NBER WORKING PAPER SERIES

NATION BUILDING

Alberto Alesina
Bryony Reich

Working Paper 18839
http://www.nber.org/papers/w18839

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2013

We thank Enrico Spolaore and participants of a seminar at UCL and Warwick for useful comments. Andrea Passalacqua provided excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Nation Building
Alberto Alesina and Bryony Reich
NBER Working Paper No. 18839
February 2013
JEL No. F3

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, 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 compare both regimes in a model where the size of countries and the degree of active homogenization in endogenous. We also offer some historical discussions of cases which illustrate our theoretical results.

Alberto Alesina
Department of Economics
Harvard University
Littauer Center 210
Cambridge, MA 02138
and IGIER
and also NBER
aalesina@harvard.edu

Bryony Reich
Department of Economics
University College London
Drayton House
30 Gordon Street
London
WC1H 0AX
b.reich@ucl.ac.uk

Nation-building*

Alberto Alesina
Harvard and IGIER Bocconi

Bryony Reich
University College London

First Draft: October 2012
Latest Revision: February 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, 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

At the time of the unification of Italy (1860) only 10% of the population spoke Italian.[1] As Massimo d'Azeglio (one of the founders of unified Italy) acknowledged: 'Italy has been made; now it remains to make Italians.' European nations, far from having preexisting nationwide characteristics, took steps to develop national identities in the 19th Century. Tilly and Tilly (1973) observe 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 elite usually came to be the official language of modern states via public education and other administrative mechanisms.' More recent history also provides examples of dictators who prohibited local cultures, often by odious means (for example the Soviet Union, China, or Franco's Spain), in order to strengthen their grip on society. The goal of this paper is to analyze the process of "homogenization" within a country, in its more or less benevolent forms. To be absolutely

*We thank Enrico Spolaore and participants of a seminar at UCL and Warwick for useful comments. Andrea Passalacqua provided excellent research assistance.

[1] Duggan (2007), p. 108.

clear, we define "nation-building" as a process that 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.

We model the equilibrium size of a country as emerging 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).[2] We depart from them, however, since we assume that the degree of divergence of preferences amongst the population is endogenous: we explicitly model the choice of the central government of how much to homogenize the population. Heterogeneity can be reduced through, for example, public education in a common language, better transport systems, or by less benevolent means such as repression of minorities and minority culture. A reduction in heterogeneity reduces the extent of the size-heterogeneity trade-off, that is it allows larger populations to stay together.[3]

When and why would a particular regime undertake such homogenization? First consider a democracy. Within a country the population only has access to one set of public goods, a catch-all term for what a public sector, i.e. "the government", 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. A democracy behaves in the interests of the majority and the latter benefits from a certain degree of homogenization. For example, better roads to the capital city improve the individuals' access to resources located there, while schooling in a common language enables better participation in the democratic process. However, since homogenization (schooling, roads, etc.) is costly, the median voter will choose to homogenize up to the point in which marginal benefits equal marginal costs. In some cases the median voter might choose to homogenize to avoid a split of the nation. For instance if it were impossible to learn each others' language a country would split, but with public schools that teach a common language the country may stay together.[4] Alternatively, communication networks may avoid distant minorities becoming isolated and disenfranchised. 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.

In comparison with a democracy, a dictator or ruling elite who is in full control of the population and faces little probability of being overthrown has different incentives. The dictator or elite will choose a "government" which matches their own preferences and 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 faces little prospect of being overthrown, thus homogenization is useless since policies and public goods already match the preferences of those in charge.

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

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

[4] 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.

The incentives of a dictator or ruling elite who faces a more substantial probability of overthrow (and the establishment of a democracy) are different again.[5] As above, they will choose public goods to perfectly match their own preferences and will tax the maximum number of people to pay for these public goods. However, they face a significant possibility of being overthrown by a democratic movement and therefore they face the prospect of living under a democratic regime in the future. The latter may choose public goods that differ from the preferences of the dictator or ruling elite; 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 allows those in charge to better preserve the status quo (their preferred policies and a single country). Thus when a dictator or elite faces a high probability of democracy and the government and borders chosen by a democracy would represent a significant deviation from the preferences of the ruling group, the latter will undertake homogenization in excess of anything that would be undertaken by a democracy. We also 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 cost on "distant" minorities, i.e. on individuals very far from the dictator's preferences/location. We show how dictators would always chose odious types of homogenization.

We then discuss cases in which the choice of homogenization of the ruler directly affects the probability of success of a democratic revolution. In one case more homogenization, when it increases welfare, may reduce the incentive to overthrow the government. 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.

In the final part of the paper we discuss some cases and review some evidence that relates to the theoretical results of the paper regarding more or less successful policies of homogenization in democracies and dictatorships in recent years and in earlier centuries.

We are not aware of any paper directly modelling endogenous homogenization, but our paper relates to the literature on the need for education in the better functioning of institutions, as in Glaeser, Shleifer and Ponzetto (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.

The results proceed as follows. Section 2 introduces the model. Section 3 examines the case of a democracy and Section 4 the case of dictators facing various probabilities of overthrow. Section 5 compares the amount of homogenization and incentives to homogenize between the different regimes. Section 6 compares different homogenization technologies. Section 7 examines an endogenous probability of democracy. Section 8 discusses historical examples and the last section concludes.

[5] One could also consider the risk of a dictator being overthrown by another dictator. We leave this point for future research.

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 countries, A and B, comprising the intervals $[0, 0.5]$ and $(0.5, 1]$ respectively.[6] Each country has a government which is located at some point $j \in [0, 1]$ inside that country. Denote by d_{ij} the distance of individual $i \in [0, 1]$ from government j in his country. By "government" we mean a set of public goods provided by an authority. We can 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 (which are funded by taxes) is k.[7] 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.[8] However, when a population splits into two countries, A and B, they are more homogenous and the policies/public goods provided in those countries can be closer to the average individual. Thus while a larger country potentially reduces the costs of public goods, increases efficiency and reduces costs of trade, the more diverse a country is the harder it is for public choices to satisfy everyone. This models a trade-off between size and heterogeneity and is the set-up used by Alesina and Spolaore (1997) to investigate equilibrium country size.

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 reduction in a within the population, where the homogenization technology is as follows. Homogenization uses government apparatus and public goods, to shift people's location in the direction of the government. Diversity within a given country can be reduced by fraction λ, where $0 \leq \lambda \leq 1$. 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, to $(1 \square \lambda)ad_{ij}$. We refer to this as homogenization of the population. A choice of $\lambda = 0$ implies no homogenization, while a choice of $\lambda = 1$ implies complete homogenization of the population such that all individuals in the country have the same ideal point and/or the cost of geographical distance is nil. Homogenization is costly: homogenizing a population of size s by λ costs $sC(\lambda)$ and the cost is born equally within the population. 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*

[6] The restriction of having at most two countries is adopted only to keep the analysis simple while still allowing for endogenous country size.

[7] 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.

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

$\lim_{\lambda \to 1} C'(\lambda) = \infty$.

Individual i's utility function includes the value he gets from the government/public good that he accesses, income (consumption in a static model) y, and taxes t (which are the same for all agents, a natural assumption since we do not focus on income differences).[9] Individual i in a country of size $s \in \{1/2, 1\}$ with the government located at j has utility:

$$U_i = g(1 \square (1 \square \lambda) a d_{ij}) + y \square t, \tag{1}$$

where $k + sC(\lambda) \leq st$. The first term measures the value of the government for individual i. We let g denote the maximum value of the public goods. The term $(1 \square \lambda) a d_{ij}$ measures how far public goods or the policies of the government are from individual i's ideal (following homogenization of degree λ). The benefit that individual i receives from the public goods g is reduced accordingly. Taxes include the cost of public goods k and the cost of the homogenization of degree λ and are split equally amongst the population of the country. Note that "homogenization", i.e. a higher λ increases welfare for the individual (although at a cost).

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 the travel costs of accessing 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 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.[10] In communist countries actual indoctrination in schools was common. For instance, Alesina and Fuchs-Schündeln (2007) present evidence of a large amount of indoctrination in East Germany. One can choose the preferred interpretation of homogenization. In order to mantain all three together, as emphasized above, one needs to make the same assumption as in Alesina and Spolaore (1997, 2003), namely 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.

Decision-making procedes as follows. Whoever is in control of the population (be it a dictator, elite or the population as a whole) has to decide, 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. We study the case of a democracy first and then the case of a non-democratic ruler.

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

[10] See Alesina and Glaeser (2004) for a discussion of these cultural differences.

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). This in fact may be the reason why different forms of forceful integration are often used instead of monetary transfers.[11]

3 Democracy

In a democracy decisions are made by majority rule with the timing of votes described above.

First note that given λ, the payoff to individual i at distance d_{ij} from the government j in a country of size $s \in \{1/2, 1\}$ is:

$$g \square (1 \square \lambda) g a d_{ij} + y \square k/s \square C(\lambda) \tag{2}$$

Let λ_{ij} be the optimal level of homogenization for person i :

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

The first order condition is:

$$g a d_{ij} = C'(\lambda_{ij}).$$

This condition implies that the marginal cost of homogenization has to equal the marginal benefit which depends on the distance of individual i from the government. Individuals who are further from the center will 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 most or more intense teaching of a national language for minorities with very different languages). The alternative case in which more costs are born by those further away from the center is analyzed below. The interpretation of 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 trasmit values and educate their children in such a way which make their children fit better in the country where they live.[12]

[11] Alesina and Spolaore (1997) discuss the issue of time inconsistency related to (reneged) promises of transfers to the periphery to keep the latter within the country. Bolton and Roland (1997) discuss the issue of rich regions separating from poor ones to avoid taxation.

[12] For models related to parents "choosing" values for children see Alesina et al. (2013) and Bisin and Verdier (2000)

First note that whether the population forms a single country or splits into two, 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.

Second, in any given country locating the government in the middle of the country beats all other possible locations in pairwise votes. See Lemma 1 in the appendix for a proof.[13] If the population forms only one country the government is located in the center at 0.5 and with two countries the two governments would also be located at the centers at 0.25 in country A and 0.75 in country B.

