting into one as that expressed in Proposition 1. Individual $i = 0.5$ in the center of the population has the highest valuation of remaining as a single country. The individuals at the center of countries A and B have the lowest valuation of forming a single country. Individual $i = 0.25$ is at the center of country A, as i increases from $i = 0.25$ to $i = 0.5$ the individual's valuation of forming a single country relative to splitting into two is increasing linearly. Similarly as i decreases from $i = 0.25$ to $i = 0$ the individual's valuation is increasing linearly (but to a strictly lower valuation at $i = 0$ than at $i = 0.5$).

the more the median voter will prefer a single country. We will use the case of a democracy as a benchmark to compare against other regimes.

Intuitively, who tends to prefer a singe country versus two? Voters near the center of the population (0.5) prefer to form a single country relative to splitting into two (equation (2)) (and they also would vote for low homogenization since they do not need it as much as people further away). As you move towards 0.25 and 0.75 voters begin to value more splitting into two countries since they would be located near the governments in countries A and B. In a model without any homogenization voters at the extremes, say with bliss points lower than 0.25 and grater than 0.75 would all prefer two countries to one. However this is not necessarily the case with endogenous homogenization since $\lambda_1^m > \lambda_2^m$. If due to the nature of the cost function $C(\lambda)$ λ_1^m is substantially greater than λ_2^m, then it is possible that some voters at the extreme (close to 0 or 1) may prefer a single country with a very high λ_1^m to two countries with a relatively low λ_2^m. In other words minorities may prefer to be in a large very homogenized country than in two countries where they would still be far from the center and not very homogenized. Intuitively this cannot happen if the cost function $C(\lambda)$ is sufficiently convex, in this case the equilibrium level of λ_1^m cannot be high enough for this "reversal" to happen.

Note that it is perfectly possible that without the option of homogenization ($\lambda = 1$) this population would decide to split into two countries, but the option of choosing $\lambda \in [0, 1]$ would lead the population to homogenize somewhat and form a single country.[29] This is in a sense the interesting case as it captures the idea of "nation-building". This population would otherwise split, but will stay together with a technology of homogenization (road building, learning a common language), i.e. "nation-building".

4 Dictators

4.1 A safe dictator

Suppose now that this population is controlled by a single ruler (or dictator, terms used synonymously here) who makes all decisions. Modeling a dictator as a single agent (technically speaking of measure zero) may be unattractive but it can be easily generalized by allowing for an elite group to rule the population. The elite group can be represented by an interval of size δ.[30] Such an extension would complicate notation and algebra with little advantage in terms of insight. The ruler is located with his ideal point within the population, $i \in [0, 1]$, with utility given by (1). The ruler controls the entire population. Obviously:

Proposition 2 *The ruler locates the government at his ideal point and undertakes no homogenization.*

The ruler chooses a government that is ideal for himself. He undertakes no homogenization since he has no incentive to increase the welfare of the population by improving their access to the public good which perfectly matches his preferences/location. He is unconcerned with the heterogeneity of the population.

[29] See the online appendix $A1$ for a formal proof.

[30] The elite group makes decisions within the elite by majority rule. Results on this point are available from the authors.

The difference with respect to a democracy is threefold. The degree of homogenization is not chosen democratically. In fact $\lambda = 0$ with the dictator while $\lambda_1^m > 0$ in a democracy. Second, the government is located at the dictator's ideal rather than the median ideal point in the population. Third, the dictator controls the entire population while a democracy might choose to split.

4.2 An unsafe dictator

Most dictators however, are not absolutely safe in office and this may affect their incentives. In particular we now examine the possibility that a ruler may be ousted by a democratic movement. For the moment we take the probability of success of a democratic insurgency as exogenous. We make it endogenous below. The timing is as follows:

Period 1 The dictator rules over the entire population, and can choose where to locate the government and the level of homogenization, denoted λ^r. He knows that with probability p democracy will prevail in period 2.

Period 2 With probability p democracy prevails and the population as a whole now decides the location of borders and governments, as well as any further homogenization if desired; or democracy is not realized, the ruler stays in power.

Let's analyze intuitively the trade off faced by the ruler in period one when choosing the level of homogenization, λ^r. First note that if democracy prevails in period 2 and the country is split the dictator always looses for two reasons: generically the new governments of the two countries are not the ideal one for the dictator and taxes are higher since the countries have split.[31] Second even if the democratic regime does not split the country into two, the ruler generically does not get his ideal government anymore; thus he pays a price which is increasing in the distance of the position of the ruler from the center. The third and more subtle consideration is that in choosing λ^r the dictator has to take into consideration several effects. One is whether the choice of λ^r will prevent a split if a democracy would prevail. The second consideration is that both in the case of a split or not, the degree of homogenization chosen by the dictator will affect the final level chosen democratically. Depending on the location of the dictator relative to the new democratic government(s) the ruler will make different choices of λ^r. It is worth remembering that with the term "dictator" we mean a ruling group, which could be identified by an ethnicity, religion, geographical location, ideology etc. Thus when we say that for instance, that dictator would be "far" from the democratic government of a new country, we mean that the dictator's group would be a minority in the democracy and thus the group's preferred policies would be far from the democratically chosen ones.

Let's now formalize this problem. The utility of individual i in period 1 in a country of size s with the government at j when homogenization λ^r takes place is

$$U_{i1} = g(1 - (1 - \lambda^r)ad_{ij}) + y - t, \tag{4}$$

[31] Note that if the dictator happens to be located at 0.25 (or 0.75) the new government of one of the two countries is the dictator's ideal but still he would pay higher taxes than with a single country with the government at 0.25 or 0.75.

where $k + sC(\lambda^r) \leq st$, as above. The utility of individual i in period 2 in a country of size s with the government at j given homogenization λ^r has taken place in the population in period 1 and further homogenization λ takes place in period 2 is

$$U_{i2} = g(1 - (1 - \lambda^r - \lambda)ad_{ij}) + y - t. \tag{5}$$

where $k + s[C(\lambda^r + \lambda) - C(\lambda^r)] \leq st$. In period 1 the ruler maximizes the sum of his utility from period 1 and his expected utility from period 2 given probability p of overthrow. In period 2 whoever is in control simply maximizes their period 2 utility.

Period 2 is the final period and so the ruler, if in power, behaves exactly as a safe dictator. Thus he adds no additional homogenization. If instead democracy prevails in period 2 then the democracy votes as described above but now taking into account any homogenization undertaken by the dictator in period 1. We are interested in the homogenization the ruler will undertake in period 1.

Let's begin to gain intuition with the simplest case of a dictator located in the middle at 0.5.

Proposition 3 *In period* 1 *the ruler at* $i = 0.5$ *locates the government at his ideal point. He undertakes homogenization* λ^r *which is weakly increasing in the probability of democracy,* p.

A ruler at the center of the population faces one of two outcomes if democracy prevails: either the democracy chooses to form a single country, in which case the ruler is located at the median of the population, the democratic government will match his ideal; or the democracy splits into two countries A and B with the ruler located at the border of those countries. The outcome in which a democracy splits is particularly bad for the ruler since both taxes are higher and he no longer enjoys his preferred government (in fact he would be at the border of country A, i.e. he would be the furthest possible from the government). However, sufficient homogenization by the dictator can prevent the population from breaking up into two countries and ensure the dictator has his ideal government and the lowest possible taxes. That is, there is some minimum level of homogenization, denoted λ^*, at which the population is sufficiently homogenized that should democracy prevail that population would choose to form a single country. Homogenization is costly, but it is clear that in this case it will improve the utility of a ruler should democracy prevail. Homogenization to a level $\lambda < \lambda^*$ also benefits the ruler, but to a lesser extent. In this case, if democracy prevails, the population will split into two countries. Homogenization of the population still benefits the dictator because his ideal will be very far from the democratically chosen government; homogenization moves population preferences towards his own preferences and so moves the position of the future democratic government closer to his own ideal point (even though it will not match it). Clearly the higher the probability of democracy the more willing the ruler is to invest in costly homogenization. At a low probability of democracy he undertakes no homogenization since it is very likely that next period he continues to choose his ideal government.[32] For a high probability of

[32] Zero homogenization at a low probability of democracy is a result of the fact that should democracy prevail, the population will split and will choose to homogenize to λ_2^m. Therefore, when at low probabilities of democracy the ruler would optimally

democracy the ruler may choose to homogenize to λ^* and ensure a single country should democracy prevail. Of course, the exact homogenization chosen depends on the parameters and so if the cost of λ^* is particularly high the ruler will instead homogenize to $\lambda < \lambda^*$ when p is high.

Note that the ruler would never homogenize above λ^* even as p increases towards 1.[33] As we highlighted above, at $\lambda^r = \lambda^*$ this ruler gets his ideal location whether the country is non-democratic or democratic. As a citizen in a democratic single country he would prefer zero homogenization thus he has no incentives to choose any level of $\lambda > \lambda^*$.

Proposition 3 generalizes to a dictator located anywhere. As above, let λ^* denote the minimum amount of homogenization by the ruler which ensures the population is homogenized enough that should democracy prevail they would choose to form a single country. Provided the cost function is sufficiently convex, it follows that if the dictator homogenizes to any $\lambda \geq \lambda^*$ a democracy would choose to form a single country. We make this assumption from now on. The proof of Proposition 4 is in the appendix and includes a discussion of a sufficient but not necessary condition on the convexity of the function $C(\lambda)$.

Proposition 4 *In period* 1 *the ruler locates the government at his ideal point. He undertakes homogenization* $\lambda^r \geq 0$ *which is weakly increasing in the probability of democracy,* p.

This generalization is complicated by the fact that depending on the position of the dictator he has different preferences if a democracy prevails. Obviously the dictator would always prefer to stay in office, but his preferences under the democratic regime, both in terms of whether to maintain a single country or to split and how much homogenization to choose, depend on his location. For instance if the democracy does not split, the dictator may prefer more or less homogenization than the median voter. This consideration would affect how much homogenization he would choose as a dictator. These issues make the general proof much more laborious than in the case of the dictator in the middle and it is worked out in the appendix.

4.3 Comparing Regimes: preliminaries

Proposition 4 implies that the greater the threat of democracy faced by a non-democratic regime the more that regime homogenizes. But does a non-democracy homogenize more or less than a democracy would? The answer to this question is that a safe dictator will homogenize less than democracy while an unsafe non-democracy will homogenize more, but we build toward this answer in steps.

Clearly the fact that an "almost safe" dictator homogenizes less than a democracy holds since for low p a dictator does not homogenize while a democracy would. The more difficult comparison is between a democracy and very unsafe ruler, namely facing a high p. Let us give

homogenize to less than λ_2^m, he undertakes zero homogenization since homogenization λ_2^m will be undertaken anyway should democracy prevail.

[33] Note that when $\lambda^* = 0$ a democracy would choose a single country the dictator would do no homogenization.

some examples of rulers at particular locations. We come back to the case of a ruler in the center of the population at $i = 0.5$. We saw above that the outcome in which a democracy splits is particularly bad for this ruler since both taxes are higher and he no longer enjoys his preferred government, in fact he would be at the border in country A. When the probability of democracy is high enough this ruler will undertake more homogenization than a democracy even up to λ^*. Take the case where homogenization λ^* is too costly for the dictator. If democracy prevails the population would split and the ruler would be at the border (in the minority) in the new country. It follows that, when in power, he wants to homogenize the population more than the median voter would to ensure this future democratic government is closer to his preferences. If λ^* is not too costly then when p is high he homogenizes to λ^* and ensures that instead of splitting a democracy would form large country with a government which matches his preferences.

Take the case of a dictator located close to the extremes of the population (close to 0 or 1). We can think of this as an ethnic minority which controls a large diverse country. This dictator would lose a lot by democratization since he would have minority status in the new country, whether the population stayed as one or split. His optimal homogenization would be much higher than the median voter in a democracy in either case. In this case when the probability of democracy is high the ruler always undertakes strictly more homogenization than a democracy.

However even when p is high there exist dictators at certain locations that undertake zero homogenization. The general intuition for this result is that if the democratic government would be close to the dictator's ideal anyway then he undertakes no homogenization since homogenization is costly and the democratic outcome will be sufficiently 'close' to his preferences anyway. In these instances Proposition 4 holds trivially. Suppose a dictatorial minority was located very close to the government of country A. This dictatorial minority prefers a dictatorship since they can set the government at their preferred position and extract taxes from the whole population. Suppose if democracy prevails the democracy would split. If homogenization were costless the dictator would completely homogenize the whole population and should democracy prevail he would get a single country with his ideal government. Instead, since homogenization is costly, the dictator undertakes no homogenization and so if democracy prevails the dictator will pay more taxes but the government of country A will represent his ideal. The reason he undertakes no homogenization is that he faces the same trade-off as everyone else between size and heterogeneity, as well as the same costs of reducing heterogeneity. That is, since the median individual in a democracy prefers a split, the dictator located at the center of country A must prefer a split even more, given the same parameters and the same costs of homogenization. In other words, the ruler cannot obtain his first best in an efficient way and so he opts for second best which in this case is a smaller country but still his ideal government. He does not need to homogenize at all to obtain this second best.

Note that this rather extreme set of assumptions do not allow the dictator to choose whom to make pay for the costs of homogenization, the dictator has no pure rents from office, nor he can affect the probability of a democracy to occur. In fact, up to here, we purposefully make what we consider to be the weakest assumptions in order to highlight the robustness

of the result that the most extreme episodes of nation building will be observed under non-democratic regimes under threat. We now move to a more realistic case in which a dictator can unequally distribute the costs of homogenization

5 Odious Homogenization: comparing regimes

Thus far we modeled homogenization as a technology such that the cost is spread equally across the population; let's label it "non-odious" homogenization. In contrast, "odious" homogenization implies a distribution of costs that falls more heavily on those who are further away from the ruling government. The repression of cultures that are different from the leading one would likely fall into the category of odious homogenization: those who face the greatest costs are those whose culture and language are different from the government. Learning the language imposed by the government is more difficult for those with a very different language spoken at home. The cost of roads may be allocated unevenly on those further away form the capital.

