Globalization and Political Structure*

Gino Gancia, Giacomo A. M. Ponzetto and Jaume Ventura

October 2021

Abstract

This paper develops a theoretical framework to study the interaction between globalization and political structure. We show that political structure adapts in a non-monotonic way to declining transport costs. Borders hamper trade. At an earlier stage, the political response to expanding trade opportunities consists of removing borders by increasing country size. At a later stage, instead, it consists of removing the cost of borders by creating international unions. This leads to a reduction in country size. Moreover, diplomacy replaces conquest as a tool to ensure market access. These predictions are consistent with historical evidence on trade, territorial changes and membership of international unions.

Keywords: Globalization, political structure, size of countries, international unions.

*Gino Gancia: Queen Mary University of London. Giacomo Ponzetto: CREI, Universitat Pompeu Fabra, IPEG and Barcelona School of Economics. Jaume Ventura: CREI, Universitat Pompeu Fabra and Barcelona School of Economics. E-mail: g.gancia@qmul.ac.uk, gponzetto@crei.cat, jventura@crei.cat. We thank Janko Heineken, Marta Santamaria and Jagdish Tripathy for excellent research assistance. We are grateful for helpful comments by Giovanni Andreottola, Pablo Beramendi, Graziella Bertocchi, Alessandra Bonfiglioli, participants at numerous conferences and seminars, the editor, Giovanni Peri, and four anonymous referees. We acknowledge financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Advanced Grant 693512 – GEPPS and Starting Grant 714905 – CITIZINGLOBAL), the Spanish Ministry of Economy and Competitiveness (RYC-2013-13838) and its Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563 and CEX2019-000915-S), the Generalitat de Catalunya through CERCA and the SGR Programme (2014-SGR-830 and 2017-SGR-1393), the Institut d'Estudis Autonòmics, and the Leverhulme Trust through a 2019 Research Fellowship.

1 Introduction

Since the Industrial Revolution, the cost of distance has been falling dramatically thanks to a stream of major technological innovations, like the railroad, the steamship, the telegraph, the jet engine, containerization and most recently the internet. Such technological progress has fundamentally transformed worldwide economic geography and fuelled a continuous expansion in the size of markets. In 1820, international trade was very modest: about 2% of world output. Over the following century, international trade grew more than four-fold to 8%. This first wave of globalization was cut short in the interwar period, which saw international trade decline to about 5% of world output. After World War II, a second wave of globalization started and still continues today. By 2010, international trade had reached unprecedented levels, surpassing 20% of world output.

These two ages of globalization saw political geography evolve in opposite directions. In the nineteenth century, economic and political integration proceeded together. Sovereign states grew larger and fewer, from 125 in 1820 to merely 54 at the eve of the Great War. Conversely, in the postwar era economic integration has been accompanied by political fragmentation. The number of countries has risen to a record high of more than 190. At the same time, there has been a proliferation of international treaties and institutions aimed especially at fostering economic integration, such as the World Trade Organization and the European Union. These trends are illustrated graphically in Figure 1, which shows the historical evolution of the number of sovereign states in the world and the number of members of the GATT/WTO, along with average exports as a share of GDP.[1]

The sharp reversal in the link between economic and political integration presents an open puzzle, which we address in this paper. We show that the observed evolution of political structure during both waves of globalization can be understood as an efficient response to the falling cost of distance. Our starting point is that borders hamper trade, so expanding trade opportunities make borders costlier. Political structure needs to adapt by removing borders or by reducing their cost. We find that this adaptation entails a non-monotonic evolution of efficient country size. At an earlier stage, the efficient political response to expanding trade opportunities is to remove borders by creating larger countries. At a later stage, it is to remove the cost of borders by creating international unions. This induces a reduction in the size of countries.

[1]The data on the number of states are from Butcher and Griffiths (2013). The trade share is from Maddison (2001). See Appendix A.1 for more details.

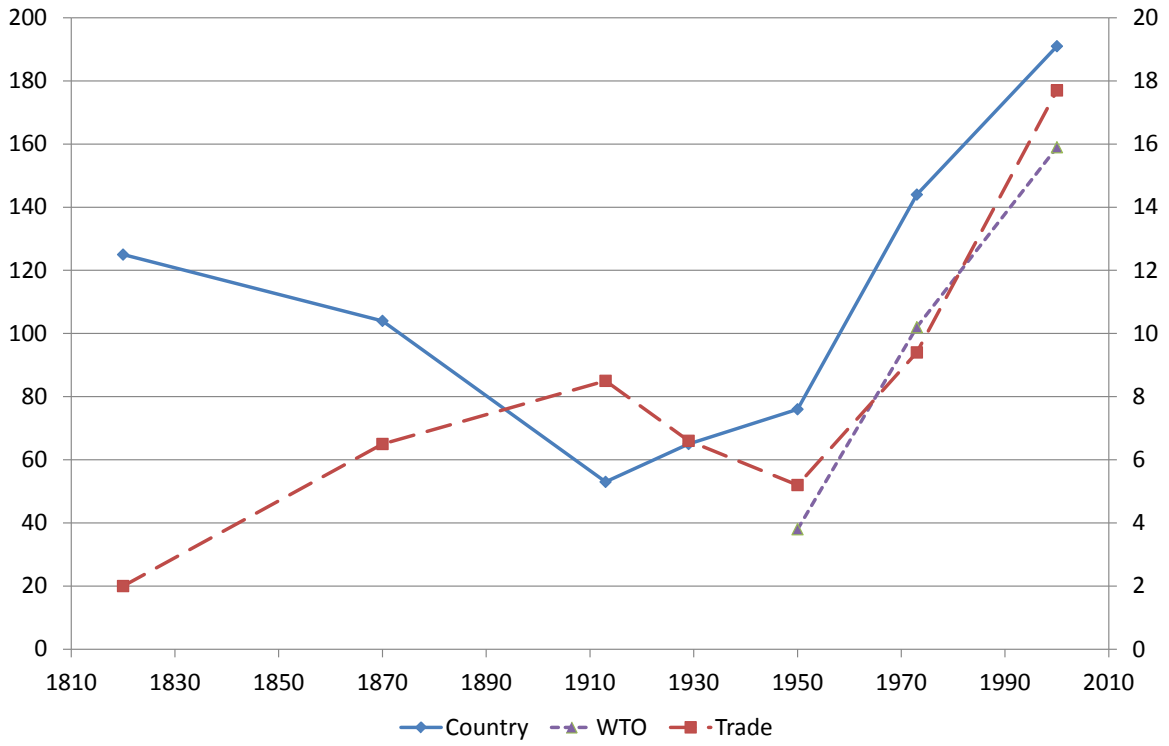

Figure 1: Economic Integration and Political Integration. The figure plots the trade share (right axis), and the number of countries and of WTO members (left axis). See Appendix A.1 for details on data.

Naturally, our theory does not imply that political geography responds to economic efficiency alone. On the contrary, our model also sheds light on historical patterns of interstate conflict. When creating international unions is too costly, we show it is appealing for the Great Powers to wage war to conquer markets. Empire-building enables them to trade with their colonies while imposing extractive institutions. As efficient market size keeps growing, however, we find that this strategy reaches a breaking point. Conflict and imperialism are no longer advantageous for the Great Powers. They prefer gaining access to world markets through peaceful diplomacy, by building supranational institutions that respect all members' political autonomy. These theoretical predictions are borne out by empirical evidence. During the first wave of globalization, increasing trade led to an expansion in country size accompanied by conflict. Both links were broken during the second wave of globalization, which saw instead the peaceful rise of supranational institutions.

To derive our results, in Section 2 we set up our model of a world with a continuum of basic geographical units, which we refer to as localities. Each contains people that share common preferences. Goods can be transported at a negligible cost within localities, but at

a positive cost across localities. Governments perform two tasks: (i) they enforce contracts, protect property rights and enact economic regulations that help markets work; and (ii) they provide public services such as education and welfare programs. We study two alternative ways of organizing the government. The first is a single-level government that performs both tasks. We refer to this single-level government as a country. The second possibility is a two-level government. The lower level provides public services and we also refer to it as a country; while the higher level regulates markets and we refer to it as an international union. Our goal is to study how the shape of government changes as the cost of distance declines.

We make two standard assumptions about the effects of government. The first is the presence of border effects. Localities that do not belong to the same country nor to the same international union have different economic regulations. Therefore, they can trade only a limited range of goods. The second assumption is policy uniformity. If two localities belong to the same country, they receive the same basket of public services even though they would prefer different baskets.[2]

Without costs of government, the optimal political structure would be a two-level government. The first level would be a continuum of country governments, one for each locality, providing each of them with its preferred basket of public services. The second level would be an international union of all localities that regulates markets and eliminates all border effects. Unfortunately, this political structure is too expensive.

We make two standard assumptions about the costs of government. The first is the presence of economies of scale. There are some costs of setting up and running a country that are fixed or independent of the number of localities that belong to it. The second assumption is the presence of economies of scope. Coordinating different levels of government is costly and, as a result, a two-level government is more expensive than a single-level government.