We can now apply the median voter theorem to the choice of λ. The level of homogenization chosen when the population forms a single country is the optimal one for the individual at median distance from the central government, that is $i = 0.25$.[14] Denote this level λ_1^m. The level of homogenization chosen when the population splits into two countries is the optimal level for the individual at median distance from the central government in country A, that is individual $i = 0.125$, and the individual at median distance from the central government in country B, that is individual $i = 0.5 + 0.125$.[15] The amount of homogenization undertaken in A and B is identical and we denote this level of homogenization 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 heterogenous populations homogenize more.

It follows that 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 a government at the center of either country and each country homogenizing to level λ_2^m. To be able to compare, for a given individual i, the value of forming a single country relative to splitting into two countries, we 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:[16]

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

which can be rewritten as

$$(1 \square \lambda_2^m)ga|0.25 \square l_i| \square (1 \square \ \lambda_1^m)gal_i + k \square [C(\lambda_1^m) \square C(\lambda_2^m)]\,. \tag{3}$$

The median voter is the voter with the median valuation of forming a single country relative to splitting into two.

[13] 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.

[14] There is of course a second median voter, at $i = 0.75$.

[15] There is another median voter at $i = 0.375$ in country A with the same preferences over homogenization and another median voter at $i = 0.5 + 0.375$ in country B.

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

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

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

from the center of the population. A democracy chooses to organize itself as one country and homogenize to degree λ_1^m when expression (3) is positive as evaluated for the median voter. A democracy chooses to split into two countries and homogenize to degree λ_2^m when expression (3) is negative as evaluated for the median voter.

Note that we have three decisions in this model with three different median voters. The decision of where to locate the government, with the median voter in the middle of the country; the decision of how much to homogenize, with median voters λ_1^m or λ_2^m; and the decision of whether to have one or two countries, with the median voter given by Proposition 1.[17] The median voter given in Proposition 1 is an individual $i \in (0.25, 0.375)$.[18]

Expression (3) encompasses three considerations for the median voter: how much closer his ideal point 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. Clearly *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, the more the median voter will prefer a single country.

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.[19] 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, i.e. "nation-building".

Finally note that we do not allow for unilateral secessions, namely a situation in which without any majority vote a group of citizens form a third country.

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

[17] 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.

[18] 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 $i = 0.5$).

[19] See appendix *A2* for a proof.

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 within the unit population.[20] Such an extension would complicate notation and algebra with little advantage in terms of insight. Thus from now on we will refer to the ruler (or the dictator) as an individual who could also represent a ruling elite.

We begin by analyzing the case of a ruler located with his ideal point at the median ideal point in the population, $i = 0.5$. Below we generalize this assumption. The ruler faces the same choices as a democracy: how to partition the population, where to locate the government, and how much homogenization to undertake.

Proposition 2 *The ruler forms a single country, locates the government at his ideal point, and undertakes no homogenization.*

The ruler forms a single country (thus reaping the benefits of a larger country) and 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 and he already has public goods which perfectly match his preferences/location.

The dictator locates the government in the same place as a democratic median voter would do in a single country. The difference with respect to a democracy is twofold: the degree of homogenization is not chosen democratically. In fact $\lambda = 0$ with the dictator while $\lambda_1^m, \lambda_2^m > 0$ in a democracy. Second, the dictator would not allow a democratic choice of whether to split the country. By forming a single country the dictator has his "ideal government" at lowest per capita costs possible. Indeed this incentive would be reinforced if the dictator could also extract rents from the population, that is he can raise taxes in addition to those needed to pay for k.[21]

4.2 An unsafe dictator

Most dictators however, are not absolutely safe in office. In particular we examine in this section 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:

- In period 1 the ruler determines the location of borders, governments and the level of homogenization λ. He knows that with probability p democracy will prevail in period 2.
- In period 2 either democracy is realized 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 and continues to make all decisions.

We assume homogenization is permanent: if the dictator homogenizes a population by λ in the first period then the heterogeneity of the population is reduced by λ from that point

[20] Results on this point are available from the authors.

[21] See Alesina and Spolaore (1997, 2003) for this extension.

onwards (that is in both the first period and second period). Any further homogenization may be undertaken by whoever is in power in the second period.[22] The cost of homogenization is paid in the period that it takes place. Specifically, the utility of individual i in period 1 in a country of size s with the government at j when homogenization λ takes place is

$$U_i = g(1 \square (1 \square \lambda)ad_{ij}) + y \square t, \tag{4}$$

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

$$U_i = g(1 \square (1 \square \lambda \square \lambda')ad_{ij}) + y \square t. \tag{5}$$

where $k + s[C(\lambda + \lambda') \square C(\lambda)] \leq st$. The costs of homogenization in the second period simply reflect that total costs of the total amount of homogenization undertaken are the same whatever date the homogenization takes place. In period 1 the ruler chooses borders, government and homogenization to maximize his total expected utility, which is the sum of his utility from period 1 and his expected utility from period 2 given probability p of overthrow (assuming no discounting of the future period). Since period 2 is the final period, if the ruler is in power he simply maximizes his period 2 utility. If democracy prevails in period 2 then the population as a whole votes by majority rule over country size, the location of the government and any further homogenization λ', given the homogenization undertaken in period 1.

Let us now analyze the choices of the ruler. A ruler faces two possible outcomes if democracy prevails: either the democracy forms a single country or the democracy splits into two countries A and B, with the government located at the center of the respective countries. A democracy that splits is worse for the dictator for two reasons: first, taxes are higher and second, the dictator located at 0.5 no longer enjoys his preferred government. A dictator, however, can homogenize in the first period to mitigate his situation in the second period should democracy prevail.

Proposition 3

When in power the ruler forms a single country and locates the government at his ideal point. In period 1 *the ruler undertakes homogenization* $\lambda^r \in [0, 1]$ *of the population and in period* 2*, if in power, he undertakes no homogenization. The homogenization undertaken by the ruler,* λ^r*, is weakly increasing in the probability of democracy,* p*.*

There exists a minimum level of homogenization which a dictator must undertake, denoted λ^*, such that the population (now of heterogeneity $(1 \square \lambda^*)a$) would vote to form a single country rather than split into two were the democratic regime to prevail in the second period. Thus 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

[22] There is never an incentive to reverse homogenize.

taxes.[23] Suppose instead that the dictator does not homogenize sufficiently ($\lambda < \lambda^*$) then, if democracy prevails, the population will split into two countries. In this situation, homogenization of the population (even if not enough to prevent the population from splitting) still benefits the dictator because homogenization moves population preferences towards his own preferences and so moves the position of the future democratic government closer to his own ideal point. Note that if the population splits into two democracies the former dictator is at the border between the two and thus prefers a higher level of homogenization than the median voter of the two democracies. The ruler faces the following trade-off: homogenization is costly and as we know he would not do any if $p = 0$, but by homogenizing the ruler can influence changes to governments and borders that occur when democracy prevails.

The intuition for Proposition 3 is clear: the higher the probability of democracy the worse the expected utility of the dictator and the more homogenization he will undertake to maintain his vested interest. Specifically, the ruler's choice of how much to homogenize in the first period is as follows. If the ruler homogenizes sufficiently such that the population would vote to form a single country should democracy prevail then the utility of the ruler in the second period is

$$g + y \square k, \tag{6}$$

whether democracy prevails or not. The cost of homogenization undertaken in the first period is $C(\lambda^r)$ and is paid in period 1. In this case the ruler would optimally set homogenization in the first period equal to λ^* and no higher since homogenization is costly and useless to the ruler once he has ensured a democracy will not split.[24] If the ruler homogenizes to degree $\lambda \in [0, \lambda^*)$, then the expected utility of the ruler is the probability of democracy, p, multiplied by the utility to the ruler if the population splits into two countries, plus the probability the ruler remains in control, $(1 \square p)$, multiplied by his utility in that case, with just one country:

$$p(g \square (1 \square \bar{\lambda})ga0.25 + y \square 2k \square [C(\bar{\lambda}) \square C(\lambda^r)]) + (1 \square p)(g + y \square k), \tag{7}$$

where $\bar{\lambda} = \lambda_2^m$ when $\lambda^r < \lambda_2^m$ and $\bar{\lambda} = \lambda^r$ otherwise. The cost of homogenization undertaken in the first period is $C(\lambda^r)$ and is paid in period 1.[25] Observe that $\lambda^* = 0$ implies the population would choose to form a single country should democracy prevail (with no homogenization needed by the dictator). The ruler chooses the optimal level of homogenization on assessing his expected utility for each $\lambda \in [0, 1]$ given by expressions (6) and (7) and the homogenization costs.

The level of homogenization chosen by a dictator depends on two things: the probability of democracy, and whether the parameters are such that were democracy to prevail (and

[23] In this case the utility of the ruler is the same as his utility when he is in power, simply because we assume the dictator takes no personal rents, an assumption which can be easily generalized, for instance for the case in which when in office as a dictator he can extract rents, i.e. taxes above and beyond k.

[24] Note however that if $\lambda^* < \lambda_1^m$ then cost $C(\lambda^*)$ is paid in the first period and with probability p, if democracy prevails, cost $C(\lambda_1^m) \square C(\lambda^*)$ is paid in the second period.

[25] The reason expression (7) seems heavy on notation is that if the level of homogenization, λ, undertaken by the ruler is less than that which would be undertaken by a democracy if the democracy splits into two (λ_2^m), then if democracy prevails the level of homogenization undertaken will be λ_2^m and not λ and this must be incorporated into any decision by the ruler.

supposing the dictator had undertaken no homogenization) the population would choose to form a single country or split into two countries. First, suppose a democracy would choose to form a single country anyway even if the ruler undertakes no homogenization (this implies $\lambda^* = 0$).[26] Then the optimal degree of homogenization for the ruler is zero whatever the probability of democracy. This result depends strictly on the ruler being located in the middle of the country. Second, suppose a democracy would choose to split into two countries were the ruler to undertake no homogenization (this implies $\lambda^* > 0$). If democracy prevails the ruler faces a particularly negative outcome and can mitigate this outcome by homogenizing the population, but he must balance the cost of homogenizing with the probability democracy prevails. When p is low the dictator undertakes no homogenization. This is because should democracy prevail and the population split they will undertake homogenization λ_2^m anyway and the the dictator does not want to pay any unnecessary costs of homogenizing since he thinks it very likely he will remain in power for the second period. At a higher probability of democracy, homogenization undertaken by the dictator may jump up from zero to $\lambda^r > \lambda_2^m$. The probability of overthrow is high enough that the dictator is willing to pay some costs in the first period to mitigate the outcome of democracy in the second. The parameters of the model determine exactly how much homogenization is undertaken. It is possible that at some p homogenization jumps from zero to λ^r where $\lambda^* > \lambda^r > \lambda_2^m$, then as the probability of democracy continues to rise, the amount of homogenization undertaken increases continuously until, at some higher probability of democracy, the optimal homogenization may jump up again this time to λ^*. There is a second jump here for the following reason: as λ moves continuously from below λ^* to λ^* there is a jump upwards in the expected utility of the dictator (because now if democracy prevails the population will not split), thus as p increases and λ increases towards λ^* at some point the marginal benefit of homogenizing to λ^* is very high and so homogenization levels jump up. Homogenization in the case of an unsafe dictator is characterized by zero homogenization at low probabilities of overthrow with jumps to higher levels of homogenization at higher probabilities of overthrow.