We label the degree of odious homogenization with μ where $1 \geq \mu \geq 0$. The utility to individual i at distance d_{ij} from a government located at j in a country of size s following odious homogenization μ is

$$U_i = g(1 - (1 - \mu)ad_{ij}) + y - \frac{k}{s} - M(\mu, d_{ij}).$$

where $M(\mu, d_{ij})$ is the cost of homogenizing by μ for the individual at distance d_{ij} from the government in a country of size s. As before, the cost of homogenizing, $M(\mu, d_{ij})$ is strictly increasing, strictly convex and continuously differentiable in the level of homogenization, $M(0, d_{ij}) = 0$, $\lim_{\mu \to 0} M_\mu(\mu, d_{ij}) = 0$ and $\lim_{\mu \to 1} M_\mu(\mu, d_{ij}) = \infty$. Note that an individual at distance d_{ij} from the government is "moved" towards the government in the amount $\mu a d_{ij}$. Thus the only difference here is we assume the cost of homogenization, $M(\mu, d_{ij})$, is increasing in d_{ij}, the distance of the individual from the government, at some rate $\alpha(\mu)$. The cost of homogenization is higher for those who are homogenized by more. We also assume the marginal cost of homogenization, $M_\mu(\mu, d_{ij})$, is increasing in distance from the government. That is, the marginal cost of homogenization is higher for those who are homogenized by more. The first order condition for individual i is now

$$g a d_{ij} = M_\mu(\mu, d_{ij}).$$

where the left-hand side and right-hand side are both increasing in the distance of individual i from government j.

To make comparisons between odious and non-odious homogenization we assume the total cost of homogenizing a population to a given degree is the same under both technologies, that

is

$$\int_{i \in country} C(\lambda) di = \int_{i \in country} M(\mu, d_{ij}) di$$

when the government is located in the center of the country.[34] Clearly this may not hold, but it is useful for comparisons. As above, we assume that a dictator or population chooses the homogenization technology followed by the amount of homogenization after borders and governments have been determined.

Proposition 5 *A dictator will always undertake odious homogenization. The median voter in a democracy weakly prefers non-odious homogenization.*

See the appendix for a proof. The intuition is simple: "odious homogenization" costs less to the dictator. The burden of homogenization shifts towards the rest of the population, at an increasing rate the more distant are individuals from the dictator himself. Depending on the location of the democratic government, a democracy is either indifferent between odious and non-odious homogenization or strictly prefers non-odious methods. We thus assume from now on that a democracy always chooses non-odious homogenization, and the result that the government will be located at the center and the value of λ^* follows as before. In general, majority rule would allocate the costs of homogenization in the way preferred by the median voter. Allowing any type of distribution of costs, any level of homogenization and any government location would make the problem intractable. Our modeling device is meant to capture the fact that a dictator can choose forms of homogenization which are advantageous to himself, regarding his location and regarding of the location of the median voter in a democracy. Thus in general, a dictator has more latitude in the allocation of costs and when in office he would take advantage of this.[35] The following proposition establishes a comparison between the degree and type of homogenization in a democracy and in a dictatorship.

Proposition 6 *(i). The amount of odious homogenization, μ^r, undertaken by a dictator is weakly increasing in the probability of democracy, p;*

(ii). A safe dictator, $p < p'$, undertakes less homogenization than a democracy;

(iii). An unsafe dictator, $p \geq p'$, undertakes at least as much homogenization as a democracy and in some cases strictly more;

where p' depends on the parameters, costs, and the location of the dictator.

The proof is in the appendix. At low probabilities of overthrow rulers undertake less homogenization than a democracy while at high probabilities rulers undertake more homogenization than a democracy. An unsafe dictator undertakes at least as much homogenization as a democracy. In fact when democracy prevails a democracy undertakes homogenization

[34] The online appendix *A2* gives an example of such a cost function and highlights details of this type of cost function.

[35] A limit on what the dictator can do in terms of allocation of cost is related to the possibility of unilateral secession of regions, or insurgencies of specific groups. This extension is left for future research.

of λ_2^m or λ_1^m. This is essentially a transfer from the center to the periphery since non-odious homogenization particularly benefits people distant from the center. In contrast a ruler can just force minorities to assimilate by odious homogenization which is cheaper for him since it taxes those further away from him by more. An unsafe ruler undertakes strictly more homogenization given the incentives which we discussed above in the previous section.

Corollary 1 *Suppose that the parameter values are such that a democracy would choose to split into two countries. An identical population controlled by a ruler, who faces a high threat of democracy, may be homogenized to such an extent that, should democracy prevail in the future, the population would choose to form a single country.*

Corollary 1 implies that two initially identical populations may both end up as democracies, but the population that has been controlled by the unsafe non-democratic regime may be homogenized by more, perhaps so much so that it is homogenous enough to form a single country. The implication is that as a result of "nation-building" by dictators, today's democracies, which followed hundreds of years of dictatorship and control by elites, may be more homogenous and bigger than they would otherwise be had democracies prevailed earlier on. Obviously if a dictator does not homogenize enough then we would observe the splitting of countries during the process of democratization, as for instance we observed after the collapse of the Soviet Union.

6 Endogenous democratization

6.1 Welfare

Overthrow may be less likely when the preferences of the population are close to that of the ruler. Suppose that the probability of overthrowing a ruler in the second period depends negatively on population welfare in that period. In this case a ruler may wish to homogenize in the first period to reduce the "distance" between himself and the population. For example, repressing or eliminating political, ethnic or other groups that are different from the dictator or elite may reduce the extent of opposition to the government and so reduce the probability of overthrow; indoctrination in schools may reduce distaste for the government. The probability of overthrow can be written as a differentiable function of homogenization in period 1, $p(\lambda^r)$, which is strictly decreasing in λ^r. Proposition 7 compares the case where the probability of overthrow depends negatively on the amount of homogenization undertaken with the case where the probability of overthrow is exogenous, for the same probability of democracy. We compare non-odious homogenization in both instances.

Proposition 7 *When the probability of overthrow depends negatively on the amount of homogenization, weakly more homogenization is undertaken relative to the case of an exogenous p.*

The proof follows a similar argument to Proposition 4 and is solved in the online appendix. Homogenization now both reduces the probability of overthrow and improves the situation of

the dictator should democracy prevail, thus the optimal amount of homogenization increases compared to the exogenous case. As in the exogenous case, at low probabilities of overthrow less homogenization is undertaken by a dictator relative to a democracy (comparing identical populations under these two different regimes) while at a high probability of overthrow more homogenization is undertaken by a dictator relative to a democracy.

When we make the probability of overthrow endogenous to homogenization, rent extraction becomes relevant to choice of homogenization. Suppose there is an additional fixed positive value to being a dictator, call it R, that we might think of as rent. This further increases incentives to homogenize in order to reduce the probability of overthrow and subsequent loss of rent, R.[36]

6.2 Divide and Rule

In some cases revolutions become more likely and more successful when a population is homogeneous since the population can communicate better and may have more similar goals. After all, the principle of "divide and rule" is meant to capture precisely this effect. Similarly, if homogenization involves education of the population, a more educated population could also increase the probability of successful overthrow. This case can be analyzed with our model.

Suppose now that the dictator, when choosing the optimal amount of homogenization λ^r, has to take into account the effect of the latter on the probability of insurgency. That is, the probability of overthrow can be written as a differentiable function of homogenization in period 1, $p(\lambda^r)$, which is increasing in λ^r. Proposition 8 compares the amount of (non-odious) homogenization undertaken in the case where the probability of overthrow increases in the amount of homogenization with the case where the probability of overthrow is exogenous, for the same probability of democracy (and non-odious homogenization). The proof follows a similar argument to Proposition 4 and is solved in the online appendix:

Proposition 8 *When the probability of overthrow depends positively on the amount of homogenization, weakly less homogenization is undertaken relative to the case of an exogenous p.*

As above, rent extraction is now relevant to the amount of homogenization chosen by the ruler. The possibility of democratization and loss of rent further reduces the incentive of the ruler to homogenize.

7 Colonizers

Thus far we have assumed that the dictator stays in the country (or countries) after the democratization. This case would not apply to a colonizer, who would simply leave the country and return home if overthrown. If the probability of democracy (and decolonization) is exogenous, the colonizer does not homogenize. In fact he does not care about the future

[36] Consider as an example the case of a dictator located at 0.5 and choosing λ^*. With an exogenous p he would not homogenize above that level but with rents he may homogenize more as an increasing function of p in order to increase the chances of staying in office.

of the country (since he will not be there) nor whether the country splits or where the new government(s) would be located.

Obviously his incentives would be different if by choosing certain policies toward nation building the colonizer could change the probability of decolonization. Probably this is a case in which the incentives to "divide and rule" are especially strong. By creating a sense of "national unity", the colonizers may indeed increase the incentive towards independence movements. In fact an interesting extension of this model would be to allow "negative nation building" namely a process in which a colonizer plays one ethnic group against the other in order to maintain power by weakening the national movement for independence.

8 Historical Examples

In this section we discuss the history of nation-building and examine some suggestive empirical evidence in light of our results.

8.1 Pre-industrial and industrializing Europe: Safe and Unsafe Elites

Proposition 2 shows that "safe" rulers not subjected to threats do not need to use resources to homogenize their populations. Rulers who face a threat of democracy, as shown in Proposition 4, undertake homogenization, the more so the greater the threat of democracy. This result is consistent with a much documented period of homogenization in European countries starting in the 19th Century. Hobsbawn argues that 'in the last third of the nineteenth Century it became increasingly manifest that the democratization or at least the unlimited decolonization of politics, were unavoidable.' Indeed the 19th Century is exactly the period of increasing intervention by states in promoting and enforcing the use of a national language, supporting public schooling and sponsoring many other aspects of national identity.[37]

8.1.1 France

Eugen Weber estimates that 'French was a foreign language for a substantial number of Frenchmen, including almost half the children who would reach adulthood in the last quarter of the [19th] Century.' The Ancien Régime had done little or nothing to create and enforce a national language. Although this regime aimed to centralize administration and imposed French at the highest administrative level,[38] there was little if any effort to foster a nation of French-speakers and 'the French Crown [showed] little concern with the linguistic conquest of the regions under its administration.'[39] In fact the ruling elites made a point of distinguishing themselves from the masses, making communication more difficult using language as a barrier (Gellner, 1983). Only after the Revolution did this policy change. Weber claims that the French 'had no uniform concept of patriotism at the Revolution...and that patriotic feelings

[37] Laitin (2007) emphasizes the importance of language in nation building as a crucial coordinating device. He also illustrates how sometimes a language not spoken by the majority became the national language.

[38] The Ordinance of Villers-Cotterêts, made law in 1539, was designed to end the use of Latin in official documents and replace it with French.

[39] Weber, 1979.

on the national level, far from instinctive, had to be learned.' There was also little interest in increasing geographic communications in France. Roads were just a means of collecting taxes and transporting troops. Only after the Revolution, in the first half of the 19th Century, was the existing network improved. According to our model, the relatively safe Kings of the Ancien Régime (at least up to the few decades before the Revolution) should not have had much interest in spending resources on homogenizing the French population, instead they would have invested to extract rents.

After the French Revolution, and increasingly throughout the 19th Century (following further major uprisings), it became clear that more and more power would be transferred from the elites to the French population as a whole. According to our model, with a greater threat of democracy we should see more efforts by French elites to homogenize the French population in order to unify the country. The region of Brittany provides a stark example. An important report on the Breton departments in the 1880s noted that 'Brittany, which was not willingly joined to France, which never wholeheartedly accepted its annexation, which still protests' had still to be merged into the nation. The report therefore urged 'Frenchify Brittany as promptly as possible...; integrate western Brittany with the rest of France'.[40]

Schooling was one way to homogenize and advance a national language. Indeed immediately after the Revolution, the Constitution of 1791 called for the establishment of free public instruction for all and the Convention (the legislative assembly from September 1792 to October 1795) decreed that in the Republic children should learn to "speak, read and write in the French language" and that "instruction should take place only in French."[41] The first serious attempt at mass schooling was made in 1833 following a period of major rebellion (the 'July Revolution', $1830-32$). In France, as elsewhere in Europe, the emergence of state intervention in schooling does not appear to have been a concession to a more demanding population; state provided schooling was, at least into the last quarter of the nineteenth Century, largely unpopular (Katznelson and Weir (1985) $p42-43$; Weber (1979) $p318-322$). Weber suggests that such policies increased in intensity with the onset of the Third Republic which was established in 1870 alongside large scale violent unrest. Efforts to promote French also involved the suppression of other languages: as late as 1890 a ministerial decree banned religious instruction in Flemmish and in 1902 the government banned Breton language sermons. The Minister of the Interior in 1891 argued that preaching in local dialects 'may endanger French unity.'

8.1.2 Italy

Italian unification was completed in the 1860s accompanied by an increase in pressure for more democracy.[42] Italy, once unified, constituted a diverse population speaking a range of languages as well as dialects very different from each other (only 10% of the population spoke Italian). Building unity and commonality was vital to the stability of the new nation.

[40] Report by the rector of the Academy of Rennes, Boudoin, who suggested Frenchification would only be possible through schooling, Weber (1979), p100.

[41] Ramirez and Boli, 1987; Weber, 1979.

[42] The largest proportion of adult males were enfranchised in Italy in 1912.

As Duggan (2007) documents, 'during the 1860s the government had embarked on extensive discussions about what form of Italian should be adopted as the national language. There was strong feeling that linguistic centralisation was needed to complement political unity.' Tuscan was chosen. As in France, linguistic homogenization was to be achieved mainly through schooling and, despite the frequent lack of popularity within the population, 'the official line remained that Italian should as far as possible be enforced, with Italian texts being used in schools and dialect literature (of which there was a distinguished tradition in many regions) being discouraged.'