Economies of scale and scope affect the optimal political structure. Economies of scale make it desirable to have a discrete number of country governments rather than a continuum of them. Localities are willing to accept public services less than ideally tailored to their preferences in order to benefit from economies of scale. Economies of scope, if large enough, make it desirable to have a single- rather than a two-level government. Localities may also

[2]We assume that local preferences differ with respect to public services but not economic regulation. Market-enabling economic regulation aims primarily at increasing efficiency, and this goal is widely shared by people with different preferences (Coase 1960; Posner [1973] 2014). Accordingly, government functions such as contract enforcement, monetary policy, or the policing of anti-competitive practices are often entrusted to non-partisan experts. Public services, however, are a focus of political tension because people have different views on how children should be educated, on the proper size and scope of the welfare system, and so on. Such preferences vary systematically across localities reflecting their distinctive history and culture.

be willing to accept higher trade costs to take advantage of economies of scope.

The equilibrium political structure balances four classic forces: border effects, preference heterogeneity, economies of scale and economies of scope. We study how a reduction in the cost of distance affects this balance. Globalization, i.e., economic integration, is an endogenous outcome in our theory. An exogenous decline in the cost of distance fosters trade directly for a given political structure. It also fosters trade indirectly by producing endogenous changes in political structure aimed at facilitating trade.

In Section 3 we assume that localities bargain efficiently in a world ruled by law and diplomacy. When trade opportunities are limited, the benefit of creating an international union does not justify sacrificing economies of scope. Thus, a single-level government is optimal. As trade opportunities expand, the incentives to remove borders grow. The number of countries declines and the mismatch between each locality's ideal and actual provision of public services grows. Eventually, this mismatch is large enough to justify the move to a two-level government. The world political structure shifts from a few large countries to many small countries within an international union. This two-level government is more expensive, but it is nonetheless desirable because it facilitates trade and improves preference-matching in the provision of public services.

In Section 4 we allow a set of core localities to wage war and build empires. War is costly but it allows core localities to remove borders with their colonies while imposing their preferences on them. Thus, we find that there is an intermediate stage of globalization in which empires are formed. Eventually this stage ends, empires collapse, and international unions are created to promote trade and peace. Core localities choose to avoid the cost of war and to replace conquered colonies with free partners in an international union. The cause of imperial collapse at a late stage of globalization is the same as the cause for the rise of empires at an early stage: namely, the desire to reap gains from trade in the most cost-efficient manner.

In Section 5 we further relax our assumptions of symmetry and allow core localities to have lower costs of trading with one another. We show that this gives rise to a richer and more gradual pattern of international integration. World-straddling empires are not replaced at once by a world union, but may rather give rise first to a regional union of core localities, which only later grows to encompass the periphery too.

In Section 6 we show three significant patterns in historical data since 1870. First, increases in trade predict territorial expansion before international unions. Second, this correlation is reversed after international unions are created. Third, membership of international

unions predicts territorial contraction. This evidence is consistent with the predictions of our theory.

Related Literature The motivating facts presented in Figure 1 were first noted by Kahler and Lake (2004) and Lake and O'Mahony (2004). They highlight the puzzling reversal in the link between economic and political integration. They stress that it poses an obstacle for efficiency-based explanations of political structure, which have so far failed to account for it and for the emergence of supranational institutions during the second but not the first wave of globalization.

Several economic theories can help explain why economic integration has been accompanied by political fragmentation since World War II (Bolton and Roland 1996, 1997; Alesina, Spolaore and Wacziarg 2000; Casella 2001; Casella and Feinstein 2002). All of them, however, predict a monotonic effect of globalization on political structure, and thus fail to explain why the first wave of globalization was accompanied by a decline in the number of countries. Nor can they explain the creation of international unions. Closest to our own work, Alesina, Spolaore and Wacziarg (2000) add border effects to Alesina and Spolaore's (1997) seminal theory of country formation based on the trade-off between preference heterogeneity and economies of scale.[3] They explain the increase in the number of countries during the second wave of globalization by interpreting globalization as an exogenous weakening of the border effect. As borders become less costly, efficient political structure reacts by creating more borders.[4]

Our model is the first to account for the non-monotonic relationship between globalization and political structure, and to explain the appearance of multi-level governance during the second wave of globalization. We obtain these results because of two innovations relative to prior work. First, we recognize that economies of scope are limited.[5] As a consequence, a broader set of political structures can be efficient. Hence, we can explain the shift from a single-level to a two-level government. Second, we consider a more primitive technological driver of globalization: expanding trade opportunities caused by a gradual decline of

[3] Desmet et al. (2011) calibrate this trade-off for European countries.

[4] Bolton and Roland (1996, 1997) focus on income distribution and find that heterogeneous countries may break up if their barriers to external trade decline. Casella (2001) and Casella and Feinstein (2002) study how preferences for public goods can endogenously become more heterogeneous as market size expands and enables greater specialization.

[5] Alesina and Spolaore (2003) offer an insightful discussion of economies of scope. Chapter 2 analyzes an arbitrary system of overlapping jurisdictions. Chapter 9 presents a system of overlapping jurisdictions constrained to form a pyramidal hierarchy.

transport costs, which makes borders costlier. This creates incentives to remove borders rather than to create them. In our theory, the weakening of the border effect occurs only endogenously as political structure adapts to new trade opportunities.[6] This is consistent with O'Rourke and Williamson's (1999) reading of historical evidence. Trade integration in the 1800s was driven overwhelmingly by declining transportation costs, while trade liberalization also became important after the 1950s when it was achieved through international institutions such as the GATT/WTO.

Our work is also related to the literature on trade and war. The idea that trade promotes peace was formalized by Polachek (1980). It is based on the premise that conflict harms trade and hence trade openness raises the opportunity cost of war (Alesina and Spolaore 2003; Rohner, Thoenig and Zilibotti 2013). The opposite idea that trade generates military conflict is instead expressed in neo-Marxist theories of imperialism. Trade can also make countries dependent on others and therefore vulnerable (Bonfatti and O'Rourke 2018).[7] Our paper suggests that these seemingly antithetical views capture two different stages of the same model. We provide a unified explanation for why territorial changes are more associated with military conflict in the first wave of globalization than in the second, an empirical pattern noted by Lake and O'Mahony (2006). Consistent with our result that international unions remove the market-access incentive for waging war, Martin, Mayer and Thoenig (2012) find evidence that regional trade agreements promote peaceful relations.

The effect of war on country formation has been studied by Alesina and Spolaore (2003), Griffiths (2014) and Gennaioli and Voth (2015), among others. These papers show that changes in military technology can explain country size, investment in state capacity and the provision of public goods. However, they find a monotonic effect of military technology on political structure. Our theory also recognizes conflict as one of the determinants of country formation and of the provision of public goods. In our model, waging war is one reason why countries grow large. We show, however, that changes in military technology alone are not sufficient to explain persuasively the observed evolution of political structure. Instead, we find that expanding trade opportunities are key to explain endogenously the

[6] One way to interpret our findings is that, through our two innovations, we provide an explanation of the decline in the border effect that drives Alesina, Spolaore and Wacziarg's (2000) results. The reaction to this decline is the same in our model as in theirs: an increase in the number of countries. However, they take the decline as exogenously given. Instead, we show it is an endogenous consequence of the creation of international unions.

[7] Martin, Mayer and Thoenig (2008) show that bilateral trade lowers the probability of conflict, but multilateral trade openness decreases dependence on any given country and hence the cost of a bilateral conflict.

creation of international unions along with the switch from a world of aggression in which countries grew large to one of diplomacy in which countries became smaller.

Finally, our work is related more broadly to the economic analysis of federalism and of the geographic structure of government. Our model embeds the key trade-off that lies at the heart of the classic theory of fiscal federalism (Oates 1972). Centralization reaps economies of scale and benefits from policy coordination, but it imposes a uniform policy on localities with different preferences. Models of political centralization and decentralization have been applied most often to the architecture of government at the sub-national level (Lockwood 2006; Treisman 2007). However, the same insights apply to the study of international unions (Hooghe and Marks 2003; Alesina, Angeloni and Etro 2005; Ruta 2005). Prior research in this field has overwhelmingly focused on the optimal size and composition of an exogenously given number of government tiers. Surprisingly, the literature has devoted much less attention to the choice between a single-level and a multi-level governance structure, which our analysis focuses on (an exception is Boffa, Piolatto and Ponzetto 2016).[8]

2 A Model of Political Structure

In this section we develop a stylized model of the world that contains the basic ingredients of our theory: geography, markets and preferences. The model mixes these ingredients imposing a high degree of symmetry. This allows us to derive our basic results about the effects of expanding trade opportunities on political structure quickly and intuitively.

The concept of locality is a key primitive in our theory. We model the world as a set of places within which there are neither geographical nor cultural distances, and we label them localities. Thus, localities consist of a group of people sharing common preferences and inhabiting a particular territory. This approach, which is common in the literature, simplifies the study of how people with different preferences interact and organize themselves into political entities. But it is silent about how these different preferences arose in the first place and how they evolve over time. It also abstracts from conflict within localities.