4.2.1 Colonizers

Note that we are assuming that the dictator will stay 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. Similarly, certain rulers may be more likely to be expelled from the country if democracy prevails. For example, an oppressive dictator may be more likely to be expelled compared to a ruling elite which expects to remain in the country after democratization.

Suppose with probability $p \in [0, 1]$ democracy prevails in period 2 and with probability $q \in [0, 1]$, given democracy prevails, the ruler leaves the country. We would expect q to be at or close to 1 for a colonizer who would return home should democracy prevail and the country become independent. Thus the ruler's expected utility in period 2, if he undertakes

[26] Expression (3) is positive for the median democracy voter, see Proposition 1, and so $\lambda^* = 0$.

homogenization $\lambda^r = \lambda^*$ in period 1 to ensure the population forms a single country should democracy prevail, is

$$p(1 \square q)(g + y \square k) + (1 \square p)(g + y \square k).$$

The cost of this homogenization, $C(\lambda^*)$, is paid in the first period.[27] The ruler's expected utility in period 2 if he undertakes homogenization $\lambda^r < \lambda^*$ so the population splits if democracy prevails is:

$$p(1 \square q)(g \square (1 \square \bar{\lambda})ga0.25 + y \square 2k \square [C(\bar{\lambda}) \square C(\lambda^r)]) + (1 \square p)(g + y \square k),$$

where $\bar{\lambda} = \lambda_2^m$ when $\lambda^r < \lambda_2^m$ and $\bar{\lambda} = \lambda^r$ otherwise. The cost of homogenization undertaken in the first period is $C(\lambda^r)$ and is paid in period 1.

It is clear from the above expressions that a higher value of q, the probability the ruler leaves the country when democracy prevails, reduces the incentives of the ruler to homogenize. For $q = 1$ (the case of a colonizer that leaves the country after a revolution) the ruler will undertake no homogenization.

5 Comparing Regimes

Proposition 3 implies that the greater the threat of democracy faced by a dictator the more he homogenizes. But will he homogenize more or less than a democracy would? A safe ruler who faces a low probability of democracy will undertake less homogenization of the population he rules over than the amount of homogenization that would be undertaken if instead the population formed a democracy. Similarly, a colonizer who faces any probability of overthrow but faces only a low probability of remaining in the country after democracy prevails will undertake less homogenization than an "equivalent" democracy.

Proposition 4 compares, for identical populations and costs, the amount of homogenization undertaken by a democracy with that undertaken by a ruler who faces a high probability of overthrow and expects to remain in the country:

Proposition 4 *When $\lambda^* > 0$, an unsafe ruler who faces a high probability of overthrow will undertake more homogenization of the population than a democracy.*

Corollary 1 *When $\lambda^* > 0$ a population that forms a democracy would choose to split into two. If instead the population was controlled by an unsafe ruler they may be homogenized to such an extent that, should democracy prevail in the future, the population would then be homogenous enough that they 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 dictator will be more homogenous,

[27] To illustrate the point without a lot of notation we have assumed $\lambda^* > \lambda_1^m$. If this was not the case the expression would simply be

$$p[(1 \square q)(g + y \square k \square [C(\lambda_1^m) \square C(\lambda^*)])] + (1 \square p)(g + y \square k)$$

with cost $C(\lambda^*)$, being paid in the first period.

perhaps so much so that they are homogenous enough to form a single country. The implication is that as a result of "nation-building" by dictators or elites, today's democracies, which followed hundreds of years of dictatorship and control by elites, may be bigger than they would otherwise be had democracies prevailed since the beginning.[28]

If $\lambda^* > 0$ the dictator does particularly badly if democracy arises because the population chooses to split into two equally sized countries and the former dictator is at the border between the two. Thus, the dictator knows he will have to undertake costly homogenization if he wants to ensure the population are sufficiently homogenous that they would instead choose to form a single country. He may be willing to undertake that level of homogenization to avoid the split, whereas a democracy would not. Why is this? While the dictator's ideal is the median ideal point in a democracy, the median person in a democracy when considering whether to form a single country or split into two is not the person in the middle. Also, if the costs of homogenizing are very high, it will be too costly even for the dictator to homogenize enough, in which case, should democracy prevail, the population would split into two. The dictator is now at the border of Country A and so would prefer more homogenization than the median voter (λ_2^m) in Country A. Under this situation a dictator still homogenizes by more than a democracy because he wants to move population preferences closer to his own preferences, and thus a future government closer to his own ideal government, even if it is not enough to ensure a single country.

The comparison of homogenization across different regimes presented in this section implies different outcomes depending on the path of rule of a country. A country that is controlled by a ruling group who foresees democracy may be homogenized to the greatest extent. Indeed more than would be possible under a democracy, since a democracy is constrained by the median voter. However, a country that is under a relatively safe ruling group or is colonized will not be homogenized much if at all. It follows that a country that jumps from a relatively safe dictatorship or colonization straight to democracy may be less homogenous and more likely to break up than a country that goes through a "slow" transition to democracy.

5.1 Varying the location of the dictator

Thus far we have assumed that a dictator is located in the middle, at 0.5. This ensures that if a democracy prevailed the dictator would prefer a large country over two small ones. Suppose however that a dictator is not located at the center of the population. Changing the dictator's ideal point varies his incentives to homogenize for the reason that it changes how far his ideal policies are from the policies that would be chosen by a democracy.

To give some examples, suppose a dictator was located close to the extreme of the population (close to 0 or 1). We can think of this as an ethnic minority which controls a large diverse country. In this case the 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. If the risk of democracy is high, his incentives to homogenize are especially strong, and he undertakes

[28] Alesina and Spolaore (2003) make a similar point but justify it by adjustment costs of changing borders when a dictatorship breaks up. Here we get the result even without any costs of adjusting borders. The latter would simply reinforce this result.

more homogenization than a democracy both when the population splits into two and when it stays together.

On the other hand, suppose a dictatorial minority was located very close to the government of country A. Obviously this dictatorial minority prefers a dictatorship since they can set the government in their preferred position and extract taxes from the whole population. However, if a democracy prevailed the former dictator would be located closer to the new democratic government if the population were to split than if the country stayed together. Thus in that case the former dictator might prefer for the single democracy to split, at least for some parameter values. This case is analyzed in Appendix $A3$. It should be noted that when we allow the dictator to implement less uniform homogenization policies (see section 6) even in the situation where his ideal point is located at the government of Country A he may homogenize more than a democracy and ensure the population forms a single country.

In summary, the location of the dictator matters because it determines how far the policies of a democratic government would be from the dictator's preferences. The results of Proposition 4 carry through to a dictator located elsewhere as follows: if the government chosen by a democracy would be far from the dictator's ideal then when the probability of democracy is high he will undertake more homogenization of the population than the amount of homogenization that would be undertaken if instead the population formed a democracy.

6 Odious Homogenization

Thus far we have assumed that apart from the fiscal cost in the government budget constraint there are no other costs of homogenization for the population.

We modeled non-odious homogenization as a homogenization technology where the cost is spread equally across the population. Odious homogenization is a technology that costs more to those who are homogenized more, in effect incorporating a personal cost. For example, minorities who do not speak the language of the ruler will be made to pay more to learn the language of the elite or to pay for roads that reach remote parts of the country.

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

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

where $M(\mu, d_{ij})$ is the cost of homogenizing by μ for the individual at distance d_{ij} from government j. As before, for a government j and any individual in that country, i, the difference between the policies of government j and i's ideal policies is reduced by $\mu a d_{ij}$. 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 $\lim_{\mu \to 1} M_\mu(\mu, d_{ij}) = \infty$. The difference between odious and non-odious homogenization is that the cost of homogenizing now falls most heavily on those who are homogenized the most, that is those furthest from the government. Note that an individual at distance d_{ij} from the government is "moved" towards

the government in the amount $\mu a d_{ij}$. We assume the cost of homogenization, $M(\mu, d_{ij})$, is linearly increasing in d_{ij}, the distance of the individual from the government. We also assume the marginal cost of homogenization, $M_\mu(\mu, d_{ij})$, is weakly increasing in distance from the government. That is, the marginal cost of homogenization cannot be smaller 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

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

where $\lambda = \mu$. 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. A democracy weakly prefers non-odious homogenization.*

The intuition is simple: "odious homogenization" costs less to the dictator. The cost burden of homogenization shifts towards the rest of the population, at an increasing rate the more distant are individuals from the dictator himself.

The next proposition shows that odious homogenization has the same qualitative properties as non-odious homogenization:[29]

Proposition 6 *(i). A safe dictator undertakes zero homogenization*

(ii). The amount of odious homogenization undertaken by an unsafe dictator is increasing in the probability of democracy, p.

(iii). When $\lambda^* > 0$ an unsafe dictator who faces a high probability of overthrow will undertake more homogenization of the population than would be undertaken if instead the population formed a democracy.

As in corollary 1, when $\lambda^* > 0$ a population that forms a democracy would choose to split into two, while a population controlled by an unsafe dictator may be homogenized to such an extent that, should democracy prevail in the future, the population would then be homogenous enough that they would choose to form a single country.

When the dictator has the option of odious homogenization and supposing a democracy that weakly prefers non-odious methods chooses to homogenize by non-odious methods in the second period then:

[29] This proposition is independent of whether a democracy uses odious or non-odious technology.