The nation building component in the motivation to introduce compulsory schooling appears, from debates and discussions of the time, to have been particularly strong. While a degree of literacy and education was desired for the wider population, there was also a fear across European countries that too much education could pose a threat. Michele Coppino, the author of the 1877 Italian compulsory education reform declared that primary schooling should ensure the masses were 'content to remain in the condition that nature had assigned them' and that the aim of elementary education should be to 'create a population ...devoted to the fatherland and the king.'[43] The imminence of democratization is apparent as a driver for these reforms in comments by Francesco Crispi, the Italian Prime Minister from $1887-1891$ and $1893-1896$: 'I do not know if we should feel regret at having broadened the popular suffrage before having educated the masses.' Politician Nicola Marselli claimed that Italy had introduced freedom before educating the masses, omitting to learn lessons from countries like Britain which had educated first.[44]

In Italy active homogenizing policies included large investment in railroads. Quite apart for their role as infrastructure they had the political goal of unifying, especially the diverse Northern and Southern parts of the country including isolated areas in the less developed South. The Minister of public works was viewed as the man who was building Italy as a nation state (Schram (1997)). Military service helped as well. Conscripts were specifically sent to regions away from home and regiments purposefully formed of soldiers from diverse parts of the country, in part with the aim to homogenize.[45] The goal of spreading the ideology of those who established the new nation is clear, Giuseppe Guerzoni, a friend of Garibaldi, explained at a conference in 1879 that 'having made Italy the army is making Italians.'[46] Nicola Marselli expressed in 1871 'I know, too, that Italy has been reunited for only ten years and is not yet established [...] I have always said that even if it had no other purpose, the army would always be a great school of Italian-ness.'[47]

8.1.3 England

The establishment of English public education also coincided with a greater threat of increasing democracy. The first appearance of public education came in 1833 following three years

[43] Duggan (2007), p280.
[44] Duggan (2007), p289.
[45] Duggan (2007), p288.
[46] Duggan (2007), p283.
[47] Duggan (2007), p274.

of widespread rioting in rural England. With further political reform in the 1860's the 'full democratization of the political realm was seen as inevitable'.[48] The link between education reform and emerging democracy is apparent in speeches of the time. Robert Lowe, a British politician (later Home Secretary and Chancellor of the Exchequer) highlighted in an address in 1867 the urgency for education reform following the 1867 Reform Act:[49] *'we cannot suffer any large number of our citizens, now that they have obtained the right of influencing the destinies of the country, to remain uneducated [...] it is a question of self preservation- it is a question of existence, even of the existence of our Constitution'*[50] In 1870 when W.E. Forster put forward the bill for his education act in Parliament his speech included the following: *'Upon this speedy provision [of elementary education] depends also, I fully believe, the good the safe working of our constitutional system. To its honour, Parliament has lately decided that England shall in future be governed by popular government [...] now that we have given [the people] political power we must not wait any longer to give them education.'*[51] In 1870 the Elementary Education Act was introduced mandating the provision of schooling.[52] Historian, Linda Colley (1986), argues that to Britain's governors 'nationalism was like Pandora's box: something which was best left alone.' Colley argues that 'dividing and ruling seemed a more attractive strategy than state-sponsored nationalism'. The fear that nationalism might increase demands by the population meant that nation building policies were enacted late in Britain: 'Only after the 1870s did Britain's governing elite commit itself to a patriotic, blatantly nationalist appeal. Not accidentally this coincided with a massive extension of the suffrage and the introduction of compulsory public education.'

8.1.4 Riots and education reforms in Europe

We have argued above that primary education reforms were a way of building the nations of Italy, France and England, especially under democratic threat. We now examine fourteen European countries between 1800 and 1875.[53] The binary variable, *education reform*, takes the value 1 if any major educational reform takes place in that country and year. Data on reforms are from Flora (1983) and include introductions and extensions of compulsory education as well as major events and laws (e.g. bringing education under state control, major introductions and changes in types of school and curricula). We do not include those relating only to university education since they are considered irrelevant to mass homogenization.

A measure of the perceived threat of democracy is undoubtedly difficult, nevertheless there is an obvious possibility. The French Revolution and recurring major uprisings in France that followed during the 19th Century had a significant impact both on France and on other European countries in two ways. First, these uprisings scared ruling elites with the prospect that populations could and were willing to overthrow the existing order. Second, they arguably

[48] Ramirez and Boli (1987).
[49] The act enfranchised a part of the male urban working-class population.
[50] From Marcham, (1973).
[51] Young and Handcock (1964).
[52] Ramirez and Boli (1987).
[53] The countries are Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Norway, Netherlands, Sweden, Switzerland, United Kingdom.

sparked uprisings in other countries.[54] In one or both of these ways, uprisings in France increased the threat of democracy as perceived by elites in other European countries. Our variable, *riot*, takes a value of 1 if a major uprising occurred in a given year in France. We use uprisings compiled by Tilly et al. (1980) between 1830 and 1875.[55] The dates recorded refer to a list of major episodes of collective violence involving a large number of people engaging 'in seizing and damaging persons or property' across a range of locations.[56] For example the year 1831 involves the continuation of smaller disturbances from the July Revolution of 1830, multiple violent demonstrations in a number of big cities including Paris, and a silk workers' insurrection in Lyon. We assume education reforms may not always follow immediately and so examine a binary variable for the years of rioting and following year, i.e. $riot(t, t-1)$ is a binary variable which takes a value of 1 if there was revolutionary activity that year or in the previous year.

Figure 1 shows that education reforms across Europe were largely concentrated in periods of insurrections in France, a proxy for democratic threats in Europe in this period.

An alternative explanation of the positive association between revolutionary activity and education reform is that rioters demanded public education and the latter was a concession under duress on the part of the rulers. However, was more education what the rioters demanded? The answer is 'probably not.' As noted previously, state-run mandatory schooling appears to have been unpopular and actually opposed by peasantry for much of the 19th Century in France. Ramirez and Boli (1987) document that in Sweden, around 1810, 'bourgeois liberals led a movement to develop mass schooling to provide national unity and purpose,' but it was primarily resistance by the peasantry that slowed the adoption of state-controlled education until 1842. In England violent and non-violent protest spread across the country in the first years of the 1830s. The Royal Commission into the Poor Laws in 1834 that was set up in part in response to this unrest asked the following question: 'Can you give the commissioners any information respecting the causes and consequences of the agricultural riots and burnings of 1830 and 1831?' The respondents were parishes and 526 replied. The only cause cited by more than 30 parishes were labor concerns (unemployment, wages and mechanization of jobs that previously provided employment), subsidies for the poor (poor law) and beer shops (where it is believed many of the protests were organized). Not a single response considered demand for education (or anything related to education) as a cause of the unrest (Holland, 2005). Similarly Charles Tilly's (1998) detailed study of episodes of collective disturbances in France $1830-1860$ provides information on the objective of the group involved in

[54] Examples from a range of countries include the following. A period of reform swept Sweden in the 1830s inspired by the uprisings in France (Ramirez and Boli, 1987). Reactionary politics 'swept Austria in the aftermath of the French Revolution' (Ramirez and Boli, 1987). The two major concentrations of violence in Germany in the 19th Century 'followed closely upon increased turbulence in neighboring France' Tilly et al. (1980) p209 (see also p247). In England, there was a feeling that events in other European countries could impact unrest in England (Holland, 2005). Hobsawn and Rudé (1969) observe, regarding the English riots in the early 1830s, 'it is doubtful whether it would have occurred on so vast a scale when it did, without the...French and Belgian revolutions abroad' p$xxiv$, $62-64$.

[55] Years of revolution: July Revolution 1830; February Revolution 1948; revolution $1870-1871$. Years of major uprisings which are not considered by Tilly to be revolutions: 1831, 1832, 1834, 1839, 1840, 1841, 1846, 1847, 1849, 1850, 1851, 1869.

[56] These dates correspond with other data in Tilly et al. (1980) describing different measures of collective violence including number of violent events, participants in collective violence and arrests in collective violence across France.

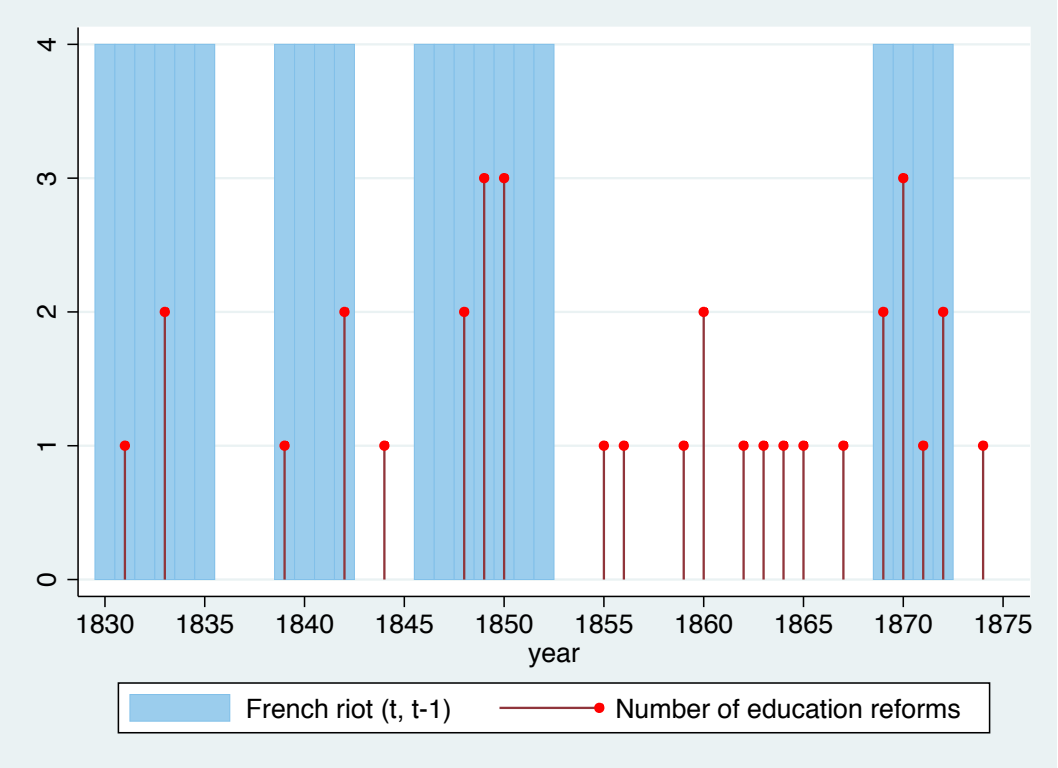

Figure 1: Years of major rioting in France and number of education reforms across Europe

the disturbance. Education is not mentioned.[57] Indeed this is consistent with evidence from modern day Brazil: Bursztyn (2012) shows that the poor prefer cash transfers to subsides for education and that their assessment of the government is negatively affected when they perceive government funding for public education to have increased but cash transfers to have decreased.

Certainly debates and discussion of political elites focused on the need for education as a means not only to educate but to nation build. If education in the 19th Century is provided with a nation-building motive then we might expect differences in the implementation of education compared with other policies such as social security or health care which are clearly redistributive. The earliest European, non-voluntary, government insurance system was introduced in 1883 the first voluntary system in 1871.[58] In contrast most countries had compulsory universal education by the time welfare reforms were introduced and in some countries it was highly developed (e.g. France).

[57] Tilly (1998).

[58] Our historical observations appear consistent with the historical discussion in Acemoglu and Robinson, (2000), on the extension of the franchise. They suggest that, in a number of cases, redistributive concessions were not credible and franchise extension was required by the elite to avoid costly overthrow. Welfare reform may then follow franchise extension (Germany being an exception, for example).

8.2 The Ottoman Empire and Turkey

Into the 19th Century the rulers of the Ottoman Empire behaved like a "safe dictator" in our model, tolerating pluralism of language, religion and culture. This is illustrated by the 'millet system' which granted non-Muslim subjects autonomy over their own communities in a number of aspects including religion, administration of law, education and taxes. As the Ottoman Empire lost more territory from external pressure and through secession, particularly towards the end of the 19th Century, homogenization policies became widespread. Perhaps it is the complexity of the Ottoman Empire that meant that a variety of nation building ideologies were proposed from the 19th Century onward. Zürcher (2010) highlights that the Ottoman constitution of 1876 had as its ideology Ottomanism, the idea that all the peoples of the empire would form a single Ottoman people if they were given full equality before the law. Two other ideologies circulated over the decades including Islamism (religious homogeneity) and Turkism (based on language, ethnicity and culture).

Nation building policies generally appeared later than in Europe. Up to 1909, instead of using the army to homogenize, Jews and Christians could pay exemptions to avoid conscription. After the constitutional revolution in 1908 by the Young Turks, military service became obligatory for all citizens of all religions. Zürcher argues that 'identity formation' in the period from 1908 'had been defined by the opposition between Muslims and non-Muslims'. In 1914, $150,000$ Greeks were expelled; while in $1915-1916$ the Armenian population of Anatolia was forcefully moved to the Syrian desert alongside massacres and death marches which killed an estimated $800,000$ Armenians.[59] The 'National Economy' programme started in 1914 favored the founding of businesses and the accumulation of capital by Muslim citizens. Political favoritism via, for example, infrastructure projects helped this aim. The financial loss to Greek and Armenian business owners from the 'National Economy' programme also encouraged their emigration.

The first years of the Atatürk rule would be considered an 'unsafe' non-democracy in the terminology of our model. Zürcher writes although a democratic parliamentarian system was in place 'emergency powers gave the government a relatively free hand from 1925 to 1929 and 1939 to 1946.' Homogenizing practices continued after the founding of the Republic in 1923, by which time due to previous government policies, as well as migration as a result of the break up of the empire, the Muslim population had dropped to 2.5% of the population from 19% in 1914.[60] The focus moved to a 'Turkish' nation. This is clear in the 1927 People's Party programme which asserted spreading the Turkish language and culture to be a guiding principle because 'among compatriots unity of language, of feelings and thoughts form the strongest tie'.[61] Zürcher writes that 'This identity was then imposed gradually on the population through a process of nation building in which, as in similar processes the world over, historiography and linguistics played a key role, as did suppression of alternative or sub-identities.' This author estimates that 'the adoption of Turkish nationalism led to the forced assimilation of the 30 per cent or so of the population which did not have Turkish

[59] Zürcher (2010).
[60] Içduygu et al. (2008), it was estimated at 19.1% in 1914 and 40% in the 1820s.
[61] Zürcher (2010).

as its mother tongue.' In 1934, a law on settlement 'closed strategic regions of the country to non-Muslim minority settlement' and aimed to encourage assimilation through strategic settlement. (Içduygu et al., 2008). In 1942 a tax was levied on all Turkish citizens to help pay for the Second World War. The tax was generally higher for non-Muslim citizens and the law resulted in a large transfer of property from the non-Muslim to Muslim population.[62] Içduygu et al. (2008) argue that such policies also increased emigration of the non-Muslim population out of Turkey.