The concept of trade opportunities is another important primitive in our theory. Geographic distance introduces trade costs across localities. In particular, we adopt the usual

[8] Federalism is a way of making public-good provision more efficient by decentralizing it to a sub-national level of government (Oates 1972). We view economic unions instead as a way of making market regulation more efficient by centralizing it to a supra-national level. Intuitively, the two structures may well coexist because they solve different problems. We derive these results formally in Appendix A.4.4, where we present an extension of our model that includes a third, sub-national, level of government.

assumption of iceberg transport costs. We examine the implications of expanding trade opportunities created by an exogenous decline in these costs.

2.1 Basic Setup

We consider a world with a continuum of atomistic localities, $l \in [0, 1]$. Each locality contains a positive measure of identical individuals. W_l denotes the welfare of the representative individual of locality l. For short, we refer to this individual as "locality l." Then, the welfare of locality l is:

$$W_l = W_l^M + W_l^G, \tag{1}$$

where W_l^M is the utility derived from the consumption of market goods and W_l^G is the utility derived from public services.

Governments provide public services and regulate markets, so government activity affects both welfare components. A political structure for this world consists of two partitions of the set of localities $[0, 1]$ into governments: a public-service partition P with typical element $P_n \in P$; and an economic-regulation partition R with typical element $R_n \in R$.

If $P = R$, we say that the world has a single-level governance structure, and we refer to the common elements of P and R as country governments or countries. Each of these countries provides both public services and market regulation to its constituent localities.

If $P \neq R$, we say that the world has a two-level governance structure. It will presently become clear that governments have a pyramidal hierarchy: if the partitions P and R do not coincide, the finer partition P is always a refinement of the coarser one R. Hence, we refer to the (smaller) elements of P as country governments or countries, and the (larger) elements of R as international unions or unions. Countries provide public services to their constituent localities, while unions regulate the markets of their constituent countries.

We develop a model of the partitions P and R, that is, a model of how localities organize themselves into countries and how countries organize themselves into unions. We start from assumptions about preferences, technology and the costs of government, and we determine how welfare W_l depends on political structure (P, R).

2.1.1 Markets and Trade

There is a continuum of industries, $i \in [0,1]$. Let $c_l(i)$ be the consumption of goods of industry i by locality l. The utility function takes the following form:

$$W_l^M = \int_0^1 \ln c_l(i)\, di. \tag{2}$$

The production of final consumption goods requires industry-specific differentiated input varieties, $m \in [0,1]$. Define $c_l(m,i)$ as the amount of inputs of the variety m for industry i used by locality l in the production of final goods. Then:

$$c_l(i) = \exp\left\{\int_0^1 \ln c_l(m,i)\, dm\right\}. \tag{3}$$

These consumption preferences and production technology are symmetric across and within industries, and we specify convenient Cobb-Douglas functional forms. As a result, each locality spends an identical fraction of its income on each variety of each industry.

To introduce gains from specialization and trade, we adopt a simple symmetric version of the Ricardian model. Each locality is endowed with one unit of labor in each industry. This unit can produce one unit of the variety with the same index as the locality ($m = l$), or $e^{-\eta}$ units of any other variety ($m \neq l$). Since $\eta > 0$, each locality has a technological advantage in its "own" variety. The parameter η measures the extent to which technologies differ across localities and, therefore, the potential gains from specialization and trade.

There are technological barriers to trade. We assume uniform iceberg transport costs across localities so that only a fraction $e^{-\tau} < 1$ of the goods shipped from l to $m \neq l$ arrives to destination. To focus on the most interesting case in which trade costs are not prohibitive and to ensure positive gains from trade, we assume that $\eta > \tau > 0$. Our measure of trade opportunities is the wedge $\gamma = \eta - \tau > 0$, which captures the potential gains from trade. This wedge increases as improvements in transportation technology reduce physical trade costs τ. Trade opportunities can thus range from $\gamma = 0$ when trade costs are prohibitive ($\tau = \eta$) to a maximum of $\gamma = \eta$ when trade costs are nil ($\tau = 0$).

Policy-induced barriers to trade (i.e., border effects) arise when different governments regulate markets. Specifically, we assume that exchanging goods in a fraction $\beta \in (0,1)$ of industries requires legal enforcement of contracts. In these industries, varieties cannot be traded between localities that have different governments regulating their markets, i.e., that belong to different elements R_n and $R_{n'}$. The reason is that foreigners correctly anticipate

that domestic courts will discriminate against them ex post. In the remaining set of industries, contracts are self-enforcing and thus varieties can be traded without restrictions. This formulation captures a simple yet realistic microfoundation for the well-known finding that borders obstruct trade.[9]

A market equilibrium is a set of prices and quantities such that individuals maximize utility and markets clear. Appendix A.3.1 shows that there exists a unique market equilibrium. Traded industries specialize in each locality's own input variety, export essentially all of their production and import all remaining input varieties. Thus, in the measure β of industries that are traded across borders, consumption of the full measure of varieties is $c_l(m,i) = e^{-\tau}$. In the remaining measure $1-\beta$ of industries, this is also the consumption in locality l of varieties from other localities it shares economic regulation with. However, locality l is forced to produce locally all input varieties that correspond to localities that do not belong to the same R_n. Thus, consumption of these nontraded varieties is $c_l(m,i) = e^{-\eta}$. Equilibrium utility from consuming market goods equals:

$$W_l^M = -\eta + \gamma\left(1 - \beta + \beta\int_0^1 I_{l=m}^R dm\right), \tag{4}$$

where $I_{l=m}^R$ is an indicator variable which takes value 1 if localities l and m belong to the same R_n, and zero otherwise. Equation (4) shows the impact of border effects. A decline in transport costs raises trade opportunities (γ) in every industry. However, border effects prevent a locality from reaping the gains from trade in a mass β of industries that require contract enforcement. As a consequence, the value of removing each border effect is proportional to $\beta\gamma$, where trade opportunities γ measure the potential gains from trade in any single industry and β the mass of industries subject to border effects.

2.1.2 Governments

Each locality has a fixed endowment of resources, normalized to one, that can be allocated to produce public services. These public services consist of a basket of differentiated varieties, $x \in [0,1]$. The basket provided to locality l is characterized by a density function $g_l(x)$ defined over these varieties, with $g_l(x) \geq 0$ and $\int_0^1 g_l(x)\,dx = 1$. The utility derived from

[9] This microfoundation is analyzed by Broner and Ventura (2011). There are other microfoundations, though. For instance, tariffs and non-tariff barriers are also policies that discriminate against foreigners and limit the range of goods that can be traded.

these public services is:

$$W_l^G = \int_0^1 \delta_l(x)\, u(g_l(x))\, dx - K, \tag{5}$$

where $\delta_l(x) \geq 0$, $u(g_l(x)) = -1/g_l(x)$ and K is a cost function to be defined shortly. We refer to the first and second terms of Equation (5) as the benefits and costs of public services respectively.

We now introduce three assumptions about governments. The first assumption is about preference heterogeneity. Each locality has a different ideal variety of public services. We define and order the basic varieties such that the ideal one for locality l is $x = l$. We assume that $\delta_l(x) = \delta > 0$ if $x = l$; and $\delta_l(x) = 0$ otherwise.

The second and third assumptions define the cost function K. First, there are economies of scale in the provision of public services. Building and maintaining a government reduces the utility from public services by a fixed total amount $\phi > 0$. This cost is equally shared among the constituent localities. Second, there are economies of scope across government functions. Membership of a union reduces the utility from public services by an amount $\kappa > 0$. This captures the costs of oversight and coordination between a country and the union.

These assumptions imply the following utility from public services:

$$W_l^G = -\frac{\delta}{g_l(l)} - \frac{\phi}{\int_0^1 I_{l=m}^P dm} - \kappa I_l^U, \tag{6}$$

where $I_{l=m}^P$ is an indicator variable that takes value 1 if if localities l and m belong to the same P_n, and zero otherwise; and I_l^U is an indicator variable that takes value 1 if locality l is a member of a union ($R_n \neq P_n$), and zero otherwise. The first term in Equation (6) means that the value of public services for locality l depends on the amount of its ideal variety that is provided. The second term means that each locality's share of the fixed cost of government declines with the size of the country. The parameter ϕ measures the magnitude of these economies of scale. The third term means that being a member of a union is costly. The parameter κ measures the magnitude of these economies of scope.

To complete the model we need to make assumptions on how localities interact. We consider law and diplomacy in Section 3, and war and conquest in Section 4. In both cases, the world's political structure is determined by the interplay of the forces that follow from our assumptions. Although these assumptions are standard in the literature, we provide some additional discussion of them in Appendix A.2.