Corollary 2 *When $\lambda^* > 0$ an unsafe ruler who faces a high probability of overthrow will choose to use odious homogenization and will undertake more homogenization of the population relative to that which would be undertaken if instead the population formed a democracy. Further, under a greater set of parameters than when non-odious homogenization is the only feasible technology, the unsafe ruler will homogenize to such an extent that should democracy prevail in the future the country will not break up.*

This corollary indicates that allowing for more flexibility in homogenization technologies, such that the burden of homogenization can be shifted towards minorities who are forced to homogenize the most, increases the amount of homogenization undertaken by the dictator and strengthens previous results.

This section shows that when dictators are free to use odious technologies we may observe even more extreme homogenization by dictators relative to the homogenization undertaken by democracies, both in the amount of homogenization undertaken and in the methods used.

7 Endogenous revolutions

7.1 Revolutions and Welfare

Suppose that the probability of overthrow depends negatively on population welfare, as measured by the median welfare in the population. Population welfare goes up with homogenization at low levels of homogenization and so the probability of overthrow goes down; whereas at high levels of homogenization welfare goes down and so the probability of overthrow goes up. Given that the cost of government is fixed and the ruler extracts no rents, he can impact welfare, and with it the probability of overthrow, only through homogenization.[30] Thus the probability of overthrow can be written as a differentiable function of homogenization, $p(\lambda)$, which is U-shaped and strictly decreasing for $\lambda < \lambda_1^m$ and strictly increasing for $\lambda > \lambda_1^m$. Welfare peaks at λ_1^m because when the dictator is in control there is a single government located at the center of the population and, as we showed above, λ_1^m is the level of homogenization which maximizes the utility of the median voter in a unified democracy.

Proposition 7 compares the case where the probability of overthrow depends negatively on the welfare of the population with the case where the probability of overthrow is exogenous, for the same probability of democracy.

Proposition 7 *When the probability of overthrow depends negatively on the welfare of the population weakly more homogenization is undertaken.*

Homogenization now both reduces the probability of overthrow (up to a point) 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

[30] Rent extraction would allow another margin for the dictator to affect welfare. The analysis of this extension is left for future research.

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.

7.2 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. This case can also be analyzed with our model.

Suppose now that the dictator, when choosing the optimal amount of homogenization λ, 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, $p(\lambda)$, which is increasing in λ. Proposition 8 compares the amount of homogenization undertaken in the case where the probability of overthrow increases in the amount of homogenization undertaken with the case where the probability of overthrow is exogenous, for the same probability of democracy.

Proposition 8 *When the probability of overthrow depends positively on the amount of homogenization undertaken weakly less homogenization is undertaken.*

This effect reduces, *ceteris paribus*, the optimal degree of homogenization. This adds an extra cost to homogenizing and so the ruler reduces the amount of homogenization he undertakes.

8 Discussion

In this section we illustrate some examples which highlight our results.

8.1 Safe Dictators

Our first result is that rulers not subjected to threats of secession do not homogenize their populations. As an example we offer the case of France under the Ancien Régime. Official figures produced in 1863 document that a quarter of France's communes spoke no French at all.[31] Even this is likely a gross exaggeration of the amount of French spoken: as an example, the same official figures document all communes in Pyrenees-Orientales as French speaking while also documenting that half of all school children in Pyrenees-Orientales spoke no French whatsoever. Based on a number of sources, 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 in the 18th Century and before had done little to impose a national language. Although the Ancien Régime aimed to centralize administration and imposed French at the highest administrative

[31] The information on France in this section is taken predominantly from Eugen Weber "Peasants into Frenchman: the modernization of rural France 1970 □ 1914".

level (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), there was little if any effort to foster a nation of French-speakers. Weber claims that 'the French Crown [showed] little concern with the linguistic conquest of the regions under its administration. Language was relevant merely as an instrument of control.' Only after the Revolution did this policy change. Further Weber claims that the French 'had no uniform concept of patriotism at the Revolution...and that patriotic feelings on the national level, far from instinctive, had to be learned.' According to our model, the relatively safe Kings of the Ancien Régime (at least up to the few years 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. Under the Ancien Régime, roads are described by Weber as a means to collect taxes and transport troops. Only after the Revolution, in the first half of the 19th Century, the existing network was improved and the first railway lines built.

8.2 Colonizers

As discussed above, colonizers who face a probability of overthrow do not behave in the same way as an internal dictator. This is because after they "leave" power a colonizer, literally speaking, "leaves" the region. Thus they care less than a "domestic" dictator about what happens if democracy prevails and do not undertake nation-building.

Indeed colonizers of Africa did not make any effort to build nation states which made sense (see Easterly and Levine (1997), Alesina, Easterly and Matuszeski (2010), and Michaolopoulos and Papaioannou (2012) amongst others). They mixed ethnicities in ways that created unstable and often failed states when they left. They, as dictators, had no interest in homogenizing and building a national identity since they knew that they were there just to extract rents and, should insurrections prevail, they knew that they would leave and so had no interest in the future of their former colonies. The lack of a national identity and ethno linguistic fragmentation are at the roots of this continent's major problems.

A further example is British rule in India. At the end of colonization by Britain, a multitude of languages were spoken in India. 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: "the creation of essentially linguistic provinces in post-independence India, and the resistance to imposition of one vernacular (Hindi)" were a result of vested interests in the different languages spoken in India on Independence. While Hindi, the most widely spoken language in India, was Ghandi's choice for a national language (Ghandi was a native speaker of Gujarati), India was a democracy and those advocating Hindi as a national language were unable to impose it on the population as a whole. Hobsbawn claims that instead 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. Contrast this with the case of France which moved slowly from control by French elites to modern democracy over the course of the nineteenth Century and which we discuss below.

8.3 Homogenizing rulers: domestic threats

Our second result concerns rulers facing a threat of democracy. The greater the threat that democracy will prevail the more homogenization rulers will undertake. This applies only to rulers or elites who expect to remain in the country after democratization.

This result relates to a much documented period of homogenization in European countries in the 19th Century. Hobsbawn states that 'in the last third of the nineteenth Century it became increasingly manifest that the democratization or at least the unlimited electoralization of politics, were unavoidable.' Indeed this is exactly the period in which historians document evidence of increasing intervention by states in national language, national schooling and other aspects of national identity.

Let us take again the concrete example of language in France. After the French Revolution, and increasingly throughout the 19th Century in France, 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, we should see more emphasis by French elites to impose the French language both for the sake of maintaining a unified country and in order to ensure that their language would be the dominant language of a democratic France. As we saw above, French was far from a national language, and indeed as France moved from the Ancien Régime, through the different Republics and towards modern democracy, the Republican elites instituted increasingly intense interventions to promote French as a national language. Weber notes that 'the French crown [showed] little concern with the linguistic conquest of the regions under its administration.' However, immediately after the Revolution 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." Much of the emphasis on training the nation in French was via schooling and the nineteenth Century saw more and more schooling measures implemented by various governments.

Eugen Weber suggests that such policies and the degree to which they were implemented increased in intensity with the onset of the Third Republic in 1870. As we saw above, in 1863 French was far from a national language: Weber cites documents that show $8,381$ of France's 37,510 communes spoke no French in 1863, i.e. at least a quarter of the country's population, and estimates that French was a foreign language for nearly half of all children at the time. The Jules Ferry laws on free, mandatory and secular public education, voted in 1881 and 1882, were a major reform. Jules Ferry, a lawyer holding the office of Minister of Public Instruction in the 1880s, is widely credited with creating the modern Republican School, l'Ecole Republicaine. The educational system, under the control of religious officials was replaced by a system characterized by state schools and secular teachers, with a curriculum including French language and literature, French geography, and the history of France. Efforts to create a national language during this time not only involved educating the population in

French but also the suppression of other languages: as late as 1890 a ministerial decree banned religious instruction in Flemmish and in 1902 the government banned (and actively policed the ban on) Breton language sermons.

Weber traces the modernization of French villages and argues that rural France went from backward and isolated to modern and possessing a sense of French nationhood during the late 19th and early 20th centuries. Emphasizing the role of republican schools, and universal military conscription he argues that schools and the army provided a common experience for Frenchmen. Increasingly exposed to French language – forced to listen to French spoken by army officers, taught French in the schools – peasants learned the official language.

Italy, once unified, also constituted a diverse population speaking a range of languages as well as dialects. Italian unification was completed in the 1860s at a time when the threat of increasing democracy was strong. Right from unification the Italian elite was concerned with the homogenization of language to build unity within the new country. 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.' 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.'

Our results also find that dictators threatened by democratic insurgency may try to homogenize their population by odious means. The Soviet Union for instance was constantly threatened by the diversity of the people it was trying to control. Indoctrination by means of public education was an important way of forcing homogenization under a Marxist Leninist doctrine. Lott (1999) notes that while the public services, like health, were lacking spending on education was especially high. In fact, this paper argues that health spending increased after democratization in former communist countries while public funding for education went down. More generally this paper shows that totalitarian regimes spend more on all those public goods and services which serve the purpose of indoctrination. 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 the breakdown of the country (Soviet Union, Yugoslavia, Czechoslovakia). In Spain, Franco's dictatorship used various forms of prohibition of local diversity, like prohibiting the use of languages other than Castillian. After the fall of the dictatorship the country did not split, but decentralization policies flourished.

8.4 Homogenizing rulers: foreign threats

Dictators may feel threatened not only from domestic insurgents but also from foreign enemies. In this case homogenization of the population may be a precondition for increasing the unity of the nation and the willingness to fight a common enemy. War, or the threat of war, may also increase the probability of democracy within a country and thus increase homogenization through this channel also. This phenomenon is documented by Aghion, Persson and Rouzet

(2012). They suggest three striking examples which are simply an illustration of a much more general pattern. They establish an interesting and strong correlation between a 'war risk' indicator and investment in public education.[32]

8.4.1 Prussia

Prussia, threatened by Napoleon, understood the importance of promoting patriotism amongst the population. Under Stein, Prussia went through an impressive series of state reforms; nevertheless, it was soon realized that these reforms were insufficient for the development of patriotism and it was decided that the low level of education of the population was to blame. Stein assigned Wilhelm von Humboldt the task of reforming the entire education system. The purpose of the system was to instil loyalty to the Crown and to train young men for the military and the bureaucracy. Humboldt ensured that schools imposed an official language, to the prejudice of ethnic groups living in Prussia. German philosopher Johann Gottlieb Fichte, an influential figure in the system, said, "If you want to influence [the student] at all, you must do more than merely talk to him; you must fashion him, and fashion him in such a way that he simply cannot will otherwise than what you wish him to will". The reformers were convinced that the population had to be educated in order to grow as a nation with a uniform goal. The reform of the Prussian education system, *Bildung*, was a key step in the aim of achieving this goal.