8.2.1 Colonization and Decolonization

Our model predicts that rulers who leave a country after they are overthrown, like colonizers, do not undertake homogenization. Colonizers of Africa did not make any effort to build cohesive nation states (see Easterly and Levine, 1997; Herbst, 2000; Alesina, Easterly and Matuszeski, 2010; and Michaolopoulos and Papaioannou, 2012, amongst others). Colonizers mixed ethnicities in ways that created unstable and often failed states when they left. Gennaioli and Rainer (2008) show that the lack of nation building in many countries in Africa had long lasting effects after decolonization with reversal to tribal based institutions.

Colonizers also used active policies of 'divide and rule', (see Christopher, 1988). Gellner (1983) notes, European colonizers were not at all inclusive of local populations in administrative and bureaucratic roles keeping themselves fully socially separate from local populations. They, as rulers, had no interest in homogenizing and building a national identity since they knew that they were there just to extract rents, especially in Africa, and, should insurrections prevail, they would leave and so had no interest in the future of their former colonies. Our model predicts that after decolonization, whether as a democracy or 'unsafe' non-democratic regime, newly independent countries would enact nation building policies. Indeed after decolonization in Africa and Asia, many leaders of the newly independent countries attempted homogenization policies to unify their populations, more or less successfully.[63]

Zambia, a British colony from the 19th Century to independence in 1964, adheres to this pattern. Colonization was 'a take-the-money-and-run affair' with relatively small welfare spending and education mainly provided by missionaries.[64] Colonization exacerbated differences among the Zambian population (Marten and Kula, 2008, on language; Phiri, 2006, on regional divisions). On independence a multitude of languages were spoken with English existing as the main language of commerce and administration. Kenneth Kuanda, the first president of Zambia, claimed that although nationalism had led to independence, national identity in Zambia was completely lacking. In politics and even within Kuanda's party, the UNIP, ethnic and regional interests were significant. Phiri (2006) writes that 'Zambia's experience in the first eight years of independence is a typical example of how most newly independent African countries grappled with the need to create a sense of national identity.' In this period the national motto 'One Zambia, One Nation' was adopted and English became the official language. Marten and Kula (2008) claim that the decision to make English the

[62] See Içduygu et al. (2008).
[63] Smith (2003).
[64] Marten and Kula (2008).

common language 'was seen as the only non-tribal alternative available to serve as a vehicle of national unity, an argument often made in post-colonial African language policies.'[65]

Countries that moved more gradually through revolutions and other petitions towards modern democracy (arguably Europe) may be larger and more homogenous today than countries who faced a different path. In particular, countries that moved straight from colonization to democracy may be less homogenous. Several fragile states in Africa are an example. We highlight also the case of India. At the end of colonization by Britain, a multitude of languages were spoken in India. As our model predicts the British had done little if anything to homogenize a diverse population, even using specific policies of divide and rule (Christopher, 1988). India then moved straight away to form a democracy. Hobsbawn claims that on Independence, the multitude of languages spoken in India made the creation of a single national language impossible since many were unwilling to accept the disadvantage of having a national language that was not their mother tongue. While Hindi was the most widely spoken language in India and was Ghandi's choice for a national language, those advocating Hindi as a national language were unable to impose it on the population as a whole. English became the 'median of national communication,' while also maintaining state level languages, because this 'was least unacceptable to Indians' as it gave no one language group a particular advantage. As well as language, the Indian National Congress was also 'committed to a single united subcontinent' but had to accept its partition into different states.

8.3 Dictatorships in the Twentieth Century: "odious homogenization"

The Nazi policies of racial purity and genocide are the most horrific example of forced homogenization.

The Soviet Union implemented Russification policies (changing the use of native languages to Russian) from the mid-1930's. More coercive homogenization was also implemented after the Second World War. In fact the Soviet regime was constantly threatened by the diversity of the people it was trying to control. Conquest (1970) describes the deportation of eight entire ethnic groups (including the Crimean Tatars, Volga Germans and Chechens) in the Soviet Union in the 1940s. They were exiled to Siberia and Central Asia, the names of their original habitats changed and their own names removed from the list of Soviet peoples. In some cases these groups had attempted autonomy and were considered a threat to unity (Chechens and Crimean Tatars), other cases were somewhat precautionary (Gorenburg, 2006).[66] Indoctrination by means of public education was an important way of forcing homogenization under a Marxist-Leninist doctrine. Lott (1999) notes that while public services like health were lacking, spending on education was especially high. In fact, the author argues that health spending increased 70% after democratization in former communist countries while public funding for education went down. For all those regimes which face a diverse population, a key aspect of indoctrination is patriotism and in many cases the transition to democracy meant

[65] Of course English was often spoken by the most urban, educated and wealthy, so it does not necessarily represent a choice independent of interest groups.

[66] The population of a block of land over which Turkey and Russia had fought over for over a century was considered to have Turkish sympathies and were arguably deported as a precaution to avoid future trouble in this area with Turkey.

the breakdown of the country (Soviet Union, Yugoslavia, Czechoslovakia).

In Spain, Franco's repressive policies against Catalan culture are well known. Even before Franco the economic prosperity and cultural differences of Catalonia and the Basque region had been seen both as a threat to Castilian political rule in Spain as well as to the unity of the country. The threat to Castilian rule and threat of secession featured heavily in the non-democratic regimes in the first half of the twentieth Century. The Spanish dictator Primo de Rivera is reported to have remarked in 1925 : 'Regions? Out of the question. A quarter of a century's silence about regions...and Spain will have been freed from one of her gravest perils.' In 1938, Franco declared a desire for national unity 'to be complete, with a single language, Castilian, and a single personality, the Spanish one.' The policies implemented in Catalonia aimed to do just this. In 1939 Barcelona was occupied and the keynote speeches of military and politicians from Franco's party were focused on crushing separatism. From 1939 'Purifying committees' sacked teachers involved in catalanism and any one thought to be associated with Catalan nationalist parties was sent to posts in distant parts of Spain. They were replaced by 700 teachers brought in from other parts of Spain and hired mainly for their lack of knowledge of Catalan. A Castilian speaking bishop was put in charge of the Church in Barcelona and given the task of eliminating Catalan from the churches in his diocese.[67]

8.4 Comparing democracies and non-democracies

Proposition 1 finds that democracies homogenize, both to improve access to public goods or the government and to avoid splitting up. Democracies exhibit many policies consistent with nation-building. In most democracies education is publicly provided rather than subsidized by vouchers for individuals to choose their own private education, and often what is taught in school is highly coordinated by governments. One of the reasons why public education is not privatized may be the fear of a loosening of sense of national unity. Lott (1987) argues that the public provision of education in almost all countries and the lack of choice and competition in schooling (even when compared to health provision within the same country) is puzzling and that schooling is used as a means to pass on desired views to students. National holidays are often designed as nation-building events. A sense of patriotism is built even in a democracy, but most of the time with less emphasis and aggressiveness than in some dictatorships.

In contrast, our model predicts that dictators and elites who face a high threat of democracy may try to homogenize, by odious means, above and beyond what a democracy would do. This insight is consistent with recent empirical findings that examine education policies across democracies and non-democratic regimes. Aghion, Persson and Rouzet (2012), using annual data on 137 countries from $1830 - 2001$ find that autocracies have higher enrollment rates in primary education than democracies and in a smaller sample also show that autocracies implement more education reforms. They also determine the period for each country in which primary enrollment rose the most sharply and find that for only 2 out of 53 countries examined does democratic transition occur before the rise in primary enrollment. This is consistent with the evidence in Mulligan, Gil and Sala i Martin (2004) who find no evidence

[67] From Jones (1976).

that democracies spend more on public education. Looking at the same data set Burstyn (2012) finds that democracies spend less on public education than non-democracies for below median income countries. It would of course be interesting to compare the content of education under a democratic regime or a dictatorship. Under the latter we would expect education to have a much higher content of indoctrination, for instance only studying Marxist economics in communist dictatorships, not teaching minority languages or history of ethnic/religious minorities.

8.5 Additional (and non mutually exclusive) Theories

8.5.1 *Foreign Threats.*

Rulers are often threatened not only by domestic insurgents but also by foreign enemies. Homogenization of the population may increase the unity of the nation and patriotism, thus the willingness to fight the enemy. This is consistent with the argument by Aghion, Persson and Rouzet (2012) who establish a correlation between a 'war risk' indicator and investment in public education. They argue that this was an investment to increase the willingness and ability of the citizens to defend the state. For instance, in the 1700s Prussia was a 'state without a nation', it had a centralized bureaucracy overseeing a fractionalized polity. In 1763, Frederick II reiterated a call for compulsory education with the aim 'to unify Prussia through state-directed education.'[68] From 1807 under Frederick III more encompassing reforms were called for and eventually enacted. One motivation was Prussia's conflict with other states (Aghion, Persson and Rouzet, 2012) and the defeat to Napoleon in 1806. The reforms served as a means of unifying the population against external threats and preparing it for military conflict. Similarly, in France, the defeat to Prussia in 1870 saw calls for education to strengthen the country as a means to avoid future defeat. Of course war, or the threat of war, may also indirectly foster the threat of domestic insurgencies and lead to an increase in the probability of democratic transition and nation building via this channel also, and so they are interlinked.

While war is important, it is difficult to fully explain nation building with the threat of war alone. The increase in homogenization policies, particularly in the last third of the 19th Century, coincides with the period in which democratization was in progress and further democratization clearly inevitable. It is not clear why external conflict would induce particularly intense homogenization towards the end of the 19th Century rather than earlier. Italy's military conscription efforts are illuminating in this period: the forming of regiments of Italians from all parts of the country and stationing recruits far from home is argued to have actually been counter-productive to producing an efficient fighting force as it would have hindered rapid mobilization.[69] Internal threats rather than external threats seem more important in other cases as well. Franco's policies of repression in Catalonia appear to be a response to threats from within Catalonia with no obvious relationship to external threats. In 1939 repressive policies in Catalonia were particularly intense,[70] while at the same time Franco reduced the

[68] Ramirez and Boli, (1987).
[69] Duggan (2007), p289.
[70] Young (1976) p236 – 241.

size of the army by three quarters, only increasing it again later as a result of the Second World War.[71] In the early Twentieth Century the deportation of Armenians in the Ottoman Empire destroyed part of the agricultural sector in Anatolia and with it the availability of strategic food supplies to the army.[72] Finally, we find that education reforms in Nineteenth Century Europe follow episodes of major collective action. Historical accounts generally do not describe these episodes of collective action as being driven by the threat of an external war but rather by internal conditions.

8.5.2 *Industrialization.*

Gellner (1983) argues that agrarian societies have no need for a "nation" in the modern sense of the word.[73] An industrial society based upon markets (as opposed to a stratified agrarian society with local markets) needs better means of communication. This requires "sustained and precise communications between strangers involving a sharing of explicit meaning transmitted in a standard idiom" (Gellner (1983) page 34). Universal schooling serves an economic purpose as well, necessary for the development of an industrial society based upon markets.[74] In other words productivity would increase in an industrial society with more homogenization relative to an agrarian society.[75] The higher geographical and social mobility of an industrial society relative to an agrarian one necessitates more homogeneity, at least in communication.

Gellner's argument is, in part, disputed (see Smith, 2003, chapter 2) because of an issue of timing. Smith (2003) argues that in many countries there was no industrial development when nationalism first emerged. Mass education reforms across countries also appear inconsistent with industrial advancement. In most countries compulsory education appears late in the 19th Century, presenting a puzzling lag between the onset of industrialization and state-sponsored education. Neither do education reforms arrive in an obvious way country by country: education reforms in England were particularly slow despite being the birthplace of the industrial revolution and thus presumably requiring a more educated population sooner (Katznelson and Weir, 1985). Industrialization may provide a better explanation for bottom-up homogenization (for example increased homogenization of language simply as a result of working with people from different regions and the need to communicate with them), a topic that we leave for future research.[76]

[71] Bowen (2007).

[72] Zürcher (2010).

[73] Gellner (1983) writes that political units in agrarian societies "can de divided into two species: local self governing communities and large empires". Neither of these type of governments represent a modern nation state.

[74] See also Bowles (1998) on this point and for a survey of other models in which preferences are endogenous and can be influenced by various institutions.

[75] See Alesina and La Ferrara (2005) for a survey of models regarding the pros and cons of diversity for productivity and development.

[76] As an example of bottom-up homogenization Weber (1979) documents the department of Vosges in France where the introduction of the cotton industry in the 1870s 'all but wiped out the local dialect when country people moved into small industrial centers.'

9 Conclusion

We examined when and to what extent a government chooses policies directed toward homogenizing its a population. We offer six key findings. One, when the probability of democracy is low a dictator undertakes no homogenization. He chooses a government that is ideal for himself and allows the population to remain heterogeneous since he faces little threat of overthrow and does not care about population welfare. Two, a democracy undertakes a positive amount of homogenization in order to improve general access to the public good chosen by majority rule. Three, a ruler who faces a high probability of overthrow may undertake significant homogenization of the population in excess of anything undertaken by a democracy. Indeed he may homogenize enough to ensure a single country where a democracy alone would instead break up. The dictator has a strong vested interest in the current government and when challenged a dictator is willing and able to undertake homogenization in order to better preserve the status quo. Fourth, contrary to a democracy, a dictator will always chooses "odious" forms of homogenization, particularly costly for minorities. Fifth, when the probability of democratic success decreases with the degree of homogeneity of the population the dictator will strategically take it into account and homogenize more. Six, in some cases the ruler may choose policies of "divide and rule" in order to reduce the feeling of national identity because the latter may increase the likelihood of national insurgencies. This may be particularly likely to occur when the rulers are foreign colonizers. Finally we offer some suggestive historical discussion which is consistent with these results.