3 Law and Diplomacy

Efficient bargaining among localities delivers Pareto efficient outcomes and constitutes a natural benchmark. In this case, the equilibrium political structure is obtained by solving the following maximization problem:

$$(P, R) = \arg\max \int_0^1 \omega_l W_l dl, \tag{7}$$

where $\{\omega_l\}_{l\in[0,1]}$ is a set of Pareto weights such that $\int_0^1 \omega_l dl = 1$. Given the symmetry of our model, we focus on the case in which the bargaining process treats all localities in the same way: $\omega_l = 1$ for all $l \in [0, 1]$. This political structure can be intepreted as the utilitarian welfare optimum because it maximizes average world welfare. We view it as the description of a world in which all localities have the right to choose their own political structure. This is a world ruled by law and diplomacy.

An implication of the maximization problem (7) is that each country P_n provides a uniform bundle that contains equal amounts of the ideal varieties of its constituent localities. That is, locality l receives the following bundle of public services:

$$g_l(x) = \begin{cases} \dfrac{1}{\int_0^1 I^P_{l=m} dm} & \text{if } I^P_{l=x} = 1 \\ 0 & \text{if } I^P_{l=x} = 0. \end{cases} \tag{8}$$

Thus, we can re-write Equation (6) as follows:

$$W_l^G = -\delta \int_0^1 I^P_{l=m} dm - \frac{\phi}{\int_0^1 I^P_{l=m} dm} - \kappa I_l^U. \tag{9}$$

The first term means that the value of public services for locality l declines with the size of its country. As more localities join the country, the public services provided are farther away from the ideal of each member locality. The parameter δ measures the importance of this preference mismatch.

Combining Equations (1), (4) and (9), we obtain:

$$W_l = -\eta + \gamma \left(1 - \beta + \beta \int_0^1 I^R_{l=m} dm\right) - \delta \int_0^1 I^P_{l=m} dm - \frac{\phi}{\int_0^1 I^P_{l=m} dm} - \kappa I_l^U. \tag{10}$$

Equation (10) shows how political structure determines welfare. It reveals the key trade-off

that underlies our theory. A desirable political structure should facilitate trade, accommodate preference heterogeneity and take advantage of economies of scale and scope. But all these goals cannot be achieved simultaneously and something must give.

3.1 Equilibrium Political Structure

Two preliminary results simplify the analysis of the maximization problem in Equation (7). The first is that P and R contain equal-sized elements. All localities prefer the same optimal country size, so the equilibrium is symmetric. We denote by S and U the sizes of each element $P_n \in P$ and $R_n \in R$ respectively.[10] The second result is that, as anticipated, P is a refinement of R. If it is ever worth paying the costs of having a two-level governance structure, this is because localities desire a lower-level government that provides public services adapted to their specific preferences, and a higher-level government that reduces border effects and facilitates trade. Thus, we can express welfare in this free world of law and diplomacy (as opposed to the world of war and conquest that we discuss in Section 4) as the following function of S and U:

$$W_l = W^F(S,U) = -\eta + \gamma(1-\beta+\beta U) - \delta S - \frac{\phi}{S} - \kappa I^U(S,U), \tag{11}$$

where I^U is an indicator function that takes value 1 if $S \neq U$, and zero otherwise.

Equation (11) implies that the equilibrium political structure features either $P = R$ or $P \neq R = \{[0,1]\}$. In the first case, the world is organized in a single-level governance structure with a set of countries that provide public services and regulate markets. In the second case, the world is organized in a two-level governance structure with countries providing public services and a world union regulating markets.[11]

We now find the equilibrium political structure in three steps. First, we compute the welfare $W^F(S_1^*, S_1^*)$ generated by the single-level governance structure, where S_1^* is the optimal country size without an international union. This political structure takes full advantage of economies of scope. Country size trades off preference heterogeneity against both economies

[10] Throughout, we disregard the constraint that the number of countries and unions, $1/S$ and $1/U$, must be a natural number. In Appendix A.3.2 we introduce this integer constraint and show that the equilibrium political strcuture remains symmetric and qualitatively analogous to the tractable approximation we use in the main text.

[11] We know that, if $P \neq R$, there is only one world economic union because the marginal cost of adding members is constant and the marginal benefit is growing with the size of the union. Thus, having many small unions is not optimal.

of scale and facilitating trade:

$$S_1^* = \sqrt{\frac{\phi}{\delta - \beta\gamma}}. \tag{12}$$

The size of countries in the absence of unions is increasing with economies of scale (ϕ) and the importance of trade ($\beta\gamma$). It is decreasing with preference heterogeneity (δ).[12]

Second, we compute the welfare $W^F(S_2^*, 1)$ generated by the two-level governance structure, where S_2^* is the optimal country size with a world union. This political structure gives up economies of scope in order to remove border effects and facilitate trade. Country size trades off preference heterogeneity and economies of scale:

$$S_2^* = \sqrt{\frac{\phi}{\delta}}. \tag{13}$$

The size of countries with a world union is increasing with economies of scale (ϕ) and decreasing with preference heterogeneity (δ). Country size is always smaller with a world union than without it. The reason is that the union removes one of the incentives for expanding country size, namely, facilitating trade.

The third step is to determine the equilibrium political structure. If $W^F(S_1^*, S_1^*) > W^F(S_2^*, 1)$, the world is partitioned into countries of size S_1^*. If $W^F(S_1^*, S_1^*) < W^F(S_2^*, 1)$, the world is partitioned into countries of size S_2^* that belong to a world union. Naturally, in the knife-edge case in which $W^F(S_1^*, S_1^*) = W^F(S_2^*, 1)$, both solutions are equilibrium political structures. A little algebra shows that the world union is preferred if and only if:

$$\kappa + 2\sqrt{\phi\delta} < \beta\gamma + 2\sqrt{\phi(\delta - \beta\gamma)}. \tag{14}$$

That is, the world union is preferred for high values of β, γ and δ; and low values of ϕ and κ. A world union is more useful if the border effect and trade opportunities are large and there is substantial preference heterogeneity. A world union is less useful if economies of scale and scope are sizable.

3.2 Globalization and Political Structure

With these results at hand, we can return to Figure 1 and ask again: Why did the first wave of globalization reduce the number of countries but not generate unions? Why did the

[12] Equation (12) assumes that $\delta > \phi + \beta\eta$, so there is enough preference heterogeneity to ensure that each country is always smaller than the whole world.

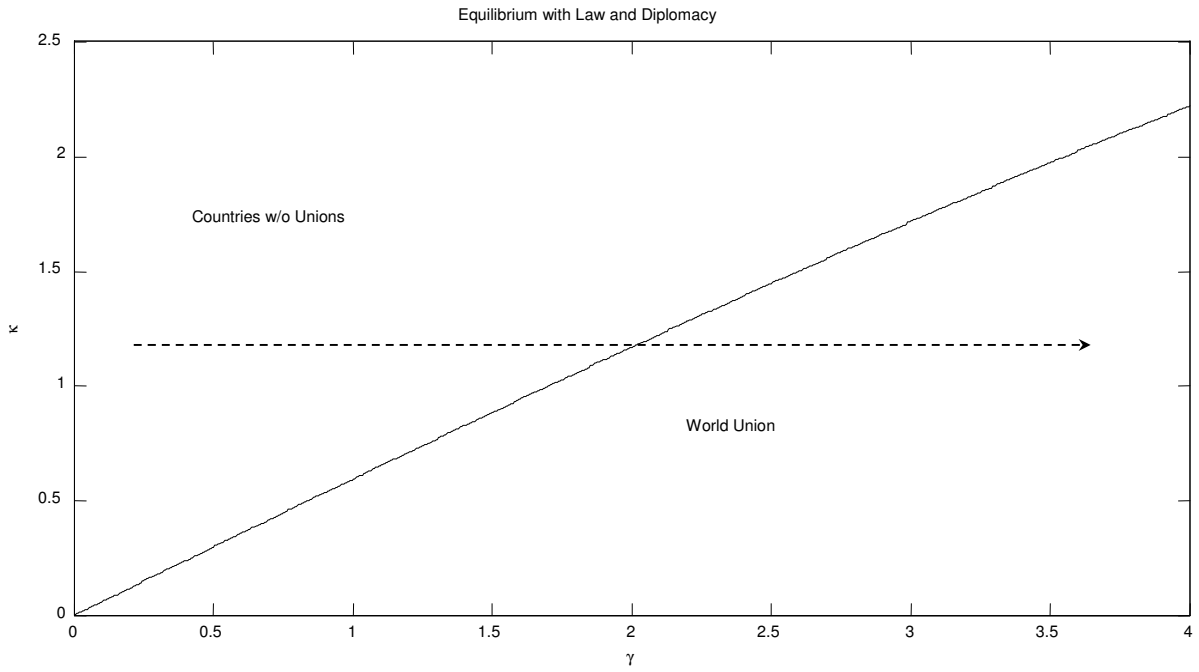

Figure 2: Globalization and Political Structure. The figure shows how equilibrium political structure depends on economies of scope (κ) and trade opportunities (γ).

second wave of globalization increase the number of countries and lead to the creation of unions? To answer these questions, we consider a gradual increase in trade opportunities γ from 0 to η, and we study how political structure changes as this process unfolds.