8.4.2 France

The defeat in 1870 against Prussia helped the state to prevail against the opposition of the Church in the battle for public education. As seen above, education reforms not only increased literacy and offered more access to schooling, but also aimed at conveying patriotic values to new generations. We discussed more extensively this case in the previous section.

8.4.3 Japan

From the 17th century, Japan was ruled by military lords of the Tokugawa dynasty. Education was a privilege of the Samurais and centered on the study of Confucian classics. However, from mid 1850s Japan came under threat from Western countries. This situation acted as a catalyst for a series of changes within the nation. In particular, the conflict led to the end of the Tokugawa dynasty, which was replaced by Meiji, who, in 1868, became Japan's 122nd emperor. The leaders of the Meiji Restoration, as this revolution came to be known,

[32] The underlying idea is that a highly disintegrated population needs to be transformed into a unified nation where people share the same patriotic values and moral principles, a spoken and written language, and a motivation to defend their country in future conflicts. In order to find empirical evidence of this hypothesis they study historical panel data on education spending and enrollment for 137 countries between 1830 and 2001. The dependent variable measures primary enrollment per capita. 'War risk' is measured in two alternative ways. The first one is a binary indicator set equal to one if a country was engaged in an interstate war in the previous 10 years. The second measure is 'military rivalry' which is a dummy variable for whether a country has a strategic rival in a given year according to Thompson (2001), which measures the risk for a country to be involved in a war based on contemporary perception of political decision-makers.

acted in the name of restoring imperial rule in order to strengthen Japan against the threat represented by the western colonial powers. The goal was to combine "western advances" with the traditional, "eastern values".

Among the major innovations was the abolition of feudalism and the reform of the education system. The education reform introduced 4 years of compulsory elementary education for all children, the establishment of the Ministry of Education and the Fundamental Code of Education. Together, these two changes in education provided the foundation for the gradual systematization of content and teaching method. A critical consequence of the Meiji reform was the development of nation-wide standards that weakened the previous decentralized prefectural system. Indeed, this was done with the goal of unifying the population, and therefore the nation, by instilling a notion of ethical behavior. Compulsory education came to be viewed as the most effective means of generating nationalism.

Despite the initial violent reactions to the introduction of this new system, the results of the reform were impressive. In just one year after the reform (1872) about 12,500 primary schools were established. Elementary education was made universal and four years of schooling were made compulsory, with each school year consisting of four months. Within the next five years the number of schools doubled to a figure not surpassed until the 1960s. As a result, elementary school attendance rose from only 28% of school age children in 1873 to over 50% in 1883, and surpassed 96% in 1906.

8.5 Homogenizing democracies

Our results in section 5 compare the amount of homogenization undertaken by different regimes. A country that moves from control by a non-democratic ruler who faces a high probability of democracy to the realization of democracy next period will see a fall in the amount of homogenization undertaken when democracy is established. This result seems counter-intuitive to the idea that, for example, democratic populations will demand more education. However, there is some evidence that the move to democracy can be accompanied by a reduction in homogenization policies. Lott (1999) finds government educational expenditures fell by 4% of GDP in the first few years after the fall of communism while government health care expenditures as a percentage of GDP increased by 70%. 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. It would of course be interesting to compare the content of education under a democratic regime or a dictatorship. Under the latter education would 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.

Nevertheless democracies homogenize as well. For instance in most democracies education is publicly provided rather than subsidized by vouchers for individuals to choose their own private education. One of the reasons why public education is not privatized may be the fear of a loosening of sense of national unity. Often what is taught in school is highly coordinated

by governments. Roads are built at large expenses even to reach remote parts of a country. A sense of patriotism is built even in a democracy, but most of the time with less emphasis and aggressiveness than in some dictatorships.

9 Conclusion

We examined when and to what extent a government would choose to homogenize a population. We study different regimes and compare the incentives to undertake homogenization and the homogenization undertaken by each regime. There are three 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 will undertake a positive amount of homogenization. A democracy provides public goods for the benefit of the population as a whole and will homogenize to some degree to improve general access to the public good. 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 that vested interest is challenged a dictator is willing and able to undertake homogenization in order to better preserve the status quo. We examine different types of homogenization technology: one which imposes an equal tax burden on the population as a whole and one which involves personal cost to minorities or those who are homogenized the most. A dictator will always impose the odious technology on the population since it transfers the cost of homogenizing from himself to those who are homogenized and this can increase the amount of homogenization he undertakes.

Appendix

Lemma 1

A democracy will locate the government at the center of the population.

Proof. First, examine how the level of homogenization chosen would differ with the location of the government on the $[0, 1]$ interval. Preferences over the amount of homogenization remain single peaked wherever the government is located. If the government is located anywhere in the interval $j \in [0.25, 0.75]$ then the median voter is individual i at distance $d_{ij} = 0.25$ from government j. Homogenization is λ_1^m. If the government is at $j < 0.25$ then the median individual is always the individual at 0.5 and the level of homogenization chosen will be

$$ga(0.5 \square j) = C'^{j}).$$

Symmetrically, if $j > 0.75$ then the median level of homogenization is

$$ga(j \square 0.5) = C'^{j}).$$

Locating the government at the center beats all other governments located in the interval $j \in [0.25, 0.75]$ in a pairwise vote. The amount of homogenization undertaken is the same when the government is located in this interval, in which case if the government is located at j on one half of the interval $[0, 1]$ then everyone in the other half strictly prefers the government located at the center. Next we have to show what happens if $j \in [0, 0.25)$. Label by d the position of the government where $d \in [0, 0.25)$ and let λ^d denote the amount of homogenization chosen in the next round if the government is located at d. Label by $l_i \in [0, 0.5]$ the distance of the individual on the left hand side interval who is indifferent between the government located at d and the government located at the center. Then

$$g \square ga(1 \square \lambda_1^m)l_i + y \square k \square C(\lambda_1^m) = g \square ga(1 \square \lambda^d)(0.5 \square d \square l_i) + y \square k \square C(\lambda^d).$$

Similarly denote by $l_i' \in [0.5, 1]$ the individual on the right hand side interval who is indifferent between the government located at d and the government located at the center. Then

$$g \square ga(1 \square \lambda_1^m)l_i' + y \square k \square C(\lambda_1^m) = g \square ga(1 \square \lambda^d)(0.5 \square d + l_i'^d).$$

The proportion of the population who prefer a government at the center is then $l_i + l_i'$. Rearrange to find

$$l_i + l_i' = \frac{1}{ga}\left(\frac{2(1 \square \lambda_1^m)}{(1 \square \lambda_1^m)^2 \square (1 \square \lambda^d)^2}\right)[C(\lambda^d) \square C(\lambda_1^m) + ga(1 \square \lambda^d)(0.5 \square d)]$$

Since $C(\cdot)$ is a convex continuously differentiable function on $\lambda \in (0, 1)$ the following holds:

$$C(\lambda^d) \square C(\lambda_1^m) \geq C'(\lambda_1^m)[\lambda^d \square \lambda_1^m] = ga0.25(\lambda^d \square \lambda_1^m)$$

and since we examine $d < 0.25$ we have $(0.5 \square d) > 0.25$, hence

$$l_i + l_i' > \frac{1}{ga}\left(\frac{2(1 \square \lambda_1^m)}{(1 \square \lambda_1^m)^2 \square (1 \square \lambda^d)^2}\right)[ga0.25(\lambda^d \square \lambda_1^m) + ga0.25(1 \square \lambda^d)] > 0.5.$$

In the same way it is easy to show that any democracy of size will locate the government at the center by showing $(l_i + l_i') > 0.5$.

Proof of Proposition 1

In choosing between remaining as one country or splitting into two, an individual is facing homogenization levels of λ_1^m and λ_2^m respectively. Each individual simply evaluates their utility in either situation. For individual i at distance $l_i \in [0, 0.25]$ from the center of the population the value of staying as one country rather than splitting into two is:

$$[g \square (1 \square \lambda_1^m)gal_i + y \square k \square C(\lambda_1^m)] \square [g \square (1 \square \lambda_2^m)ga(0.25 \square l_i) + y \square 2k \square C(\lambda_2^m)] \quad (8)$$

For $l_i \in [0.25, 0.5]$ this is:

$$[g \square (1 \square \lambda_1^m)gal_i + y \square k \square C(\lambda_1^m)] \square [g \square (1 \square \lambda_2^m)ga(l_i \square 0.25) + y \square 2k \square C(\lambda_2^m)] \quad (9)$$

Expressions (8) and (9) can be rewritten respectively as

$$\square((1 \square \lambda_2^m) + (1 \square \lambda_1^m))gal_i + (1 \square \lambda_2^m)ga0.25 + k \square [C(\lambda_1^m) \square C(\lambda_2^m)] \quad (10)$$

$$((1 \square \lambda_2^m) \square (1 \square \lambda_1^m))gal_i \square (1 \square \lambda_2^m)ga0.25 + k \square [C(\lambda_1^m) \square C(\lambda_2^m)]. \quad (11)$$

Expression (10) is at a maximum when $l_i = 0$ and decreasing until $l_i = 0.25$; while expression (11) is increasing from this point at $l_i = 0.25$ to a maximum at $l_i = 0.5$. Thus there exist 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 $l_i' + (0.5 \square l_i'') = 0.25$, who are the median voters. Thus l_i' solves:

$$\square((1 \square \lambda_2^m) + (1 \square \lambda_1^m))gal_i' + (1 \square \lambda_2^m)ga0.25 + k \square C(\lambda_1^m) + C(\lambda_2^m)$$

$$= ((1 \square \lambda_2^m) \square (1 \square \lambda_1^m))ga(0.25 + l_i') \square (1 \square \lambda_2^m)ga0.25 + k \square C(\lambda_1^m) + C(\lambda_2^m).$$

Hence

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

Since $\lambda_1^m > \lambda_2^m$, the median voters are at $0.125 < l_i' < 0.25$ and $0.375 < l_i'' < 0.5$. or $i \in (0, 0.125), (0.25, 0.375), (0.625, 0.75), (0.875, 1)$.