Appendix

Lemma 1 : *A democracy will locate the government at the center of the population.*

Proof. Label by $j \in [0,1]$ the position of the government. If $j \in [0.25, 0.75]$ then the median voter is individual i at $d_{ij} = 0.25$ and homogenization is λ_1^m. Locating the government at the center clearly beats all other $j \in [0.25, 0.75]$ in a pairwise vote. If $j \in [0, 0.25)$ (the argument for $j \in (0.75, 1]$ is symmetric) then the median individual is always the individual at $i = 0.5$ and the level of homogenization chosen satisfies $ga(0.5 - j) = C'(\lambda)$, denoted λ^j. Label by $\hat{l}_i \in [0, 0.5]$ the distance from the center of the individual on the interval $i \in [0, 0.5]$ who is indifferent between the government at some $j \in [0, 0.25)$ and a government at the center.[77] Similarly denote by $\hat{\hat{l}}_i \in [0, 0.5]$ the individual that satisfies the same condition on the interval $i \in [0.5, 1]$. Then

$$g - ga(1 - \lambda_1^m)\hat{l}_i + y - k - C(\lambda_1^m) = g - ga(1 - \lambda^j)(0.5 - j - \hat{l}_i) + y - k - C(\lambda^j) \quad (6)$$

$$g - ga(1 - \lambda_1^m)\hat{\hat{l}}_i + y - k - C(\lambda_1^m) = g - ga(1 - \lambda^j)(0.5 - j + \hat{\hat{l}}_i) + y - k - C(\lambda^j).$$

[77] This person must be located to the right of j. If not then $\hat{l}_i > 0.25$. Since λ_1^m is optimal for $i = 0.75$ while the government at j is further away from $i = 0.75$ and homogenization λ^j is less than optimal, then $\hat{\hat{l}}_i > 0.25$ and the result follows.

The proportion of the population who prefer a government at the center is $\hat{l}_i + \hat{\hat{l}}_i$. Rearrange to find

$$\hat{l}_i + \hat{\hat{l}}_i = \frac{1}{ga}\left(\frac{2(1-\lambda_1^m)}{(1-\lambda_1^m)^2-(1-\lambda^j)^2}\right)[C(\lambda^j) - C(\lambda_1^m) + ga(1-\lambda^j)(0.5-j)]$$

Since $C(\lambda)$ is a convex continuously differentiable function on $\lambda \in (0,1)$ then $C(\lambda^j) - C(\lambda_1^m) \geq C'(\lambda_1^m)[\lambda^j - \lambda_1^m] = ga0.25(\lambda^j - \lambda_1^m)$ and since we examine $j < 0.25$ we have $ga(1-\lambda^j)(0.5-j) > ga(1-\lambda^j)0.25$. Using these inequalities it can be seen that $\hat{l}_i + \hat{\hat{l}}_i > 0.5$. In the same way it is easy to show that any democracy of size γ will locate the government at the center by showing $\gamma(\hat{l}_i + \hat{\hat{l}}_i) > \gamma 0.5$.

Proof of Proposition 1

Following Lemma 1, if a democracy forms a single country the government is located at the center with homogenization λ_1^m; if a democracy forms two countries the government is located at the center of those countries with homogenization λ_2^m. The vote over one or two countries is determined as follows. Expression (2) for individual i at distance $l_i \in [0, 0.25]$ from the center of the population and individual i at distance $l_i \in [0.25, 0.5]$ can be rewritten respectively as

$$-((1-\lambda_2^m) + (1-\lambda_1^m))gal_i + (1-\lambda_2^m)ga0.25 + k - [C(\lambda_1^m) - C(\lambda_2^m)] \tag{7}$$

and

$$((1-\lambda_2^m) - (1-\lambda_1^m))gal_i - (1-\lambda_2^m)ga0.25 + k - [C(\lambda_1^m) - C(\lambda_2^m)]. \tag{8}$$

Expression (7) is at a maximum when $l_i = 0$ and decreasing until $l_i = 0.25$; while expression (8) is increasing from this point at $l_i = 0.25$ to a maximum at $l_i = 0.5$. Thus there exist uniquely two individuals, $l_i' \in [0, 0.25]$ and $l_i'' \in [0.25, 0.5]$, with the same value of staying as one country relative to splitting into two countries and such that $l_i' + (0.5 - l_i'') = 0.25$, who are the median voters. Thus l_i' solves:

$$\begin{aligned}&-((1-\lambda_2^m) + (1-\lambda_1^m))gal_i' + (1-\lambda_2^m)ga0.25 + k - C(\lambda_1^m) + C(\lambda_2^m)\\ &= ((1-\lambda_2^m) - (1-\lambda_1^m))ga(0.25 + l_i') - (1-\lambda_2^m)ga0.25 + k - C(\lambda_1^m) + C(\lambda_2^m).\end{aligned}$$

Hence

$$l_i' = 0.25\frac{(1-\lambda_1^m) + (1-\lambda_2^m)}{2(1-\lambda_2^m)} \qquad l_i'' = 0.25(1 + \frac{(1-\lambda_1^m) + (1-\lambda_2^m)}{2(1-\lambda_2^m)})$$

Since $\lambda_1^m > \lambda_2^m$, the median voters are at $0.125 < l_i' < 0.25$ and $0.375 < l_i'' < 0.5$. A median voter is found in each of the intervals $(0, 0.125), (0.25, 0.375), (0.625, 0.75), (0.875, 1)$.

Assumption on convexity of the cost function: *For $C(\lambda)$ sufficiently convex, there exists λ^* such that for all $\lambda < \lambda^*$ undertaken by the dictator a democracy would choose to form two countries and for all $\lambda \geq \lambda^*$ a democracy would choose to form a single country.*

We prove the above statement and highlight the convexity requirement. The only action of the ruler in period 1 that affects the choice of a democracy in period 2 is the level of homogenization, λ^r. Given λ^r, in period 2 if a democracy forms a single country, the government is located in the center and additional homogenization $\lambda = \max\{0, \lambda_1^m - \lambda^r\}$ is undertaken at cost $C(\lambda + \lambda^r) - C(\lambda^r)$. If a democracy forms two countries the governments are located at the respective centers with additional homogenization $\lambda = \max\{0, \lambda_2^m - \lambda^r\}$. A proof of these statements follows in the same way as Lemma 1 and Proposition 1 and can be found online in Appendix *A*3. We can then determine the outcome if democracy prevails in period 2, given λ^r. A democracy forms a single country when

$$[g - (1 - \lambda_{11})gal_i + y - k - C(\lambda_{11})] - [g - (1 - \lambda_{22})ga(0.25 - l_i) + y - 2k - C(\lambda_{22})] \geq 0 \quad (9)$$

where $l_i = 0.25((1 - \lambda_{11}) + (1 - \lambda_{22}))/2(1 - \lambda_{22})$ is the median voter following the proof of Proposition 1; $\lambda_{22} = \lambda_2^m$ if $\lambda^r \leq \lambda_2^m$ and $\lambda_{22} = \lambda^r$ otherwise; $\lambda_{11} = \lambda_1^m$ if $\lambda^r \leq \lambda_1^m$ and $\lambda_{11} = \lambda^r$ otherwise.

It can be seen immediately that if (9) is weakly positive for some $\lambda' \in [\lambda_1^m, 1]$ then (9) is positive for all $\lambda > \lambda'$ and if (9) is zero for some $\lambda' \in [\lambda_1^m, 1]$ then (9) is negative for all $\lambda < \lambda'$. Suppose (9) is strictly negative for $\lambda^r = \lambda_1^m$, then sufficient convexity implies (9) is also negative for all $\lambda^r < \lambda_1^m$. Observe that since $-(1 - \lambda_1^m)ga0.25 + k = -C$ for some $C > 0$, for any λ^r (9) can be rewritten

$$\begin{aligned} &- (1 - \lambda_1^m)ga(l_i - 0.25) + (1 - \lambda^r)ga(0.25 - l_i) - [C(\lambda_1^m) - C(\lambda^r)] - C \\ &\leq (\lambda_1 - \lambda^r)ga0.25\frac{(1 - \lambda_1)}{2(1 - \lambda^r)} - C \end{aligned}$$

where the inequality follows by the convexity and continuous differentiability of the cost function. This shows (9) is negative for $\lambda^r \in [\lambda_2^m, \lambda_1^m)$ not too far from λ_1, i.e. $C''(\lambda)$ high enough in the interval $[\lambda_2^m, \lambda_1^m]$. Now examine the case where (9) is weakly positive at $\lambda^r = \lambda_1^m$. The derivative of (9) with respect to λ_2 is

$$-ga0.25\frac{(1 - \lambda_1^m)^2 + (1 - \lambda^r)^2}{2(1 - \lambda^r)^2} + C'(\lambda^r). \quad (10)$$

where (10) is increasing at rate

$$-ga0.25[\frac{(1 - \lambda_1^m)^2}{(1 - \lambda^r)^3}] + C''(\lambda^r). \quad (11)$$

Since (10) is negative at λ_2^m and zero at λ_1^m then (11) must be strictly positive for some $\lambda^r \in [\lambda_2^m, \lambda_1^m]$. If $C''(\lambda)$ is high enough in the interval $\lambda^r \in [\lambda_2^m, \lambda_1^m]$ then (11) is positive in this interval and (10) negative. In which case (9) is positive for all $\lambda \in [\lambda_2^m, \lambda_1^m]$.

Finally note that if (9) is positive at λ_2^m then this represents a situation where parameters are such that were the population a democracy it would choose to form a single country. Suppose at $\lambda^r = \lambda_1^m$ that (9) is negative. A contradiction since then (9) is negative at λ_2^m. It follows that at $\lambda^r = \lambda_1^m$ that (9) is positive, hence (9) is positive for all λ^r and so $\lambda^* = 0$. If (9) is negative at λ_2^m this represents a situation where parameters are such that were the population a democracy it would choose to form two countries. Suppose (9) is positive at $\lambda^r = \lambda_1^m$ then we have a contradiction. Thus (9) is negative at $\lambda^r \in [\lambda_2^m, \lambda_1^m]$ and so $\lambda^* > \lambda_1^m$.

Proof of Proposition 4

The ruler's expected utility in period 2 does not depend on his choice of location in period 1. See Appendix *A*3 online for details. Thus in period 1 the ruler locates the government at his ideal point and homogenizes to λ^r. His utility in period 1 is $g + y - k - C(\lambda^r)$. In period 2 with probability $1 - p$ democracy does not prevail and his utility is $g + y - k$. With probability p democracy prevails and his utility depends on the borders, government and homogenization chosen by a democracy. If $\lambda^r \geq \lambda^*$, a period 2 democracy forms a single country, the government is located in the center, and additional homogenization $\lambda = \max\{0, \lambda_1^m - \lambda^r\}$ is undertaken at cost $C(\lambda + \lambda^r) - C(\lambda^r)$. If $\lambda^r < \lambda^*$, a period 2 democracy forms two countries, the governments are located at the respective centers, with additional homogenization $\lambda = \max\{0, \lambda_2^m - \lambda^r\}$ at cost $C(\lambda + \lambda^r) - C(\lambda^r)$. Thus to determine λ^r we examine the ruler's expected utility in the second period including the cost of homogenization in the first period.

Suppose $\lambda^* = 0$ and denote by λ_1^r the optimal homogenization undertaken by the ruler. If democracy prevails in the second period a democracy will homogenize to λ_1^m, then $\lambda_1^r \notin (0, \lambda_1^m]$ since expected utility would always be strictly lower than $\lambda_1^r = 0$. Thus a ruler located at distance l_i from the center will homogenize to $\lambda_1^r = 0$ if his utility given by expression (12) is greater than his utility given by the maximized expression in (13), otherwise he homogenizes to $\lambda_1^r = \tilde{\lambda}$, where

$$p(g - (1 - \lambda_1^m)gal_i + y - k - C(\lambda_1^m)) + (1 - p)(g + y - k). \tag{12}$$

$$\tilde{\lambda} = \underset{\lambda \in (\lambda_1^m, 1]}{\arg\max}[p(g - (1 - \lambda)gal_i + y - k) + (1 - p)(g + y - k) - C(\lambda)] \tag{13}$$

Homogenization is weakly increasing in p.

Suppose instead $\lambda^* > 0$. Examine the optimal homogenization in the interval $[0, \lambda^*)$, denoted λ_2^r. Similar to above, since a democracy will homogenize to λ_2^m, a ruler located at distance l_i from the center will homogenize to $\lambda_2^r = 0$ when his utility given by expression (14)

is greater than his utility given by the maximized expression in (15), otherwise $\lambda_2^r = \tilde{\tilde{\lambda}}$, where

$$p(g - (1 - \lambda_2^m)ga|0.25 - l_i| + y - 2k - C(\lambda_2^m)) + (1 - p)(g + y - k). \tag{14}$$

$$\tilde{\tilde{\lambda}} = \underset{\lambda \in (\lambda_2^m, \lambda^*)}{\arg\max}[p(g - (1 - \lambda)ga|0.25 - l_i| + y - 2k) + (1 - p)(g + y - k) - C(\lambda)] \tag{15}$$

Homogenization λ_2^r is weakly increasing in p. The optimal $\lambda \in [\lambda^*, 1]$, which thus ensures a democracy would not split, is denoted $\hat{\lambda} = \max\{\lambda^*, \tilde{\lambda}\}$ where $\tilde{\lambda}$ is given by (13) and so $\hat{\lambda}$ is weakly increasing in p. It remains to show that if a ruler does better by homogenizing to $\hat{\lambda}$ than λ_2^r at some p, then this is true for all p higher. The dictator's expected utility from homogenizing to $\hat{\lambda} \geq \lambda_1^m$ is $p(g - (1 - \hat{\lambda})gal_i + y - k) + (1 - p)(g + y - k) - C(\hat{\lambda})$. Expected utility from homogenizing to λ_2^r is as above. Differentiation with respect to p of the expected utility of homogenizing to one country minus the expected utility of homogenizing to two is given by[78]

$$-(1 - \hat{\lambda})gal_i + (1 - \lambda_2^m)ga|0.25 - l_i| + k + C(\lambda_2^m) \tag{16}$$

or

$$-(1 - \hat{\lambda})gal_i + (1 - \lambda_2^r)ga|0.25 - l_i| + k. \tag{17}$$

Terms with $\partial\hat{\lambda}/\partial p$ and $\partial\lambda_2^r/\partial p$ cancel out.[79] Since $\hat{\lambda} \geq \lambda^* > \lambda_1^m$ (from above) we know that $-(1 - \hat{\lambda})ga0.25 + k \geq 0$ and it follows that (16) is positive.