Figure 2 shows how equilibrium political structure depends on the two parameters that measure economies of scope (κ) and trade opportunities (γ). For a given κ, the world chooses a single-level political structure if γ is low and a two-level structure if γ is large. The dashed arrow in Figure 2 displays a trajectory for the world economy, along a path of incrasing trade opportunities with constant economies of scope. Equilibrium political structure is initially single-level, but eventually shifts to two-level as trade opportunities cross a threshold value γ_U defined as follows:

$$\beta\gamma_U + 2\sqrt{\phi\left(\delta - \beta\gamma_U\right)} = \kappa + 2\sqrt{\phi\delta}. \tag{15}$$

If economies of scope are nil, the smallest trade opportunities leads to the formation of a world union ($\kappa = 0$ implies $\gamma_U = 0$). If economies of scope are prohibitive, the world union is never an equilibrium ($\kappa > \beta\eta + 2\sqrt{\phi\left(\delta - \beta\eta\right)} - 2\sqrt{\phi\delta}$ implies $\gamma_U > \eta$). The comparative statics of this threshold follow directly from our analysis of Equation (14). The larger the border effect (β) and preference heterogeneity (δ), the smaller γ_U. The larger economies of

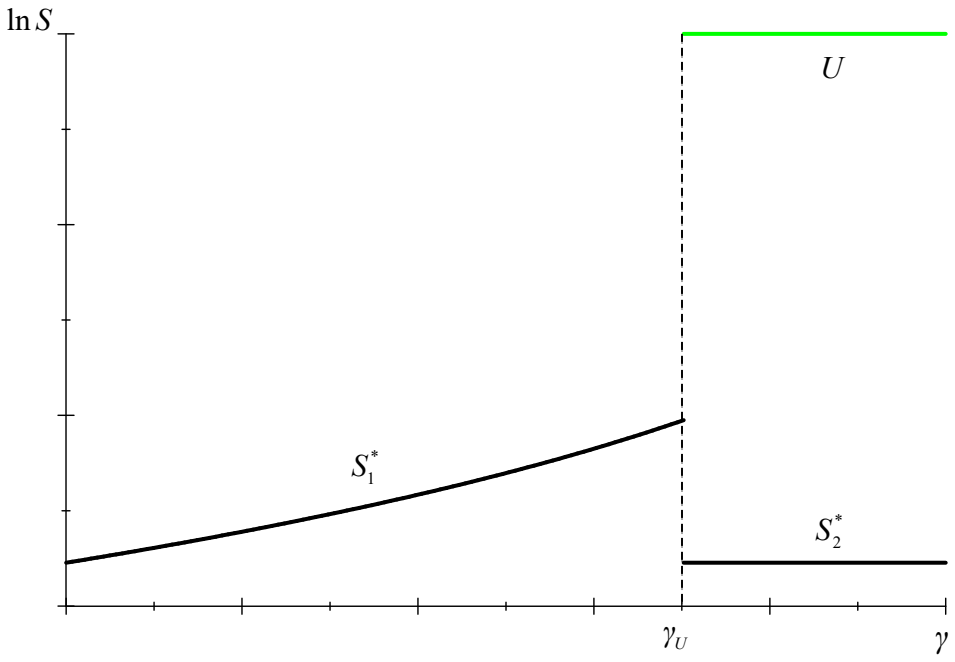

Figure 3: Globalization, Countries and Unions. The figure shows how the world political structure changes with trade opportunities (γ). The black line is the size of each country, the green line is the world union.

scale (ϕ) and scope (κ), the larger γ_U.

Figure 3 displays how political structure changes along the trajectory shown by the dashed arrow in Figure 2, by plotting the equilibrium size of countries and unions as a function of trade opportunities (γ). In the first wave of globalization ($\gamma < \gamma_U$), it is too expensive to create a world union, so increases in γ lead to an increase in country size. The cost of reaping additional gains from trade is a growing preference mismatch. Eventually, the preference mismatch has grown so large that it becomes cost-effective to create a world union. In the second wave of globalization ($\gamma > \gamma_U$), the cost of reaping additional gains from trade is the loss of economies of scope. The creation of a world union allows countries to revert to a smaller size and reduce the preference mismatch. Further increases in γ have no effect on political structure.

The pattern we derive as a consequence of declining transport costs is borne out by historical evidence on the two waves of globalization. In our theory, before the creation of international unions country borders impose constant trade frictions (equal to the border

effect β). Consistent with this prediction, O'Rourke and Williamson (1999) argue that the first wave of globalization was driven overwhelmingly by declining trade costs and not by the liberalization of cross-country trade. In our model, the creation of international unions endogenously eliminates the trade frictions created by country borders between their members. Consistent with this notion, O'Rourke and Williamson (1999) also suggest that trade liberalization became important after the 1950s—when it was achieved through international institutions such as the GATT/WTO.[13]

In deriving our results, we have taken transport costs as exogenous. This need not be the case. For instance, the model could be extended by adding the possibility of investing in trade-promoting technologies. Suppose governments can adopt a new technology with a lower τ, hence a higher γ, after paying a fixed cost. Suppose also that the new technology, once available, is adopted by all countries. This could be through coordination or simply because new technologies, once discovered, can be copied by others. As we can see from Equation (11), the benefit from improving the trade technology is proportional to the size of countries or unions. Once more, the reason is that the value of lowering trade costs is proportional to the size of the market that is affected. In turn, a better trade technology increases the size of countries, but also makes unions relatively more attractive. This suggests the existence of a complementarity between technological and political globalization. The adoption of a better trade technology can trigger the formation of a union. But the union can also make the adoption of such technology cost effective. Moreover, the rise of large countries may promote the adoption of trading technologies that could ultimately lead to their collapse. This suggests that the secular increase in trade opportunities need not be a gradual process, but rather can be marked by abrupt changes.

4 War and Conquest

We explore next how war and conquest affect the relationship between globalization and political structure. To do so, we assume that the world is divided into core and periphery. The core contains a measure π of localities with a superior military technology that can be

[13] It is instructive to compare our results to Alesina, Spolaore and Wacziarg's (2000) finding that expanding trade opportunities always reduce country size. Their analysis differs from ours in two crucial ways. First, they assume prohibitive economies of scope (in our model, $\kappa \to \infty$), so the world always has a single-level governance structure. Second, they model expanding trade opportunities as an exogenous reduction in the border effect (in our model, a decline in β with constant γ). Such a reduction always reduces country size, so their model can explain why the second wave of globalization led to smaller countries, but it cannot explain why the first wave of globalization led to larger countries.

used to conquer other localities and form empires. The periphery contains the remaining localities that do not have this military technology. We keep all assumptions regarding preferences, technology and the costs of government. Thus, the model of the previous section applies in the limit as $\pi \to 0$.

Empires are an alternative form of government that provides public services and regulates markets. Each empire contains a metropolis and its colonies. The metropolis consists of core localities that unite to conquer periphery localities, which then become their colonies.

One upside of empire-building is that the metropolis can force its colonies to finance an imperial government that provides none of their ideal public services. Instead, an empire provides the ideal public services of localities in its metropolis, at a lower cost to them thanks to economies of scale:

$$g_l(x) = \begin{cases} \dfrac{1}{\int_0^1 I^M_{l=m} dm} & \text{if } I^M_{l=m} = 1 \\ 0 & \text{if } I^M_{l=m} = 0, \end{cases} \tag{16}$$

where $I^M_{l=m}$ is an indicator variable that takes value 1 if localities l and m belong to the same metropolis and zero otherwise.

Another gain from empire is that the metropolis expropriates market goods from it colonies. We assume it loots a share $\lambda \in [0,1]$ of the colonies' output in the industries that do not rely on contract enforcement (a share $1-\beta$ of the total).[14] As a result, its consumption of the full measure of varieties in those industries rises to $c_l(m,i) = e^{-\tau}\left(1-\lambda+\lambda \int_0^1 I^E_{l=n} dn / \int_0^1 I^M_{l=n} dn\right)$ where $I^E_{l=n}$ is an indicator variable that takes value 1 if localities l and n belong to the same empire and zero otherwise.

The last benefit of building an empire is that it enables trade in the industries that require legal enforcement of contracts (a share β of the total). The metropolis enjoys the gains from trade creation that result from common market regulation in its empire, while it incurs reduced costs of preference mismatch.

The downside of empire-building for the metropolis is that waging war and holding the empire together reduces the utility that it derives from public services by an amount $\omega > 0$. This cost captures the diversion of government resources from providing public services in the metropolis to waging colonial wars.

[14] This assumption simplifies the algebra, but our results remain qualitatively unchanged if we assume that looting can take place in all industries, as we show in Appendix A.4.1.