9.1 Proof of Proposition 3

First observe that in period 2, if democracy prevails, the population choose to homogenize to λ_1^m for one country or λ_2^m for two countries, if the amount of homogenization undertaken in the first period by the dictator is less.

Second, examine where the dictator would locate the country, government and how much homogenization would need to be undertaken if he wants to ensure the population would not choose to split in the second period. Let λ^* be the minimum value that induces the following expressions to equal zero. If $\lambda^* < \lambda_1^m$ then

$$[g \square (1 \square \lambda_1^m)gal_i + y \square k \square [C(\lambda_1^m) \square C(\lambda^*)]] \square [g \square (1 \square \lambda^*)ga(0.25 \square l_i) + y \square 2k] = 0 \quad (12)$$

where $l_i = 0.25((1 \square \lambda_1^m) + (1 \square \lambda^*))/2(1 \square \lambda^*)$. If $\lambda^* \geq \lambda_1^m$ then

$$[g \square (1 \square \lambda^*)gal_i + y \square k] \square [g \square (1 \square \lambda^*)ga(0.25 \square l_i) + y \square 2k] = 0 \quad (13)$$

where $l_i = 0.25$. There is no cheaper way for the dictator to ensure a democracy chooses not to split. He could split the population into two countries and homogenize only Country A, but this both defeats the point and requires at least homogenization λ^* of Country A which is just as costly to the dictator since it would cost $0.5C(\lambda^*)$ but now taxes are shared only between half the population.

Recall that for the ruler the optimal choice of homogenization in the interval $[\lambda^*, 1]$ is λ^* since this is the cheapest way of ensuring a democracy will form a single country. This is unaffected by p. The optimal choice of homogenization in the interval $[0, \lambda^*)$ at a given p is $\lambda_2^r \in \{0, \tilde{\lambda}\}$ where

$$\tilde{\lambda} = \underset{\lambda \in (\lambda_2^m, \lambda^*)}{\arg\max}\, p(g \square (1 \square \lambda) ga 0.25 + y \square 2k \square C(\lambda)) + (1 \square p)(g + y \square k \square C(\lambda)). \tag{14}$$

Note that since a democracy homogenizes to λ_2^m if democracy prevails, the dictator will never choose $\lambda_2^r \in (0, \lambda_2^m]$. A choice of $\lambda_2^r = 0$ gives utility

$$p(g \square (1 \square \lambda_2^m) ga 0.25 + y \square 2k \square C(\lambda_2^m)) + (1 \square p)(g + y \square k)$$

and a choice of $\lambda_2^r = \tilde{\lambda}$ gives utility

$$p(g \square (1 \square \tilde{\lambda}) ga 0.25 + y \square 2k \square C(\tilde{\lambda})) + (1 \square p)(g + y \square k \square C(\tilde{\lambda})).$$

The optimal choice depends on which is larger. Observe that λ_2^r is weakly increasing in p.

Now show that the value of homogenizing enough to ensure a democracy chooses to form a single country is also increasing as p increases. The dictator's utility from homogenizing to λ^* is

$$g + y \square k \square C(\lambda^*)$$

if $\lambda^* \geq \lambda_1^m$ (which does not change with p); and it is

$$p(g + y \square k \square C(\lambda_1^m)) + (1 \square p)(g + y \square k \square C(\lambda^*)).$$

if $\lambda^* < \lambda_1^m$. While his expected utility from homogenizing to λ_2^r is

$$p(g \square (1 \square \lambda_2^m) ga 0.25 + y \square 2k \square C(\lambda_2^m)) + (1 \square p)(g + y \square k).$$

when p is low and

$$[p(g \square (1 \square \lambda_2^r) ga 0.25 + y \square 2k \square C(\lambda_2^r)) + (1 \square p)(g + y \square k \square C(\lambda_2^r))].$$

when p is high. His expected utility from homogenizing to λ_2^r is decreasing in p faster than his expected utility from homogenizing to λ^*. This comes from noting that his expected utility from homogenizing to λ^* is decreasing in p at rate $\square[C(\lambda_1^m) \square C(\lambda^*)]$ (or zero), while his expected utility from homogenizing to λ_2^r is decreasing in p at rate

$$\square(1 \square \lambda_2^r) ga 0.25 \square k + \frac{\partial \lambda_2^r}{\partial p}(pga 0.25 \square C'(\lambda_2^r))$$

(where the last term in the above expression is zero since $\lambda_2^r = \tilde{\lambda}$) or decreasing at rate $\square(1 \square \lambda_2^m)ga0.25 \square k \square C(\lambda_2^m)$. But we know from expression (12) that $\square(1 \square \lambda_1^m)gal_i + (1 \square \lambda^*)ga(0.25 \square l_i) + k = C(\lambda_1^m) \square C(\lambda^*)$ and so, since $\lambda_2^r < \lambda^*$ and $\lambda_2^m < \lambda^*$ then $(1 \square \lambda_2^r)ga0.25 + k > C(\lambda_1^m) \square C(\lambda^*)$ and $(1 \square \lambda_2^m)ga0.25 + k + C(\lambda_2^m) > C(\lambda_1^m) \square C(\lambda^*)$. It follows that the valuation of homogenizing enough to ensure a democracy forms a single country is also increasing in p.

9.2 Proof of Proposition 4

The median voter in a democracy faces the following valuation of staying as a single country relative to splitting into two

$$[g \square (1 \square \lambda_1^m)gal_i + y \square k \square C(\lambda_1^m)] \square [g \square (1 \square \lambda_2^m)ga(0.25 \square l_i) + y \square 2k \square C(\lambda_2^m)] \quad (15)$$

where the median voter is at $l_i = 0.25\frac{(1\square\lambda_1^m)+(1\square\lambda_2^m)}{2(1\square\lambda_2^m)}$. When $p = 1$ the ruler faces the following valuation of homogenizing enough to ensure a democracy chooses a single country versus homogenizing less such that a democracy would choose to split into two

$$[g + y \square k \square C(\bar{\lambda})] \square [g \square (1 \square \lambda_1^m)ga0.25 + y \square 2k \square C(\lambda_1^m)] \quad (16)$$

where $\bar{\lambda} = max\{\lambda^*, \lambda_1^m\}$. When $p = 1$ the optimal degree of homogenization for a dictator if he homogenizes below λ^* optimizes $g \square (1 \square \lambda)ga0.25 + y \square 2k \square C(\lambda)$, which is λ_1^m, thus a dictator always undertakes at least as much homogenization as a democracy that would split. Finally note that expression (16) is strictly positive for $\lambda^* \leq \lambda_1^m$ and thus the set of parameters for which (16) is positive is strictly larger than the set of parameters for which expression (15) is positive. Finally observe that there is no p at which homogenization is equal.

Proof of Proposition 5

We need to determine the degree of odious homogenization. Then, given we know how much odious and non-odious homogenization would be undertaken, we find the voter with the median preference for odious versus non-odious homogenization. When the population remains a single country the optimal level of odious homogenization for person i at distance d_{ij} from government j is

$$\mu_{1,i} = \arg\max_{\mu\in[0,1]}(g \square (1 \square \mu)gad_{ij} + y \square k \square M(\mu, d_{ij})$$

otherwise written

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

If $M(\mu, d_{ij})$ is linearly increasing in d_{ij} we can write

$$M(\mu, d_{ij}) = \beta(\mu)d_{ij} + \alpha(\mu).$$

We assume the total cost of homogenizing to the same degree is identical under odious and non-odious homogenization, hence

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

Hence $\beta(\mu)0.25+\alpha(\mu) = C(\mu)$ for all μ and it follows that $\beta'(\mu)0.25+\alpha'(\mu) = C'(\mu)$. We know that $\alpha'(\mu) > 0$ and $\alpha(\mu)$ continuous in μ since $M(\mu, 0)$ is strictly increasing and continuous in μ. Then $\beta(\mu)$ must also be continuous in μ. Also $\beta'(\mu) \geq 0$ since $M_\mu(\mu, d_{ij}) = \beta'(\mu)d_{ij} + \alpha'(\mu)$ is weakly increasing in d_{ij}. Similarly $\lim_{\mu\to 0}\beta'(\mu) = 0$ and $\lim_{\mu\to 0}\alpha'(\mu) = 0$. We can then rewrite the first order condition as

$$gad_{ij} = \beta'(\mu)d_{ij} + \alpha'(\mu)$$

Thus for $d_{ij} = 0$, individual i chooses $\mu = 0$ and as d_{ij} increases from 0 the optimal μ is increasing. Further preferences are single peaked over μ. Since $\alpha'(\mu) > 0$ then $\beta'(\mu) < ga$ and it follows that μ is increasing in d_{ij}. Hence $d_{ij} = 0.25$ is the median voter and we label the median homogenization μ_1^m. Observe that for the median individual under non-odious homogenization

$$ga0.25 = C'(\lambda_1^m)$$

and under odious homogenization

$$ga0.25 = M_\mu(\mu_1^m, 0.25) = \beta'(\mu_1^m)0.25 + \alpha'(\mu_1^m)$$

Thus $C'(\lambda_1^m) = \beta'(\mu_1^m)0.25 + \alpha'(\mu_1^m)$ but then we saw above that $\lambda_1^m = \mu_1^m$. That $\lambda_2^m = \mu_2^m$ follows as above.

Each individual evaluates the difference between their utility in the case of non-odious homogenization and their utility in the case of odious homogenization:

$$[\square(1 \square \lambda_1^m)gad_{ij} \square C(\lambda_1^m)] \square [\square(1 \square \mu_1^m)gad_{ij} \square M(\mu_1^m, d_{ij})]. \tag{17}$$

But when $M(\mu, d_{ij})$ is linear in d_{ij} we saw $\lambda_1^m = \mu_1^m$ and so expression (17) becomes

$$\square C(\lambda_1^m) + M(\mu_1^m, d_{ij}).$$

Since $C(\lambda_1^m) = M(\mu_1^m, 0.25)$ (from above) we see that individuals i at distance $d_{ij} > 0.25$ prefer non-odious homogenization and those at distance $d_{ij} < 0.25$ prefer odious homogenization. Thus $i = 0.25$ is the median voter when deciding between odious and non-odious homogenization and he is indifferent between the two. Similarly when the population chooses to split into two countries, the median voter is at 0.125 and is indifferent between the two technologies with those closer to the new government preferring odious homogenization and those further away preferring non-odious.