Proof of Proposition 5

Without loss of generality we can write $M(\mu, d_{ij}) = \beta(\mu) + \alpha(\mu)d_{ij}$. Since total costs of different technologies are equalized for a government located in the center of the country

$$2\int_0^{0.5} [\beta(\mu) + \alpha(\mu)x]\, dx = C(\mu).$$

Hence $\beta(\mu) + \alpha(\mu)0.25 = C(\mu)$ for all μ and it follows that $\beta'(\mu) + \alpha'(\mu)0.25 = C'(\mu)$. Rewrite the first order condition $gad_{ij} = \beta'(\mu) + \alpha'(\mu)d_{ij}$. Since $M(\mu, 0) = \beta(\mu)$, then $\beta'(\mu) > 0$, which is continuous by assumption and therefore so is $\alpha'(\mu)$, and since $M_\mu(\mu, d_{ij})$ is increasing in d_{ij}, $\alpha'(\mu) > 0$. Thus $ga > \alpha'(\mu) > 0$ and it follows that the optimal μ for an individual i is increasing in d_{ij}. Also preferences are single peaked over μ.

For a government at $j = 0.5$ it follows that for the level of odious homogenization $d_{ij} = 0.25$ is the median voter. Median homogenization, denoted μ_1^m, satisfies $ga0.25 = \beta'(\mu_1^m) + \alpha'(\mu_1^m)0.25$. But since also $ga0.25 = C'(\lambda_1^m)$ then $C'(\lambda_1^m) = \beta'(\mu_1^m) + \alpha'(\mu_1^m)0.25$, and from

[78] There exists a $p = p'$ at which (15) and (14) are equal. For $p < p'$ then, differentiation with respect to p of the expected utility of homogenizing to one country minus the expected utility of homogenizing to two gives (16) and if $p \geq p'$ this is (17).

[79] The term including $\partial\hat{\lambda}/\partial p$ is $+\frac{\partial\hat{\lambda}}{\partial p}[pgal_i - C'(\hat{\lambda})]$. There exists a p denoted p' at which $\lambda_1^r = \lambda^*$, then for all $p < p'$ $\frac{\partial\hat{\lambda}}{\partial p} = 0$ and for all $p \geq p'$ $pgal_i - C'(\hat{\lambda}) = 0$. In (17) the term $\frac{\partial\lambda}{\partial p}[pga|0.25 - l_i| - C'(\lambda)]$ is zero since $pga|0.25 - l_i| = C'(\tilde{\tilde{\lambda}})$.

above $\lambda_1^m = \mu_1^m$. Each individual evaluates the difference between their utility in the case of non-odious homogenization and their utility in the case of odious homogenization:

$$[-(1-\lambda_1^m)gad_{ij} - C(\lambda_1^m)] - [-(1-\mu_1^m)gad_{ij} - M(\mu_1^m, d_{ij})]. \tag{18}$$

Since $\lambda_1^m = \mu_1^m$, expression (18) is increasing in d_{ij} and $d_{ij} = 0.25$ is the median voter when deciding between odious and non-odious homogenization. He is indifferent between the two.

Where the government is located may affect the type and level of homogenization chosen. We show that for a government located other than $j = 0.5$, a democracy is either again indifferent between the homogenization methods or strictly prefers non-odious homogenization. If $j \in [0.25, 0.75]$ the median voter over the level of odious homogenization is still at $d_{ij} = 0.25$ and μ_1^m is chosen. Similarly the median voter in the choice over odious or non-odious is still at $d_{ij} = 0.25$ and is indifferent. For $j \in [0, 0.25)$, the median voter when choosing the level of odious or non-odious homogenization is $i = 0.5$. From above we saw $M_\mu(\mu, 0.25) = C'(\mu)$ hence for all $d_{ij} > 0.25$ we have $M_\mu(\mu, d_{ij}) > C'(\mu)$. For $j \in [0, 0.25)$, since the median voter is at $d_{ij} = 0.5 - j > 0.25$ then the median level of odious homogenization chosen will be lower than the median level of non-odious. Each individual evaluates the difference between their utility in the case of non-odious homogenization and their utility in the case of odious homogenization: $[-(1-\lambda^r)gad_{ij} - C(\lambda^r)] - [-(1-\mu^r)gad_{ij} - M(\mu^r, d_{ij})]$. Since $\mu^r < \lambda^r$, this is increasing in d_{ij}. The median individual in the choice between odious or non-odious homogenization is also $d_{ij} = (0.5 - j)$. He must prefer non-odious homogenization since he could always choose the same level of odious homogenization but use non-odious methods and do strictly better.

Proof of Proposition 6

In period 2 if democracy prevails any further homogenization is undertaken by non-odious means (that a democracy weakly prefers non-odious can be shown as the proof of Proposition 5). As with non-odious homogenization the only action of the ruler in period 1 that can directly affect the democratic outcome is the level of homogenization chosen, denoted μ^r. The only difference compared to non-odious homogenization is the cost to the ruler of homogenizing in period 1 : in period 1 the ruler's utility is $g + y - k - M(\mu^r, 0)$, while the effect of μ^r on the democratic outcome is exactly the same as the equivalent amount of non-odious homogenization described in the proof of Proposition 4. Part (i) follows by the same arguments as Proposition 4 and is contained in the online appendix.

To see that a dictator will always undertake at least as much homogenization as a democracy when p is high, examine the case $p = 1$. In a democracy the population homogenizes to λ_2^m if the population splits and λ_1^m otherwise, at cost $C(\lambda)$. Now if a ruler homogenizes by μ in period 1 this costs the ruler $M(\mu, 0)$. If democracy prevails in period 2 and a democracy would split then $\mu < \lambda_2^m$ cannot be optimal since then period 2 homogenization is $\lambda_2^m - \mu$ and costs the ruler $-C(\lambda_2^m) + C(\mu)$. This cost is uniquely minimized at $\mu = \lambda_2^m$ by noting that $M_\mu(\mu, 0) < C'(\mu)$. It is clear this is also true for an interval $p \in [1 - \epsilon, 1]$. Similarly if a democracy would form a single country $\mu < \lambda_1^m$ cannot be optimal.

To see that in some cases when p is high a dictator will undertake strictly more homogenization and will ensure a democracy will form a single country where it would have otherwise split, take the example of the ruler at $i = 0.5$ where a democracy prefers to split and will homogenize to λ_2^m. At $p = 1$, optimal homogenization $\mu_2^r \in [0, \lambda^*)$ satisfies $ga0.25 = M_\mu(\mu, 0)$ which is strictly greater than λ_1^m since $M_\mu(\mu, 0) < C'(\mu)$. The optimal $\mu \in [\lambda^*, 1]$ is λ^* since this ensures a single country and the ruler's ideal government at the cheapest cost. Thus the dictator always undertakes strictly more homogenization than a democracy. Under some parameters he may also choose to form a single country where a democracy otherwise would not. Observe that his value of homogenizing to λ^* versus μ_2^r is

$$[g + y - k - M(\lambda^*, 0)] - [g - (1 - \mu_2^r)ga0.25 + y - 2k - M(\mu_2^r, 0)].$$

This is positive for $\lambda^* \in (\lambda_1^m, \gamma]$, where $\gamma > \lambda_1^m$.

10 References

Acemoglu, D. and J. A. Robinson (2000) "Why did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective. " *Quarterly Journal of Economics*, 115(4) 1167-99.

Aghion, P., T. Persson and D. Rouzet (2012)"Education and Military Rivalry."NBER Working Paper No. 18049, National Bureau of Economic Research, Inc.

Alesina, A., W. Easterly and J. Matuszeski (2011) "Artificial States." *Journal of the European Economic Association*, European Economic Association, 9(2), 246 − 77, 04.

Alesina, A. and N. Fuchs-Schündeln (2007) "Good bye Lenin (or not?): The effect of Communism on people's preferences." *American Economic Review*, 97, 1507 − 28.

Alesina, A. and E. Glaeser (2005) *Fighting Poverty in the US and Europe: A World of Difference.* Oxford University Press, USA.

Alesina, A., J. Harnoss and H. Rapoport (2013) "Birthplace Diversity and Economic Prosperity."NBER Working Papers 18699, National Bureau of Economic Research, Inc.

Alesina A. and E. La Ferrara (2005) "Ethnic Diversity and Economic Performance" Journal of Economic Literature 721-61.

Alesina, A. and E. Spolaore (1997) "On the number and size of nations." *Quarterly Journal of Economics*, 112(4), 1027 − 56.

Alesina, A. and E. Spolaore (2003) *The size of nations.* MIT Press, Cambridge MA.

Alesina, A., E. Spolaore and R. Wacziarg (2000) "Economic integration and political disintegration." *American Economic Review*, 90(5), 1276 − 96.

Anderson, B. (1983) *Imagined communities: reflections on the origins and spread of nationalism.* Verso, London, UK.

Algan, Y., T. Mayer and M. Thoenig (2013)"The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France."CEPR discussion paper No. DP9416.

Asprachs-Bracons, O., J. Costa-Font, I. Clots-Figueras and P. Masella (2008) "Compulsory

Language Educational Policies and Identity Formation." *Journal of the European Economic Association*, $6(2-3)$, $434-444$.
Bisin, A. and T. Verdier (2000) "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits." *The Quarterly Journal of Economics*, 115(3), $955-88$.
Bourgignon F. and T. Vedier (2000) "Oligarchy Democracy Inequalty and Growth." *Journal of Development Economics*, 62(2), $285-313$.
Besley T. and T. Persson (2010) "State Capacity, Confict and Development." *Econometrica*, 78(1), 1-34.
Besley T. and T. Persson (2011) "Fragile States and Development Policy." *Journal of the European Economic Association*, 9(3), $371-98$.
Bowen, W.H. (2007) "The Spanish Military during World War Two." in *A Military History of Modern Spain*, W.H. Bowen and J.E. Alvarez eds. Praeger, USA.
Bowles S. (1998) "Endogenous Preferences: The Cultural Consequences of Markets and other Economic Institutions" *Journal of Economic Literature* 75-111.
Bursztyn Leonardo (2013) "Poverty and the Political Economy of Public Education Spending: Evidence from Brazil" UCLA Working Paper
Bolton, P. and G. Roland (1997) "The Breakup of Nations: A Political Economy Analysis." *Quarterly Journal of Economics*, 112(4), 1057-90.
Christopher, A. J. (1988) "'Divide and Rule': The Impress of British Separation Policies." Wiley, 20(3), 233-249.
Conquest, R. (1970) *The Nation Killers.* The University Press, Glasgow, UK.
Colley, L (1986) "Whose Nation? Class and National Consciousness in Britain $1750-1830$." *Past and Present*, 113, 97-117.
Duggan, C. (2007) *The Force of Destiny: A History of Italy since* 1796. Penguin Books, London, UK.
Easterly W. and R. Levine (1997) "Africa's Growth Tragedy: Policies and Ethnic Division." *Quarterly Journal of Economics*, 112(4), 1203-50.
Flora, P. et al. (1983) *State, economy, and society in Western Europe 1815-1975 : a data handbook in two volumes. Vol.1.* Campus Verlag, Frankfurt.
Gellner E. (1983) Nations and Nationalism *Cornell University Press.*
Gennaioli N. and I Rainer (2007) The modern Impact of PreColonila Centralization in Africa" *Journal of Econimic Growth* 12, 185-234
Glaeser E. G Ponzetto and A. Shleifer (2007) "Why Democracy needs Education?" *Journal of Econimic Growth* 12(4), 77-99.
Gorenburg, D. (2006) "Soviet Nationalities, Policy and Assimilation" in *Rebounding Identities: The Politics of Identity in Russia and Ukraine*, D. Arel, B. A. Ruble, Woodrow Wilson Center Press.
Herbst (2000) States and Power in Africa Princton University Press Princeton NJ
Hobsbawn, E. J. (1990) *Nations and nationalism since 1780: programme, myth, reality.* Cambridge University Press, UK.
Holland, M. (2005) "The Captain Swing Project" in *Swing Unmasked*, M. Holland. FACHRS