Overall, the welfare of a member of an imperial metropolis is:

$$W_l = -\eta + (1-\beta)\left[\gamma + \ln\left(1 - \lambda + \lambda\frac{\int_0^1 I^E_{l=m}dm}{\int_0^1 I^M_{l=m}dm}\right)\right] + \beta\gamma\int_0^1 I^E_{l=m}dm - \delta\int_0^1 I^M_{l=m}dm - \frac{\phi}{\int_0^1 I^E_{l=m}dm} - \omega. \quad (17)$$

From the perspective of the conquered colonies, instead, the gains from trade creation are dwarfed by the costs of looting and especially of subjugation to an imperial government that generates an unbounded preference mismatch. This division of the surplus captures in our model the extractive nature of imperialism: the metropolis enjoys the gains from trade, while the conquered localities suffer under an exploitative colonial administration.

Periphery localities are incapable of resisting the core's superior military technology. Thus, a periphery locality remains free if and only if no core locality wishes to colonize it. This may be the case in equilibrium, because each empire has a limited reach. We assume that a metropolis of size M can subjugate a set of colonies of size θM. This assumption captures two considerations. The first is technological. Country size is important for military success, so one of the reasons countries grow large is to prepare for war. The second is ideological. As the disproportion between the size of the metropolis and its colonial targets grows, the aggressive and undemocratic nature of imperialism becomes more obvious, and eventually unpalatable on moral grounds for the citizens of the metropolis. Either force alone suffices for our results.

The equilibrium size of an empire does not depend only on the metropolis's ability and willingness to use violent repression in its colonies. It is also limited by competition between rival empires, which may clash in the attempt to conquer the same territories. If the sum of desired colonies is larger than the periphery, empires cannot avoid clashing. We assume that each metropolis can target which periphery localities it tries to colonize. If a periphery locality is targeted by multiple empires, each of them has an equal probability of colonizing it.

4.1 Equilibrium Political Structure

With war and conquest, equilibrium political structure need no longer be globally efficient. Formally, equilibrium political structure is now determined in two stages.

1. Core localities simultaneously choose whether to build empires and wage wars. Each

metropolis chooses which periphery localities (if any) it tries to conquer, taking the choice of other metropolises as given. In the ensuing Nash equilibrium, localities in the periphery may become colonies or remain free.

2. Localities that do not belong to an empire choose their political structure through efficient bargaining, as in our baseline model.

The world's equilibrium political structure now consists of a set of empires that have a combined size $1-F$; plus two partitions (P,R) of the free world which itself has a combined size F. We solve for this equilibrium political structure in two steps. First, we determine the political structure of the free world (P,R) for a given size F. Second, we determine the number and size of empires and therefore the size of the free world F.[15]

4.1.1 The Free World

The analysis of the free world is essentially the same as in the previous section. The only difference is that now the combined size of the free world is F rather than 1. Efficient bargaining ensures that free localities choose the optimal political structure. Equation (11) still applies and, as a result, there are two cases to consider: $S=U$ and $S<U=F$. The optimal country sizes in these cases are still given by Equations (12) and (13), respectively.[16]

The union of the free world is preferred if and only if:

$$\kappa+2\sqrt{\phi\delta}<\beta\gamma F+2\sqrt{\phi\left(\delta-\beta\gamma\right)}. \tag{18}$$

Condition (18) generalizes Condition (14) for the case of a free world of size F. The main difference is that empires reduce the size of the free world. This reduction lowers the welfare associated with a union of free localities. This union still costs κ to each member. But it is now less efficient at removing border effects (βF instead of β).

4.1.2 Empires

Core localities must first decide whether to wage war and build an empire, or to forego war and enter the free world. We start our analysis with three observations. First, a metropolis always targets for conquest all the colonies it can given its own size. Adding extra colonies

[15] For simplicity, we do not consider the possibility that empires could be members of international unions. Appendix A.4.2 shows that our results are robust to relaxing this assumption.

[16] We now assume that $\delta>\phi F^{-2}+\beta\eta$ to ensure that countries are always smaller than the free world.

is desirable from its perspective because it lowers the cost of government, enables looting and facilitates trade without creating any preference mismatch (for the metropolis) in the provision of public services. Second, in a Nash equilibrium the overlap between the targets of rival metropolises is minimized. In particular, if the sum of desired colonies is smaller than the entire periphery, rival metropolises competely avoid clashing with one another. Third, the Nash equilibrium generically features either no empire, or empires whose metropolises span the entire core.

These three observations imply that there are two cases. If $\pi \leq 1/(1+\theta)$, each metropolis conquers colonies of size θM, but empires do not need to clash because there remains a free world $F = 1-(1+\theta)\pi \geq 0$ that no empire is capable of subjugating. If instead $\pi > 1/(1+\theta)$, there is conflict between empires and the entire periphery is colonized ($F = 0$). We focus on symmetric Nash equilibria such that every metropolis carves out a share of the periphery proportional to its share of the core, so its colonies have size $(1-\pi)M/\pi$. Summarizing, this implies that the size of each metropolis relative to its empire is:

$$\frac{M}{E} = \mu = \max\left\{\frac{1}{1+\theta}, \pi\right\} < 1. \tag{19}$$

The welfare of a core locality that builds an empire is:

$$W_l = W^E(E) = -\eta + (1-\beta)\left[\gamma + \ln\left(1+\lambda\frac{1-\mu}{\mu}\right)\right] + \beta\gamma E - \delta\mu E - \frac{\phi}{E} - \omega. \tag{20}$$

The size of an empire trades off preference heterogeneity against both economies of scale and facilitating trade:

$$E^* = \sqrt{\frac{\phi}{\delta\mu - \beta\gamma}}, \tag{21}$$

where E^* is the optimal empire size (for the the core localities).[17]

Comparing Equation (21) to Equations (12) and (13), we see immediately that empires are larger than peaceful countries. The reason is that the metropolis does not internalize the cost of the preference mismatch it imposes on its colonies: hence, $\delta\mu$ appears instead of δ in the denominator. An intuitive implication is that an empire is larger when its metropolis can be smaller relative to its colonies (μ). Hence empires are smaller if the core is so large (π) that empires clash in equilibrium; or if they are sufficiently democratic they cannot repress

[17] Equation (21) assumes that $\delta\mu > \phi(\mu/\pi)^2 + \beta\eta$, so there is enough preference heterogeneity to ensure that the metropolis is smaller than the whole core and a fortiori the empire is smaller than the whole world.

a large share of their population (θ).

The equilibrium size of empires is also decreasing with preference heterogeneity (δ) and increasing with the importance of trade ($\beta\gamma$) and with economies of scale (ϕ). These comparative statics are analogous to those for peaceful countries. An empire is distinguished by its extractive institutions, but these determine the equilibrium size of empires only through economies of scale for the imperial government.

When are empires formed? If core localities wage war and build empires, their welfare is $W^E(E^*)$. If core localities instead refrain from waging war and choose to form countries and unions by efficient bargaining, their welfare is $\max\left\{W^F(S_1^*, S_1^*), W^F(S_2^*, 1)\right\}$.

If $W^E(E^*) > \max\left\{W^F(S_1^*, S_1^*), W^F(S_2^*, 1)\right\}$, empire-building is a dominant strategy for core localities. Thus, in equilibrium they form $\pi/(\mu E^*)$ empires of size E^*, and the size of the free world is reduced to $F = 1 - \pi/\mu$. Some algebra shows that empire-building is the unique equilibrium if and only if:

$$\omega - (1-\beta)\ln\left(1 + \lambda\frac{1-\mu}{\mu}\right) + 2\sqrt{\phi(\delta\mu - \beta\gamma)} < \min\left\{2\sqrt{\phi(\delta - \beta\gamma)}, \kappa + 2\sqrt{\phi\delta} - \beta\gamma\right\}. \quad (22)$$

Unsurprisingly, core localities wish to build empires if colonial wars are effective enough: i.e., if their military cost (ω) is low and their looting benefits (λ) are high, while the size of the metropolis (μ) is small relative to the empire, because the core (π) is small enough empires do not need to clash and their government is sufficiently autocratic (θ) to subjugate a large share of the empire's population.

If instead $W^E(E^*) \leq \max\left\{W^F(S_1^*, S_1^*), W^F(S_2^*, 1)\right\}$, there is always an equilibrium in which no empires are formed, diplomacy rules and the size of the free world is $F = 1$. This peaceful equilibrium need not be unique, though. A Pareto-inefficient second equilibrium also exists if $W^F(S_2^*, 1) \geq W^E(E^*) \geq \max\left\{W^F(S_1^*, S_1^*), W^F(S_2^*, 1 - \pi/\mu)\right\}$. In this case, Pareto efficiency requires all core localities to forego empire-building and create a peaceful world union. Yet, if core localities expect other core localities to build empires, their best response is to build an empire themselves. Empire-building is then a coordination failure that lowers the welfare of every locality in the world; but it is also an equilibrium because once it happens there is no incentive for a single core locality to abandon its empire and join the free world.