Proof of Proposition 6

For $l_i \in [0, 0.25]$, where l_i measures the distance of individual i from the center, the value of staying as one country rather than splitting into two is:

$$\square((1 \square \mu_2^m) + (1 \square \mu_1^m))gal_i + (1 \square \mu_2^m)ga0.25 + k \square [M(\mu_1^m, l_i) \square M(\mu_2^m, 0.25 \square l_i)] \quad (18)$$

For $l_i \in [0.25, 0.5]$ this is:

$$((1 \square \mu_2^m) \square (1 \square \mu_1^m))gal_i \square (1 \square \mu_2^m)ga0.25 + k \square [M(\mu_1^m, l_i) \square M(\mu_2^m, l_i \square 0.25)]. \quad (19)$$

Expression (18) is at a maximum when $l_i = 0$ and linearly decreasing until $l_i = 0.25$. Expression (19) may be linearly decreasing or increasing in l_i from $l_i = 0.25$ to $l_i = 0.5$. If expression (19) is decreasing then the median individual is at $l_i = 0.25$. If expression (19) is increasing then there exist 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 $l_i' + (0.5 \square l_i'') = 0.25$, who are the median voters. Substituting this into expressions (18) and (19) we find:

$$l_i' = 0.25\frac{(1 \square \mu_1) + (1 \square \mu_2)}{2(1 \square \mu_2)} + \frac{1}{2ga(1 \square \mu_2)}[M(\mu_1^m, 0.25 + l_i') \square M(\mu_1^m, l_i')] \quad (20)$$

$$\square \frac{1}{2ga(1 \square \mu_2)}[M(\mu_2^m, l_i') \square M(\mu_2^m, 0.25 \square l_i')]. \quad (21)$$

This is a fixed point equation in l_i. The first two terms are positive. The third term is weakly positive if $l_i \leq 0.125$ and negative otherwise. Suppose $l_i \leq 0.125$. A contradiction since the first term is ≥ 0.125. Thus $l_i' > 0.125$.

Suppose a dictator homogenizes by μ^r, how much further homogenization will be undertaken by a democracy? Suppose a democracy would choose to form a single country then individual i's optimal level of homogenization is μ which maximizes

$$g \square (1 \square \mu^r \square \mu)gal_i + y \square k \square [M(\mu^r + \mu, l_i) \square M(\mu^r, l_i)].$$

The optimal μ satisfies

$$gal_i = M_\mu(\mu^r + \mu, l_i)$$

and hence $\mu^r + \mu = \mu_1^m$. Suppose a democracy chooses to form two countries then individual i's optimal level of homogenization is μ which maximizes

$$g \square (1 \square \mu^r \square \mu)ga(0.25 \square l_i) + y \square k \square [M(\mu^r + \mu, 0.25 \square l_i) \square M(\mu^r, 0.25 \square l_i)].$$

The final term in square brackets is written like this for the following reason: the first term says how much it costs to homogenize to $\mu^r + \mu$. But homogenization μ^r has already been done so that part of the cost is already paid even though it originally cost the individual at l_i amount $M(\mu^r, l_i)$. The optimal homogenization for the individual is μ such that $\mu^r + \mu = \mu_2^m$.

Let μ^* be the level of homogenization which, if undertaken by the dictator in period 1, ensures a democracy will form a single country. Examine how much homogenization is undertaken by a dictator. If the dictator undertakes homogenization $\mu \leq \mu_2^m$ his expected utility is thus

$$p[g \square (1 \square \mu_2^m)ga0.25 + y \square 2k \square M(\mu, 0) \square [M(\mu_2^m, 0.25) \square M(\mu, 0.25)]]$$
$$+(1 \square p)[g + y \square k \square M(\mu, 0)]$$

and the optimal level of homogenization satisfies

$$M_\mu(\mu, 0) = pM_\mu(\mu, 0.25). \tag{22}$$

There will be some weakly positive amount of homogenization when p is small since $M_\mu(\mu, 0) \leq M_\mu(\mu, 0.25)$. If the dictator undertakes homogenization μ such that $\mu^* > \mu > \mu_2^m$, then the optimal μ satisfies

$$pga0.25 = M_\mu(\mu, 0) \tag{23}$$

Given the parameters the dictator chooses the optimal $\mu < \mu^*$ accordingly.

Denote by μ^* the value of μ that satisfies

$$[g \square (1 \square \mu^*)ga0.25 + y \square k] \square [g \square (1 \square \mu^*)ga(0.25 \square l_i) + y \square 2k] = 0$$

where $l_i = 0.25$ or

$$[g \square (1 \square \mu_1^m)gal_i + y \square k \square [M(\mu_1^m, l_i) \square M(\mu^*, l_i)]] \square [g \square (1 \square \mu^*)ga(0.25 \square l_i) + y \square 2k] = 0 \tag{24}$$

if $\mu^* < \lambda_1^m$, where $l_i \in (0.125, 0.25]$ is the median voter similarly as above. A dictator will not homogenize above μ^* since it increases costs and does not improve the outcome. If $\mu^* \geq \mu_1^m$ his utility is

$$g + y \square k \square M(\mu^*, 0)$$

if $\mu^* < \mu_1^m$ it is

$$p[g + y \square k \square M(\mu_1^m, 0)] + (1 \square p)[g + y \square k \square M(\mu^*, 0)]$$

Examine the choice of the dictator between μ^* and μ_2. Note that the value of homogenizing enough to form a single country either does not change with p or is decreasing in p at rate $\square[M(\mu_1^m, 0) \square M(\mu^*, 0)]$ if $\mu^* < \mu_1^m$. The value of homogenizing less than μ^* is decreasing in p at rate $\square(1 \square \mu)ga0.25 \square k$ (or $\square(1 \square \mu_2^m)ga0.25 \square k \square [M(\mu_2^m, 0.25) \square M(\mu, 0.25)]$). To show that the value of forming a single country relative to splitting into two is increasing in p we need to show that $(1 \square \mu)ga0.25 + k > [M(\mu_1^m, 0) \square M(\mu^*, 0)]$ when $\mu^* < \mu_1^m$, which follows from (24) and the fact that the marginal cost of homogenizing cannot be decreasing in distance from the government.

Compare the homogenization undertaken by a dictator at $p = 1$ with a democracy. If a democracy would form a single country then the dictator chooses λ to optimize

$$p[g + y \square k \square M(\mu_1^m, 0)] + (1 \square p)[g + y \square k \square M(\mu, 0)].$$

Thus he optimally undertakes no homogenization. If a democracy would split, the democracy homogenizes to a maximum of μ_1^m where $ga0.25 = M_\mu(\mu_1^m, 0.25)$. When $p = 1$ a dictator chooses between μ^* and μ_2 where μ_2 satisfies $ga0.25 = M(\mu_2, 0)$. Since $M_\mu(\mu, d_{ij})$ is weakly increasing in d_{ij} it follows that $\mu_2 \geq \mu_1^m > \mu_2^m$. Thus a dictator will always want to undertake more homogenization than a democracy and it follows that the set of parameters for which a dictator will ensure a single country is strictly higher than for a democracy.

Since a democracy is indifferent between homogenization technologies we should also compare a dictator using odious homogenization and a democracy using non-odious. This follows much the same as above, noting that the maximum amount of homogenization undertaken by a democracy is $\lambda_1^m \leq \mu_2$ and that now a dictator may undertake weakly positive homogenization at low probabilities of democracy and if a democracy would not split.

9.3 Proof of Corollary 2

A democracy uses non-odious technology and so λ^*, as above, is the amount of homogenization that ensures a democracy would form a single country. A democracy homogenizes to λ_2^m if they split and λ_1^m if they form a single country. When $p = 1$ and using odious technology a dictator who homogenizes to $\mu < \lambda^*$ optimizes

$$p[g \square (1 \square \mu)ga0.25 + y \square k \square M(\mu, 0)] + (1 \square p)[g + y \square k \square M(\mu, 0)],$$

so μ satisfies $ga0.25 = M_\mu(\mu, 0)$ when $p = 1$. Denote this by μ_2 and note it is weakly greater than the optimal amount of homogenization less than λ^* when $p = 1$ using non-odious methods (λ_1^m) since the marginal cost of homogenization is weakly increasing in distance from the government. We next show that when $p = 1$ a dictator using odious methods will always homogenize to λ^* if a dictator using non-odious methods would do so. A dictator using odious methods homogenizes to λ^* when $p = 1$ when the following is positive

$$[g + y \square k \square M(\hat{\lambda}, 0)] \square [g \square (1 \square \mu_2)ga0.25 + y \square 2k \square M(\mu_2, 0)]. \tag{25}$$

A dictator using non-odious methods homogenizes to λ^* when $p = 1$ when the following is positive

$$[g + y \square k \square C(\hat{\lambda})] \square [g \square (1 \square \lambda_1^m)ga0.25 + y \square 2k \square C(\lambda_1^m)] \tag{26}$$

where $\hat{\lambda} = \max \lambda_1^m, \lambda^*$. Subtracting expression (26) from (25) we get

$$\square M(\hat{\lambda}, 0) + (1 \square \mu_2)ga0.25 + M(\mu_2, 0) + C(\hat{\lambda}) \square (1 \square \lambda_1^m)ga0.25 \square C(\lambda_1^m). \tag{27}$$

Since costs are convex then $(\mu_2 \square \lambda_1^m)ga0.25 \leq C(\mu_2) \square C(\lambda_1^m)$ and expression (27) is greater than

$$[C(\hat{\lambda}) \square M(\hat{\lambda}, 0)] \square [C(\mu_2) \square M(\mu_2, 0)]$$

which is weakly positive since the marginal cost of homogenization is weakly increasing in distance from the government.