publications, UK.
Içduygu, A. S. Toktas and B. Ali Soner (2008) "The Politics of Population in a Nation-Building Process: Emigration of non-Muslims from Turkey" *Ethnic and Racial Studies*, 31(2), 358-389.
Jones, N. L. (1976) "The Catalan Question since the Civil War" in *Spain in Crisis* ed. Paul Preston. The Harvest Press Ltd, UK.
Laitin D (2007) Nations Sates and Violence Oxford University Press, Oxford, U Press
Lott, J. R. Jr. (1999) "Public Schooling, Indoctrination, and Totalitarianism." *Journal of Political Economy*, 107, 127-57.
Lott, J. R. Jr. (1990) "An Explanation for Public Provision of Schooling: the Importance of Indoctrination." *Journal of Law and Economics*, 33(1), 199-231.
Lott, J. R. Jr. (1987) "Why is Education Publicly Provided? A Critical Survey." *Cato Journal*, 7(2), 475-501.
Marcham, A. J. (1973) "Educating our masters: political parties and elementary education 1867 to 1870." *British Journal of Educational Studies*, 21(2), 180-191.
Marten, L. and N. Kula (2008) *Zambia: One Zambia, One Nation, Many Languages.* In Simpson, Andrew, (ed.), Language and National Identity in Africa. Oxford: Oxford University Press.
Michaopolous S. and E. Papaioannou (2012) "The Long Run Effects of the Scramble for Africa." NBER Working Paper.
Miquel, E., S. Satyanath and E. Sergenti (2012) "Economic Shocks and Civil Conflict: an Instrumental Variables Approach." *Journal of Political Economy*, 112(4), 725-753.
Mulligan C., R. Gil and X. Sala-i-Martin (2004) "Do Democracies have Different Public Policies than Nondemocracies" Journal of Economic Perspectives 18(1) 51-74.
Phiri, B. J. (2006) *A Political History of Zambia: From the Colonial Period to the 3rd Republic.* Africa World Press.
Ramirez, F.O. and J. Boli (1987) "The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization." *Sociology of Education*, 60(1), 2-17.
Richards, M. (1998) *A Time of Silence: Civil War and the Culture of Repression in Franco's Spain,* 1936 – 1945. Cambridge University Press, UK.
Schram Albert (1997) *Railways and the formation of the Italian State in the Nineteenth Century.* Cambridge University Press, UK
Smith, A. D. (2003) *Nationalism and Modernism,* Routledge.
Tilly, C. (1975) "On the History of European State Making" in *The Formation of National States in Western Europe,* C. Tilly. Princeton University Press.
Tilly, C. (1998) "Disturbances in France, 1830-1860 and 1930-1960: intensive sample." Inter-University Consortium for Political and Social Research, Ann Arbor, MI. doi:10.3886/ICPSR00051.v1.
Tilly, C., L. Tilly and R. Tilly (1980) *The Rebellious Century* 1830 – 1930 Harvard University Press, Cambridge MA.
Weber, E. (1979) *Peasants into Frenchmen.* Chatto and Windus, London.
Young, G. M. and W. D. Handcock eds. (1964) *English Historical Documents,* 1833 – 1874. Eyre and Spottiswoode, London.
Zürcher, E. J. (2010) *The Young Turk Legacy and Nation Building: From the Ottoman Empire*

to Atatürk's Turkey. I.B. Taurus, London.

For online publication

A1

When there is no option to homogenize ($\lambda = 1$), the value of forming a single country versus splitting into two for the median voter is $-ga0.25 + k$. When homogenization is possible, the value of forming a single country versus splitting into two for the median voter is

$$[g - (1 - \lambda_1^m)gal_i + y - k - C(\lambda_1^m)] - [g - (1 - \lambda_2^m)ga(0.25 - l_i) + y - 2k - C(\lambda_2^m)] \quad (19)$$

where $l_i = 0.25((1 - \lambda_1^m) + (1 - \lambda_2^m))/2(1 - \lambda_2^m)$. To see that (19) is strictly higher than when there is no option to homogenize, observe that the expression

$$\begin{aligned}[g - (1 - \lambda_2^m)gal_i + y - k - C(\lambda_2^m)] - [g - (1 - \lambda_2^m)ga(0.25 - l_i) + y - 2k - C(\lambda_2^m)] \\ = -(1 - \lambda_2^m)ga(2l_i - 0.25) + k\end{aligned}$$

is strictly higher than when no homogenization is allowed. Then differentiating expression (19) with respect to λ_1^m to get

$$ga0.25\frac{(1 - \lambda)}{(1 - \lambda_2^m)} + ga0.25 - C'(\lambda) \quad (20)$$

we see this is positive for all $\lambda \in [\lambda_2^m, \lambda_1^m]$.

A2

The homogenization technology is such that the cost of homogenization by μ is $M(\mu, d_{ij})$ which is increasing in d_{ij} at rate $\alpha(\mu)$. Let us give an example of such a technology. This cost function will take the form $M(\mu, d_{ij}) = \beta(\mu) + \alpha(\mu)d_{ij}$. Let us take the particular function where $\beta(\mu) = 0.5C(\mu)$ for all μ. Intuitively this implies that the cost of this technology is divided into two parts. A cost that we can think of as the cost of using government apparatus (this could be enforcement for example), $0.5C(\mu)$, which is divided equally among the population through taxes. The rest of the cost burden is shouldered in greater proportion by those who are homogenized by more. One way to think of this is as a personal cost, for example of being forced not to speak one's own language. Since we assume $\beta(\mu) = 0.5C(\mu)$ and from Proposition 5 we know $M(\mu, 0.25) = C(\mu)$, then we calculate that $\alpha(\mu) = 2C(\mu)$. Thus in this example $M(\mu, d_{ij}) = 0.5C(\mu) + 2C(\mu)d_{ij}$. It is clear this satisfies all conditions: $M(0, d_{ij}) = 0$; $M_\mu(\mu, d_{ij}) = 0.5C'(\mu) + 2C'(\mu)d_{ij} > 0$; it is then clear that $\lim_{\mu\to 0} M_\mu(\mu, d_{ij}) = 0$, $\lim_{\mu\to 1} M_\mu(\mu, d_{ij}) = \infty$ and the marginal cost is higher for those who are homogenized by more; also $M_{\mu\mu}(\mu, d_{ij}) > 0$. Finally it can be seen that for $j = 0.5$ the total cost sums to $C(\mu)$.

Note that if we fix the homogenization technology to be increasing at rate $\alpha(\mu)$ in d_{ij}, then for a different country size, i.e $s = 1/2$, the fixed cost component must change to equate the total costs of odious and non-odious homogenization. Take the example above, where $\alpha(\mu) = 2C(\mu)$. For two countries to equate the costs of odious and non-odious homogenization we have $\beta(\mu) + 2C(\mu)0.125 = C(\mu)$ thus $\beta(\mu) = 0.75C(\mu)$.

A3

In the first period the ruler locates the government at j and undertakes homogenization λ^r. First period utility for any individual i is $g-(1-\lambda^r)gad_{ij}+y-k-C(\lambda^r)$. In the second period, if the ruler is in power, then he behaves like a safe ruler and utility for any individual i is $g-(1-\lambda^r)gad_{ij}+y-k$. If democracy prevails in the second period a population chooses either one country or two, where to locate the government and how much additional homogenization to undertake. Suppose one country is formed. For a government located at j, the utility of i satisfies $g-(1-\lambda^r-\lambda)gad_{ij}+y-k-[C(\lambda^r+\lambda)-C(\lambda^r)]$ where λ is any additional homogenization undertaken in period 2. Differentiating with respect to λ we get $gad_{ij}-C'(\lambda^r+\lambda)$ and thus the choice of λ is single-peaked and weakly increasing in d_{ij}. The level of homogenization chosen by majority rule if a single country is formed with a government at j is as follows. If $j\in[0.25.0.75]$, the median voter is at $d_{ij}=0.25$ and chooses $\lambda=\max\{0,\lambda_1^m-\lambda^r\}$. If $j\in[0,0.25)$ (symmetrically for $j\in(0.75,1]$), the median voter is at $0.5-j$, where optimal homogenization satisfies $ga(0.5-j)=C'(\lambda^j)$, thus $\lambda=\max\{0,\lambda^j-\lambda^r\}$. We now show that the democracy will always locate the government at the center. If $\lambda^r\leq\lambda_1^m$ then the median λ chosen when the government is located at j results in total homogenization of $\lambda+\lambda^r$, equal to the median homogenization in Lemma 1 when the government is located at j. The results of Lemma 1 follow. If $\lambda^r>\lambda_1^m$, then for all $j\in[0.25-x,0.75+x]$, $\lambda=0$ is the median homogenization chosen, where $j=0.25-x$ and $j=0.75+x$ are the governments for which $ga(0.5-j)=C'(\lambda^r)$. Thus $j=0.5$ beats all other j in this interval in a pairwise vote. For any $j\in[0,0.25-x)$ (and symmetrically for $j\in(0.75+x,1]$) the median homogenization chosen by a democracy is $\lambda=\lambda^j-\lambda^r$. The voters $\hat{l}_i\in[0,0.5]$ and $\hat{\hat{l}}_i\in[0.5,1]$ are indifferent between a government at 0.5 and a government at $j\in[0,0.25-x)$ where

$$g-ga(1-\lambda^r)\hat{l}_i+y-k=g-ga(1-\lambda^j)(0.5-j-\hat{l}_i)+y-k-[C(\lambda^j)-C(\lambda^r)]$$

$$g-ga(1-\lambda^r)\hat{\hat{l}}_i+y-k=g-ga(1-\lambda^j)(0.5-j+\hat{\hat{l}}_i)+y-k-[C(\lambda^j)-C(\lambda^r)].$$

then

$$\hat{l}_i+\hat{\hat{l}}_i=\frac{1}{ga}\left(\frac{2(1-\lambda^r)}{(1-\lambda^r)^2-(1-\lambda^j)^2}\right)[C(\lambda^j)-C(\lambda^r)+ga(1-\lambda^j)(0.5-j)]$$

where $C(\lambda^j)-C(\lambda^r)\geq C'(\lambda^r)[\lambda^j-\lambda^r]>ga0.25[\lambda^j-\lambda^r]$ and the result follows as Lemma 1. Thus if a single country is formed, the government is located at the center with additional homogenization $\lambda=\max\{0,\lambda_1^m-\lambda^r\}$. By an analogous argument, if two countries are formed the government is located at the center of each country with additional homogenization $\lambda=\max\{0,\lambda_2^m-\lambda^r\}$. Finally, a democracy will form a single country, following λ^r, if the following is weakly positive and split if it is negative:

$$[g-(1-\lambda_{11})gal_i+y-k-C(\lambda_{11})]-[g-(1-\lambda_{22})ga(0.25-l_i)+y-2k-C(\lambda_{22})]$$

where $l_i = 0.25((1-\lambda_{11})+(1-\lambda_{22}))/2(1-\lambda_{22})$is the median voter following the proof of Proposition 1; $\lambda_{22} = \lambda_2^m$ if $\lambda^r \leq \lambda_2^m$ and $\lambda_{22} = \lambda^r$ otherwise; $\lambda_{11} = \lambda_1^m$ if $\lambda^r \leq \lambda_1^m$ and $\lambda_{11} = \lambda^r$ otherwise. Observe that the only choice in period 1 that directly impacts the outcome in period 2 is the level of homogenization, λ^r. The cost of λ^r is the same to the dictator wherever he locates the government. Thus he optimally locates the government at his ideal point.

Proof of Proposition 6(i)

For notational simplicity we incorporate the cost of homogenizing in period 1 in the period 2 expected utility. Suppose $\lambda^* = 0$. A ruler located at distance l_i will homogenize to $\mu_1^r = \bar{\bar{\mu}}$ if the maximized expression in (21) is greater than the maximized expression in (22) and $\mu_1^r = \bar{\mu}$ if vice versa, where

$$\bar{\bar{\mu}} = \underset{\mu\in[0,\lambda_1^m]}{\arg\max}\, p(g-(1-\lambda_1^m)gal_i+y-k-[C(\lambda_1^m)-C(\mu)])+(1-p)(g+y-k)-M(\mu,0) \quad (21)$$

$$\bar{\mu} = \underset{\mu\in(\lambda_1^m,1]}{\arg\max}\, p(g-(1-\mu)gal_i+y-k)+(1-p)(g+y-k)-M(\mu,0). \quad (22)$$

Homogenization is weakly increasing in p.[80]

Suppose $\lambda^* > 0$. The optimal choice of $\mu_2^r \in [0,\lambda^*)$ is $\tilde{\mu}$ when the maximized utility in (23) is greater than (24) or $\tilde{\tilde{\mu}}$ otherwise, where

$$\tilde{\mu} = \underset{\mu\in[0,\lambda_2^m]}{\arg\max}\, p(g-(1-\lambda_2^m)ga|0.25-l_i|+y-2k-[C(\lambda_2^m)-C(\mu)])+(1-p)(g+y-k)-M(\mu,0) \quad (23)$$

$$\tilde{\tilde{\mu}} = \underset{\mu\in(\lambda_2^m,\lambda^*)}{\arg\max}\, p(g-(1-\mu)ga|0.25-l_i|+y-2k)+(1-p)(g+y-k)-M(\mu,0). \quad (24)$$

Homogenization μ_2^r is again non-decreasing in p. The optimal value of $\mu^r \geq \lambda^*$ is $\hat{\mu} = \max\{\lambda^*,\bar{\mu}\}$ and is non-decreasing in p. The argument to show that if $\hat{\mu}$ is optimal at some p then it is optimal for all p higher follows exactly the argument in the proof of Proposition 4.

Proof of Proposition 7

Let λ_{en} denote the utility maximizing $\lambda \in [0,1]$ for the endogenous dictator. Then the probability of democracy is $p(\lambda_{en})$. We set the probability of democracy in the exogenous case $p_{exg} = p(\lambda_{en})$ also and let λ_{exg} denote the utility maximizing $\lambda \in [0,1]$ for the exogenous dictator when $p = p_{exg}$. The first order conditions in the endogenous case are as follows and proceed in a similar manner to the exogenous case.