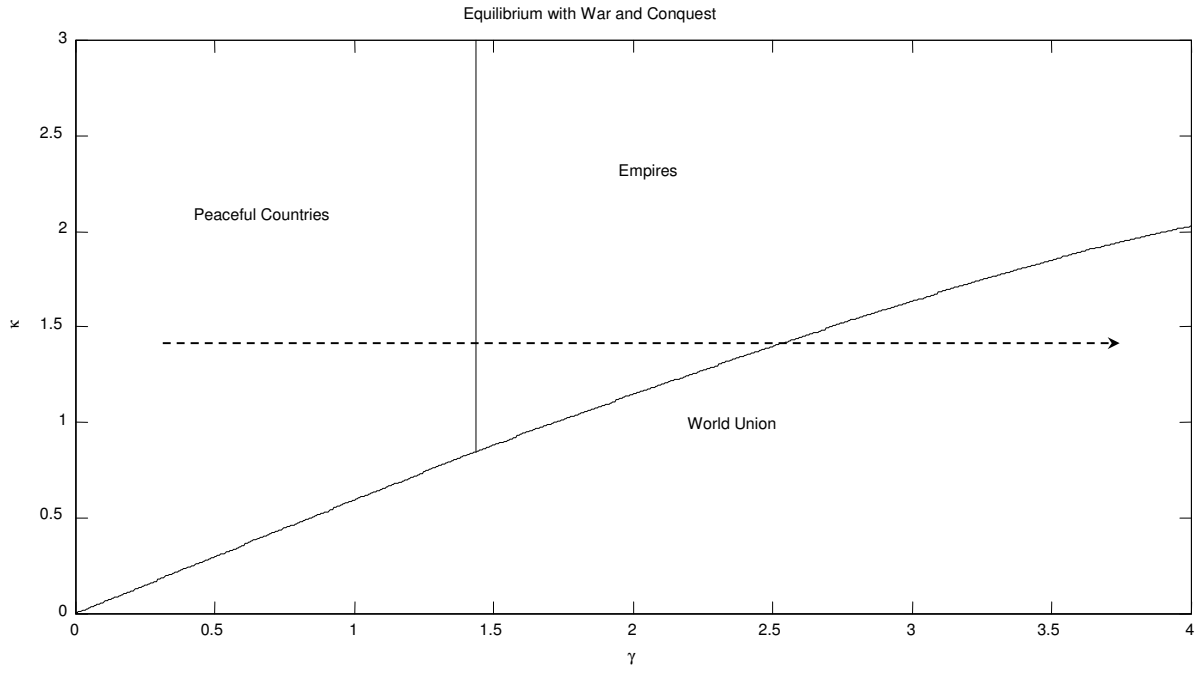

Figure 4: Globalization, Conflict and Political Structure. The figure shows how equilibrium political structure depends on economies of scope (κ) and trade opportunities (γ).

4.2 Globalization and Political Structure

Returning to the relationship between globalization and political structure, Figure 4 shows again how equilibrium political structure depends on the two parameters that measure economies of scope and trade opportunities (i.e., κ and γ). For simplicity, we disregard Pareto-dominated equilibria and focus throughout on the unique Pareto-dominant equilibrium. Then there is an age of empires if and only if the following condition holds:

$$\omega < (1-\beta)\ln\left(1+\lambda\frac{1-\mu}{\mu}\right)+2\sqrt{\phi}\left(\sqrt{\delta-\beta\gamma_U}-\sqrt{\delta\mu-\beta\gamma_U}\right). \tag{23}$$

Otherwise, empires are never built and the world is always ruled by law and diplomacy.

The dashed arrow in Figure 4 displays again a trajectory for the world economy along a path of incrasing trade opportunities, with constant economies of scope. Initially the whole world is free and there is no union. As trade opportunities reach a first threshold value $\gamma_L > 0$, empires are formed. Eventually, as trade opportunities reach a a second threshold value $\gamma_H < \eta$, empires are abandoned and a world union is formed. This evolution is not

generic, though. A necessary and sufficient condition for $\gamma_L > 0$ is that:

$$\omega > (1-\beta)\ln\left(1+\lambda\frac{1-\mu}{\mu}\right) + 2\sqrt{\phi\delta}\,(1-\sqrt{\mu})\,. \tag{24}$$

If this condition holds, core localities do not conquer colonies for the extractive benefits alone. Empires become cost-effective for the metropolis only when they have sufficient additional benefits from trade creation: hence, after trade opportunities have grown enough.

Likewise, a necessary and sufficient condition for $\gamma_H < \eta$ is that:

$$\omega > (1-\beta)\ln\left(1+\lambda\frac{1-\mu}{\mu}\right) + 2\sqrt{\phi}\left(\sqrt{\delta} - \sqrt{\delta\mu - \beta\eta}\right) - \beta\eta + \kappa. \tag{25}$$

If this condition holds, expanding trade opportunities make empires grow so large that the preference mismatch within the metropolis alone becomes enough to justify a move to a two-level governance structure. Perhaps surprisingly, the cause of imperial collapse at a late stage of globalization is the same as the cause for the rise of empires at an early stage: namely, the desire to remove border effects and reap the gains from trade. It is just that, at some point, it becomes more cost-efficient to replace conquered colonies with free partners in a union.

Figure 5 displays how political structure changes along the trajectory shown by the dashed arrow in Figure 4. At low levels of globalization ($\gamma < \gamma_L$), the world contains only free countries. There are no empires nor unions. As trade opportunities expand, the size of countries grows. When trade opportunities cross the first threshold ($\gamma_L < \gamma < \gamma_H$), core localities prefer to build empires. Empires are larger than countries and keep growing as globalization proceeds. Eventually, trade opportunities cross the second threshold ($\gamma > \gamma_H$). Empires collapse and countries revert to a smaller size. A world union is created. After this, there are no further changes in political structure.[18]

4.3 Other Drivers of Political Structure

Our analysis has focused on growing opportunities for trade as the explanation for the non-monotonic evolution of political structure. However, several additional factors shape the

[18] Figure 5 depicts the case in which, during the age of empires, the free world always adopts a single-level governance structure and the shift to the two-level governance structure coincides with the collapse of empires. This need not be the case. If the size of the free world is large enough, an economic union of free countries may co-exist with empires.

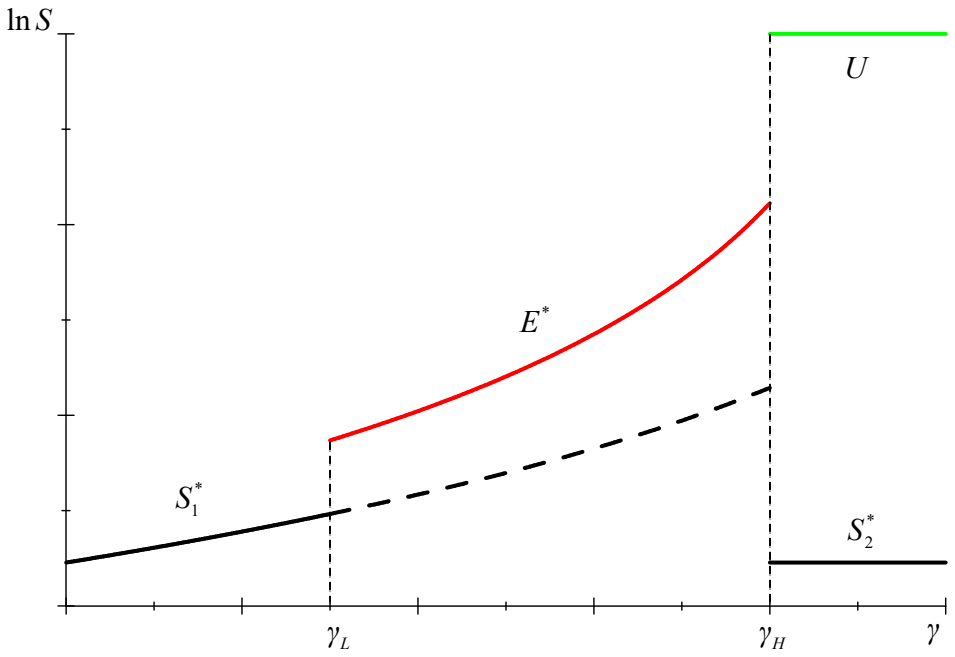

Figure 5: Countries, Empires and Unions. The figure shows how the world political structure changes with expanding trade opportunities (γ). The black line is the size of peaceful countries, the red line is the size of empires, the green line is the world union.

number of countries and the decision to form colonial empires or international unions. A natural question is whether these other factors alone could be driving the patterns we observe in historical data. In particular, could the three-stage evolution of political structure observed in history and depicted in Figure 4 be explained purely by changes in the costs and benefits of war and conquest?

To answer this question, Figure 6 shows how equilibrium political structure depends on the cost of war (ω) as well as trade opportunities (γ). The rise and subsequent fall of empires could be explained by a decline and subsequent increase in ω alone, i.e., by movements along a vertical trajectory in Figure 6. Such a non-monotonic evolution of the cost of war plausibly fits the historical consequences of advancing military technology. In the nineteenth century, it enabled the great powers to make easy colonial conquests. Conversely, in the twentieth century it led them to incur horrific costs in two World Wars, and eventually to develop nuclear weapons and the doctrine of mutually assured destruction.