Proof of Proposition 7

Let λ_{en} denote the utility maximizing $\lambda \in [0, 1]$ for the endogenous dictator. The first order conditions in the endogenous case are as follows. If $\lambda^* = 0$, then λ_{en} satisfies

$$p'(\lambda)[C(\lambda) \square C(\lambda_1^m)] = (1 \square p(\lambda))C'(\lambda). \tag{28}$$

If $\lambda^* > 0$ and $\lambda^* > \lambda_{en} > \lambda_2^m$ then λ_{en} satisfies

$$p'(\lambda)[\square(1 \square \lambda)ga0.25 \square k] + p(\lambda)ga0.25 = C'(\lambda) \tag{29}$$

which we denote by $\tilde{\lambda}$, and if $\lambda_{en} \leq \lambda_2^m$ then λ_{en} satisfies

$$p'(\lambda)[\square(1 \square \lambda_2^m)ga0.25 \square k \square [C(\lambda_2^m) \square C(\lambda)]] = (1 \square p(\lambda))C'(\lambda) \tag{30}$$

which we denote by $\tilde{\tilde{\lambda}}$. The first order conditions in the endogenous case all have an extra positive term on the left hand side such that all else equal the λ that satisfies each expression is strictly higher than when $p'(\lambda) = 0$.

The endogenous dictator compares his utility from λ^*, $\tilde{\lambda}$, and $\tilde{\tilde{\lambda}}$ taking into account in his decision how the probability of overthrow changes. Given the choice of λ_{en}, the probability of overthrow is fixed at $p(\lambda_{en})$ for the exogenous dictator who also compares his utility from λ^*, $\tilde{\lambda}$, and $\tilde{\tilde{\lambda}}$ where $p(\lambda) = p(\lambda_{en})$ and $p'(\lambda) = 0$ in (28), (29) and (30). Thus it remains to show that if $\lambda_{en} < \lambda^*$ then the exogenous dictator chooses $\lambda_{ex} < \lambda^*$ where the exogenous probability is fixed at the relevant value of $p(\lambda_{en})$, and if $\lambda_{en} \leq \lambda_2^m$ then the exogenous dictator chooses $\lambda_{ex} \leq \lambda_2^m$ where the exogenous probability is fixed at the relevant value of $p(\lambda_{en})$. This can be shown simply by comparing the payoffs of the exogenous dictator from choosing λ^*, $\tilde{\lambda}$, and $\tilde{\tilde{\lambda}}$ where $p'(\lambda) = 0$ and the exogenous probability is fixed at the relevant value of $p(\lambda_{en})$.

Proof of Proposition 8

The first order conditions are given by (28), (29) and (30). The first order conditions in the endogenous case all have an extra negative term on the left hand side such that all else equal the λ that satisfies each expression is weakly lower than when $p'(\lambda) = 0$. Similar to above, it remains to show that if $\lambda_{en} = \lambda^*$ then $\lambda_{ex} = \lambda^*$ and if $\lambda^* > \lambda_{en} > \lambda_2^m$ then $\lambda_{ex} > \lambda_2^m$. As above, this can be shown easily by comparing the payoffs of the exogenous dictator from choosing λ^*, $\tilde{\lambda}$, and $\tilde{\tilde{\lambda}}$ when $p'(\lambda) = 0$ where the exogenous probability is fixed at the relevant value of $p(\lambda_{en})$.

A1

Alesina and Spoloare (1997) analyze the formation of countries and, in doing so, have to make some assumptions on how countries are formed. They suppose that any partition of the population into countries must be such that no member of a country would prefer to be a member of another country and that countries are stable to small perturbations of their borders. Impose this assumption here, so that when countries are formed, before any homogenization has been undertaken, no member of a country would prefer to be a member of another country and countries are stable to small perturbations of their borders. First, note that if no individual wants to change country and countries are stable to small perturbations before homogenization then the same conditions are satisfied post homogenization. Second, if the population splits into two non-disjoint countries of size s_1 and s_2, the individual at the border of these two countries must be indifferent between joining either country. Individuals at the border are indifferent when $s_1 = s_2$ and $s_1 s_2 = \frac{2k}{ga}$. Stability (before homogenization) requires that countries are of equal size and the maximum number of stable countries is N where $N < \sqrt{ga/2k}$. The assumption that the population can split into a maximum of two countries is then essentially a restriction on the parameter space.

A2

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 $\square ga0.25 + k$. When homogenization is possible, the value of forming a single country versus splitting into two for the median voter is

$$[g \square (1 \square \lambda_1^m)gal_i + y \square k \square C(\lambda_1^m)] \square [g \square (1 \square \lambda_2^m)ga(0.25 \square l_i) + y \square 2k \square C(\lambda_2^m)] \quad (31)$$

where $l_i = 0.25((1 \square \lambda_1^m) + (1 \square \lambda_2^m))/2(1 \square \lambda_2^m)$. To see that (31) is strictly higher than when there is no option to homogenize, observe that the expression

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

$$= \square(1 \square \lambda_2^m)ga(2l_i \square 0.25) + k$$

is strictly higher than when no homogenization is allowed. Then differentiating expression (31) with respect to λ_1^m to get

$$ga0.25\frac{(1 \square \lambda)}{(1 \square \lambda_2^m)} + ga0.25 \square C'(\lambda) \quad (32)$$

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

A3

Compare the amount of homogenization undertaken by dictators located at different points $i \in [0, 1]$. To do so examine the choice of homogenization of a dictator located at distance

$l_i \in [0, 0.5]$ from the center. His optimal homogenization if a democracy would form a single country given no previous homogenization is zero if $l_i \leq 0.25$ and satisfies λ_1 such that $gal_i = C'(\lambda_1)$ if $l_i > 0.25$. Now suppose a democracy would split into two given no previous homogenization. Then the optimal level of homogenization for a dictator located at $l_i \in [0, 0.25]$ when $p = 1$ is $\hat{\lambda} = \max\{\lambda^*, \lambda_1^m\}$ (since if $\lambda^* < \lambda_1^m$ then if the dictator homogenizes to λ^* when democracy prevails with probability 1 the democracy will further homogenize to λ_1^m) if (33) is negative and λ_2 if (33) is positive, where $\lambda_2 = \lambda_2^m$ if $l_i \in [0.125, 0.25]$ and λ_2 satisfies $ga(0.25 \square l_i) = C'(\lambda_2)$ if $l_i \in [0, 0.125)$:

$$[g \square (1 \square \hat{\lambda})gal_i + y \square k \square C(\hat{\lambda})] \square [g \square (1 \square \lambda_2)ga(0.25 \square l_i) + y \square 2k \square C(\lambda_2)] \qquad (33)$$

Expression (33) is decreasing in l_i. The optimal level of homogenization for a dictator at $l_i \in (0.25, 0.5]$ is $\hat{\lambda} = \max\{\lambda^*, \lambda_1\}$ if the following is negative and λ_2 otherwise:

$$[g \square (1 \square \hat{\lambda})gal_i + y \square k \square C(\hat{\lambda})] \square [g \square (1 \square \lambda_2)ga(l_i \square 0.25) + y \square 2k \square C(\lambda_2)] \qquad (34)$$

where $\lambda_2 = \lambda_2^m$ if $l_i \in (0.25, 0.375]$ and satisfies $ga(l_i \square 0.25) = C'(\lambda_2)$ otherwise. When $l_i \in (0.25, 0.375]$ we see that the first term is decreasing in l_i at a slower rate (at $\square(1 \square \lambda^*)ga$) compared to the second term (decreasing at $\square(1 \square \lambda_2^m)ga$). When $l_i \geq 0.375$, the first term is decreasing at rate $\square(1 \square \lambda^* ga)$ while the second term is decreasing at $\square(1 \square \lambda_2)ga + \frac{\partial \lambda_2}{l_i} ga(l_i \square 0.25) \square \frac{\partial \lambda_2}{l_i} C'(\lambda_2)$ but since λ_2 is a function of l_i such that $ga(0.25 \square l_i) = C'(\lambda_2)$, these last two terms cancel (and similarly if $\hat{\lambda} = \lambda_1$). Thus we see that λ_2 and the value of expressions (33) and (34) are increasing as the location of the dictator increases and decreases from location $i = 0.25$.

Examine a dictator at $i = 0.25$. If a democracy forms a single country then the dictator's optimal homogenization is λ_1^m which will be undertaken anyway so the dictator undertakes zero homogenization. If a democracy splits into two then the dictator will homogenize to $\hat{\lambda} = \max\{\lambda^*, \lambda_1^m\}$ if the following expression is positive and zero otherwise (at $p = 1$):

$$[g \square (1 \square \hat{\lambda})ga0.25 + y \square k \square C(\hat{\lambda})] \square [g + y \square 2k \square C(\lambda_2^m)].$$

The second term in square brackets is strictly higher for a dictator than a democracy. If $\lambda^* = \lambda_1^m$ then the first term in square brackets is strictly lower for a dictator, but since λ_1^m is optimal for a dictator $\lambda^* > \lambda_1^m$ reduces this term further. Thus if a democracy would split a dictator will never choose to homogenize to keep them together.

10 References

Acemoglu, D. and J. A. Robinson (2008) "Persistence of power, elites and institutions." *American Economic Review*, 98(1), 267 □ 93.

Aghion, P., T. Persson and D. Rouzet (2012)"Education and 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. 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.

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.

Easterly W. and R. Levine (1997) "Africa's Growth Tragedy: Policies and Ethnic Division." *Quarterly Journal of Economics*, 112(4), 1203-50.

Bolton, P. and G. Roland (1997) "The Breakup of Nations: A Political Economy Analysis." *Quarterly Journal of Economics*, 112(4), 1057-90.

Duggan, C. (2007) *The Force of Destiny: A History of Italy since* 1796. Penguin Books, London, UK.

Glaser E. G Ponzetto and A. Shleifer (2007) "Why Democracy needs Education?" *Journal of Economic Growth*, 12(4), 77-99.

Hobsbawn, E. J. (1990) *Nations and nationalism since 1780: programme, myth, reality.* Cambridge University Press, UK.

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.

Michaopolous S. and E. Papaioannou (2012) "The Long Run Effects of the Scramble for Africa." NBER Working Paper

Tilly, C. and L. Tilly (1973) *The Rebellious Century* Harvard University Press, Cambridge MA.

Weber, E. (1979) *Peasants into Frenchmen.* Chatto and Windus, London, UK.