Let $\lambda^* = 0$. A ruler located at distance l_i from the center will homogenize to λ_1^r equal to $\tilde{\tilde{\lambda}}$ if his utility from the maximized expression in (25) is greater than his utility from (26), if

[80] Suppose at some p $\mu' > 0$ maximizes (21) but is not necessarily unique. It is straightforward to see that for p higher, μ' gives a strictly higher utility than any $\mu < \mu'$.

not he homogenizes to $\tilde{\lambda}$, where

$$\tilde{\tilde{\lambda}} = \underset{\lambda\in[0,\lambda_1^m]}{\arg\max}\, p(\lambda)(g-(1-\lambda_1^m)gal_i+y-k-C(\lambda_1^m))+(1-p(\lambda))(g+y-k-C(\lambda)). \quad (25)$$

$$\tilde{\lambda} = \underset{\lambda\in(\lambda_1^m,1]}{\arg\max}[p(\lambda)(g-(1-\lambda)gal_i+y-k-C(\lambda))+(1-p(\lambda))(g+y-k-C(\lambda))] \quad (26)$$

Differentiate these two expressions with respect to λ to get respectively

$$-p'(\lambda)((1-\lambda_1^m)gal_i+[C(\lambda_1^m)-C(\lambda)]) = (1-p(\lambda))C'(\lambda). \quad (27)$$

$$-p'(\lambda)(1-\lambda)gal_i+p(\lambda)gal_i = C'(\lambda) \quad (28)$$

In expression (27) $\tilde{\tilde{\lambda}} \geq 0$. Expression (28) has extra positive term on the left hand side compared to the exogenous case. To show when $\lambda^* = 0$ that $\lambda_{exg} \leq \lambda_{en}$, it remains to show that if $\lambda_{en} < \lambda_1^m$ then $\lambda_{ex} < \lambda_1^m$. If $\lambda_{en} < \lambda_1^m$ then

$$p(\tilde{\tilde{\lambda}})(g-(1-\lambda_1^m)gal_i+y-k-C(\lambda_1^m))+(1-p(\tilde{\tilde{\lambda}}))(g+y-k-C(\tilde{\tilde{\lambda}})) > \quad (29)$$
$$p(\tilde{\lambda})(g-(1-\tilde{\lambda})gal_i+y-k-C(\tilde{\lambda}))+(1-p(\tilde{\lambda}))(g+y-k-C(\tilde{\lambda}))$$

If $\lambda_{ex} < \lambda_1^m$ then

$$p(\tilde{\tilde{\lambda}})(g-(1-\lambda_1^m)gal_i+y-k-C(\lambda_1^m))+(1-p(\tilde{\tilde{\lambda}}))(g+y-k) > \quad (30)$$
$$p(\tilde{\tilde{\lambda}})(g-(1-\lambda')gal_i+y-k-C(\lambda'))+(1-p(\tilde{\tilde{\lambda}}))(g+y-k-C(\lambda'))$$

where λ' maximizes the right hand side of (30). First note that the left hand side of (30) is weakly greater than the left hand side of (29). Show that the right hand side of (30) is weakly lower than the right hand side of (29). Since $\lambda' > \lambda_1^m$ then $\lambda' > \tilde{\tilde{\lambda}}$ and so $p(\lambda') < p(\tilde{\tilde{\lambda}})$ so

$$\begin{aligned}
&p(\tilde{\tilde{\lambda}})(g-(1-\lambda')gal_i+y-k-C(\lambda'))+(1-p(\tilde{\tilde{\lambda}}))(g+y-k-C(\lambda')) \\
\leq &p(\lambda')(g-(1-\lambda')gal_i+y-k-C(\lambda'))+(1-p(\lambda'))(g+y-k-C(\lambda')) \\
\leq &p(\tilde{\lambda})(g-(1-\tilde{\lambda})gal_i+y-k-C(\tilde{\lambda}))+(1-p(\tilde{\lambda}))(g+y-k-C(\tilde{\lambda}))
\end{aligned}$$

since $\tilde{\lambda}$ optimizes this expression.

Let $\lambda^* > 0$. A ruler located at distance l_i from the center will homogenize to $\lambda_2^r \in [0, \lambda^*)$ equal to $\bar{\bar{\lambda}}$ if his utility from the maximized expression in (31) is greater than his utility from (32), if not he homogenizes to $\bar{\lambda}$, where

$$\bar{\bar{\lambda}} = \underset{\lambda\in[0,\lambda_2^m]}{\arg\max}\, p(\lambda)(g-(1-\lambda_2^m)ga|l_i-0.25|+y-k-C(\lambda_2^m))+(1-p(\lambda))(g+y-k-C(\lambda)) \quad (31)$$

$$\bar{\lambda} = \underset{\lambda\in(\lambda_2^m,\lambda^*)}{\arg\max}[p(\lambda)(g-(1-\lambda)gal_i+y-k-C(\lambda))+(1-p(\lambda))(g+y-k-C(\lambda))]. \quad (32)$$

Differentiating these expressions with respect to λ gives

$$-p'(\lambda)[(1-\lambda_2^m)ga|l_i-0.25|+[C(\lambda_2^m)-C(\lambda)]=(1-p(\lambda))C'(\lambda)$$

$$-p'(\lambda)(1-\lambda)gal_i+p(\lambda)gal_i=C'(\lambda). \tag{33}$$

as above $\bar{\bar{\lambda}} \geq 0$ and (33) has an extra positive term compared to the exogenous case. It remains to show that if $\lambda_{en} < \lambda^*$ then also $\lambda_{ex} < \lambda^*$, and if $\lambda_{en} \leq \lambda_2^m$ then also $\lambda_{ex} \leq \lambda_2^m$.

First show that if the endogenous dictator chooses $\lambda_{en} \leq \lambda_2^m$ then $\lambda_{ex} = 0$. The argument follows exactly as above. Now show that if $\lambda_{en} < \lambda^*$ then also $\lambda_{ex} < \lambda^*$. Examine the choice of the ruler between homogenizing to $\hat{\lambda} = \max\{\lambda^*, \tilde{\lambda}\}$, where $\tilde{\lambda}$ is as given in (26), or $\bar{\lambda}$. If he chooses $\bar{\lambda}$ then

$$p(\bar{\lambda})[g-(1-\bar{\lambda})ga|0.25-l_i|+y-2k-C(\bar{\lambda})]+(1-p(\bar{\lambda}))[g+y-k-C(\bar{\lambda})] > \tag{34}$$

$$p(\hat{\lambda})[g-(1-\hat{\lambda})gal_i+y-k-C(\hat{\lambda})]+(1-p(\hat{\lambda}))[g+y-k-C(\hat{\lambda})].$$

If an endogenous dictator prefers $\bar{\lambda}$ to $\bar{\bar{\lambda}}$ then for the exogenous ruler we could have either $\lambda_2^r = 0$ or $\lambda_2^r > \lambda_2^m$. Suppose $\lambda_2^r > \lambda_2^m$ then an exogenous dictator prefers to split into two if

$$p(\bar{\lambda})[g-(1-\lambda')ga|0.25-l_i|+y-2k-C(\lambda')]+(1-p(\bar{\lambda}))[g+y-k-C(\lambda')] > \tag{35}$$

$$p(\bar{\lambda})[g-(1-\hat{\hat{\lambda}})gal_i+y-k-C(\hat{\hat{\lambda}})]+(1-p(\bar{\lambda}))[g+y-k-C(\hat{\hat{\lambda}})]$$

where λ' maximizes the left hand side of (35) and $\hat{\hat{\lambda}} = \max\{\lambda^*, \lambda\}$ where λ maximizes the right hand side of (35) given $p(\bar{\lambda})$. It is clear that the left hand side of (35) is weakly higher than the left hand side of (34) since the exogenous dictator can always set $\lambda = \bar{\lambda}$. The right hand side of (35) is weakly lower than the right hand side of (34) since $\hat{\hat{\lambda}} > \bar{\lambda}$ and so $p(\hat{\hat{\lambda}}) < p(\bar{\lambda})$ then using the same argument as earlier. If the exogenous dictator does better by choosing $\lambda = 0$ instead of $\lambda' > \lambda_2^m$ then the same argument holds. The same argument also holds if the endogenous dictator does better by choosing $\bar{\bar{\lambda}} \leq \lambda_2^m$ rather than $\bar{\lambda}$.

Proof of Proposition 8

The first order conditions are as given above. The first order conditions in the endogenous case have an extra negative term on the left hand side and the λ that satisfies each expression is weakly lower than when $p'(\lambda) = 0$.

First show that when $\lambda^* = 0$ if $\lambda_{en} > \lambda_1^m$ then $\lambda_{ex} > \lambda_1^m$. If $\lambda_{en} > \lambda_1^m$ then

$$p(\tilde{\lambda})(g-(1-\tilde{\lambda})gal_i+y-k-C(\tilde{\lambda}))+(1-p(\tilde{\lambda}))(g+y-k-C(\tilde{\lambda})) > \tag{36}$$

$$p(\tilde{\tilde{\lambda}})(g-(1-\lambda_1^m)gal_i+y-k-C(\lambda_1^m))+(1-p(\tilde{\tilde{\lambda}}))(g+y-k-C(\tilde{\tilde{\lambda}}))$$

If $\lambda_{ex} > \lambda_1^m$ then

$$p(\tilde{\lambda})(g-(1-\lambda')gal_i+y-k-C(\lambda'))+(1-p(\tilde{\lambda}))(g+y-k-C(\lambda'))> \tag{37}$$

$$p(\tilde{\lambda})(g-(1-\lambda_1^m)gal_i+y-k-C(\lambda_1^m))+(1-p(\tilde{\lambda}))(g+y-k)$$

Clearly the left hand side of (36) is lower than the left hand side of (37). From the first order condition we know $\tilde{\tilde{\lambda}}=0$ and since $\tilde{\lambda}>\tilde{\tilde{\lambda}}$ then $p(\tilde{\lambda})>p(\tilde{\tilde{\lambda}})$, and the right hand side of (36) is higher than the right hand side of (37).

Now if $\lambda^*>0$ it remains to show that if $\lambda_{en}\geq\lambda^*$ then $\lambda_{ex}\geq\lambda^*$ and if $\lambda^*>\lambda_{en}>\lambda_2^m$ then $\lambda_{ex}>\lambda_2^m$. Suppose the endogenous dictator has chosen $\lambda^*>\lambda_{en}>\lambda_2^m$ then

$$p(\bar{\lambda})[g-(1-\bar{\lambda})ga|0.25-l_i|+y-2k-C(\bar{\lambda})]+(1-p(\bar{\lambda}))[g+y-k-C(\bar{\lambda})]. \tag{38}$$

$$>p(\bar{\bar{\lambda}})[g-(1-\lambda_2^m)ga|0.25-l_i|+y-2k-C(\lambda_2^m)]+(1-p(\bar{\bar{\lambda}}))[g+y-k]$$

where $\bar{\bar{\lambda}}=0$ from the first order condition. The exogenous dictator chooses $\lambda_{ex}>\lambda_2^m$ if

$$p(\bar{\lambda})[g-(1-\lambda')ga|0.25-l_i|+y-2k-C(\lambda')]+(1-p(\bar{\lambda}))[g+y-k-C(\lambda')] \tag{39}$$

$$>p(\bar{\lambda})[g-(1-\lambda_2^m)ga|0.25-l_i|+y-2k-C(\lambda_2^m)]+(1-p(\bar{\lambda}))[g+y-k]$$

where λ' maximizes the right hand side of the inequality. The left hand side of (39) is weakly greater than the left hand side of (38) since the exogenous dictator could always choose $\bar{\lambda}$. The right hand side of (39) is lower than the right hand side of (38) since $p(\bar{\lambda})>p(\bar{\bar{\lambda}})$.

Examine the choice of the exogenous ruler between homogenizing to $\hat{\lambda}=\max\{\lambda^*,\tilde{\lambda}\}$ where $\tilde{\lambda}$ satisfies (26) or λ_2^r. If $\lambda_2^r>\lambda_2^m$ and he chooses $\hat{\lambda}$ then

$$p(\hat{\lambda})[g-(1-\hat{\lambda})gal_i+y-k-C(\hat{\lambda})]+(1-p(\hat{\lambda}))[g+y-k-C(\hat{\lambda})] \tag{40}$$

$$>p(\bar{\lambda})[g-(1-\bar{\lambda})ga|0.25-l_i|+y-2k-C(\bar{\lambda})]+(1-p(\bar{\lambda}))[g+y-k-C(\bar{\lambda})].$$

An exogenous dictator prefers $\lambda>\lambda^*$ if

$$p(\hat{\lambda})[g-(1-\hat{\bar{\lambda}})gal_i+y-k-C(\hat{\bar{\lambda}})]+(1-p(\hat{\lambda}))[g+y-k-C(\hat{\bar{\lambda}})] \tag{41}$$

$$>p(\hat{\lambda})[g-(1-\lambda')ga|0.25-l_i|+y-2k-C(\lambda')]+(1-p(\hat{\lambda}))[g+y-k-C(\lambda')].$$

where $\hat{\bar{\lambda}}=\max\{\lambda^*,\lambda\}$ and λ maximizes the left hand side of (41) given $p(\hat{\lambda})$, and λ' maximizes the right hand side (note if the endogenous ruler chooses $\bar{\lambda}$ then $\lambda_{ex}>\lambda_2^m$ as discussed above). The left hand side of (41) must be weakly higher than (40) since $\hat{\lambda}$ could always be chosen by the exogenous ruler. The right hand side of (41) is lower than the right hand side of (40):

$$p(\hat{\lambda})[g-(1-\lambda')ga|0.25-l_i|+y-2k-C(\lambda')]+(1-p(\hat{\lambda}))[g+y-k-C(\lambda')]$$

$$\leq p(\lambda')[g - (1-\lambda')ga|0.25 - l_i| + y - 2k - C(\lambda')] + (1 - p(\lambda'))[g + y - k - C(\lambda')]$$
$$\leq p(\bar{\lambda})[g - (1-\bar{\lambda})ga|0.25 - l_i| + y - 2k - C(\bar{\lambda})] + (1 - p(\bar{\lambda}))[g + y - k - C(\bar{\lambda})]$$

since $\bar{\lambda}$ maximizes the expression. Suppose instead for the endogenous ruler $\lambda_2^r = 0$, then the right hand side of (40) is

$$p(0)[g - (1-\lambda_2^m)ga|0.25 - l_i| + y - 2k - C(\lambda_2^m)] + (1 - p(0))[g + y - k].$$

Then if the exogenous dictator chooses zero the right hand side of (41) is strictly lower since everything is the same except the probability of overthrow is higher. If the exogenous dictator in this case chooses $\lambda > \lambda_2^m$ then the right hand side of (41) is the same as above but we already know this is less than the right hand side of (40) by the same argument as for $\bar{\lambda}$.