Yet, changes in the cost of war alone cannot explain why what follows empires (interna-

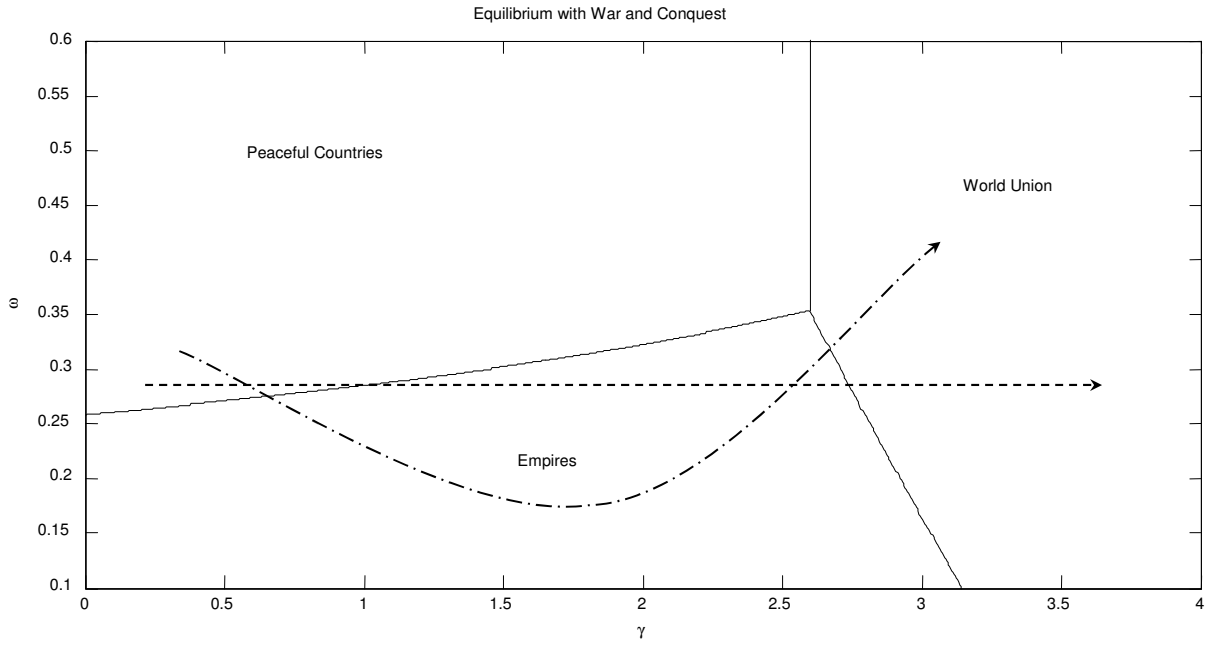

Figure 6: Trade, War and Political Structure. The figure shows how equilibrium political structure depends on the cost of war (ω) and trade opportunities (γ).

tional unions) is different from what preceded them (peaceful countries). Expanding trade opportunities are necessary to account for this pattern. They are also sufficient, as Figure 6 confirms by plotting once again as a dashed arrow the straight trajectory along a path of increasing trade opportunities, with a constant cost of war. Nonetheless, joint changes in γ and ω along a U-shaped trajectory—displayed as a dash-dotted arrow in Figure 6—paint a more realistic picture. The two drivers of political structure are complementary: changes in the cost of war accelerate the consequences of changes in trade opportunities. In the nineteenth century, empires arise earlier as conquering colonies becomes cheaper. In the twentieth century, international unions arise earlier as war becomes costlier.

This complementarity sheds light on the relationship between peace, trade and international organizations. In our model, a goal of international unions is to reduce the appeal of costly wars. Empirically, this was a key real-world motivation for creating supranational institutions like the United Nations and the European Union. Figure 6 helps understand why they were created only after World War II. By that point, trade opportunities had grown large enough that a positive shock to the cost of war sufficed to trigger a switch from the age of empires to that of peaceful unions.[19]

[19] Our theory also points to a non-monotonic complementarity between military and trade technologies.

The rise and fall of empires may reflect not only the gradual improvement of military technology, but also its gradual spread. Our model captures the traditional view that conflict among rival empires was caused by the spread of industrialization and the attendant rise of new great powers on the world scene. As more localities acquire the technology to conquer empires, the size of the core (π) expands. So long as the core remains small enough ($\pi < 1/(1+\theta)$), rival powers can secure their desired colonies without clashing. Accordingly, Britain and France could carve out two world-straddling empires without fighting each other, the Fashoda incident notwithstanding. In this case, the spread of great-power technology increases the number of empires without affecting their size. However, once technology diffusion reaches a critical threshold ($\pi > 1/(1+\theta)$) rival powers cannot avoid costly clashes as they try to secure scarce colonies for themselves and deny them to each other. Accordingly, the rise of imperial powers eventually came to be marked by wars like the Spanish-American war and the Russo-Japanese war—and it was arguably a cause of both World Wars.

In our model, such clashes unambiguously reduce the size of the empire each core locality can conquer ($\partial E^*/\partial\pi < 0$). At the same time, they force a metropolis to remain relatively large, so as to remain competitive with rivals in the military domain ($\partial M^*/\partial\pi > \partial E^*/\partial\pi$).[20] Intuitively, costly great-power rivalry thus makes imperialism less appealing ($\partial W^E(E^*)/\partial\pi < 0$). Just as an increase in the cost of warfare, a greater prevalence of clashes between imperial powers (such as the World Wars) hastens the demise of empires ($\partial\gamma_H/\partial\pi < 0$). Once again, though, the simultaneous expansion of trade opportunities is not only a complementary driver of decolonization, but also the factor required to explain the creation of international unions.

Finally, recall that the parameter θ can be interpreted as the limit to imperialism arising from moral concerns among citizens of the metropolis. A decline in this limit can represent the spread of democratic values. In our model, a decline in θ makes empires weakly smaller ($\partial E^*/\partial\theta \geq 0$) and less desirable ($\partial W^E(E^*)/\partial\theta \geq 0$). This effect is always present if

In an early stage of globalization ($\gamma < \gamma_U$), trade-enabling technology (higher γ) complements offensive war technology (lower ω). An improvement in either tempts core localities to forego diplomacy and peace and embark instead on imperial conquest, thereby raising the value of an increase in the other. In a later stage of globalization ($\gamma < \gamma_U$) the pattern switches to complementarity between trade-enabling technology and defensive war technology (higher ω), both of which hasten the demise of empires and the advent of peaceful unions. While a dynamic theory of endogenous technological progress remains beyond the scope of our model, these considerations suggest that a realistic U-shaped trajectory in Figure 6 could derive from directed technical change, with causal feedback loops between trading technology and different types of military technologies.

[20] Great-power rivalry can prove so demanding that an imperial metropolis needs to expand while its empire shrinks: $\partial M^*/\partial\pi > 0 > \partial E^*/\partial\pi$ if $\pi > \max\left\{1/(1+\theta), \sqrt{2\beta\gamma/\delta}\right\}$.

the binding constraint on imperialism arises from moral concerns in the metropolis ($\pi < 1/(1+\theta)$). If instead the binding constraint is competition between the great powers, the effect only arises if the increase in democratic values is large enough. Since competition between empires grew over time, our theory suggests that democratization may have played a greater role in slowing the rise of empires in the nineteenth century than in hastening their demise in the twentieth.

5 Regional Unions

Perhaps the most unrealistic aspect of our baseline model is its lack of gradualism. If core localities join a union, it is a world union. If a world union is created, country size is reduced at once to its autarky level. However, a simple generalization of our theory shows that this need not be the case. Sometimes, regional unions precede world unions.[21]

To introduce regional unions, we now consider a slightly richer treatment of geography. Assume that, in addition to having a superior military technology, core localities are closer to each other. Transportation costs for core-core trade are $\tau - \rho$ (with $\rho > 0$), so the gains from core-core trade are now $\gamma + \rho$. Transportation costs for core-periphery and periphery-periphery trade remain τ, so the gains from these types of trade are still γ. The rest of our assumptions remain those of Section 4.

Appendix A.3.3 provides a full analysis of this model. Here we just state the new results it generates. The key implication of this new assumption is that Equation (4), indicating the utility from consuming market goods, must be replaced by:

$$W_l^M = -\eta + \gamma\left(1 - \beta + \beta\int_0^1 I_{l=m}^R dm\right) + \rho I_l^C\left[(1-\beta)\,\pi + \beta\int_0^\pi I_{l=m}^R dm\right], \qquad (26)$$

where I_l^C is an indicator variable that equals 1 if l belongs to the core and zero otherwise. Border effects are larger for core-core trade ($\beta(\gamma+\rho)$) than for either core-periphery or periphery-periphery trade ($\beta\gamma$). Other things equal, core localities prefer sharing economic regulation with other core localities. Periphery localities, instead, are indifferent about which localities they share regulation with and care only about their total number.

With this additional assumption, a mixed political structure with a two-level governance structure in the core and a single-level governance structure in the periphery becomes possi-

[21] An alternative form of gradualism is the creation of unions with increasing depth of international integration. Appendix A.4.3 provides a formalization of this idea.

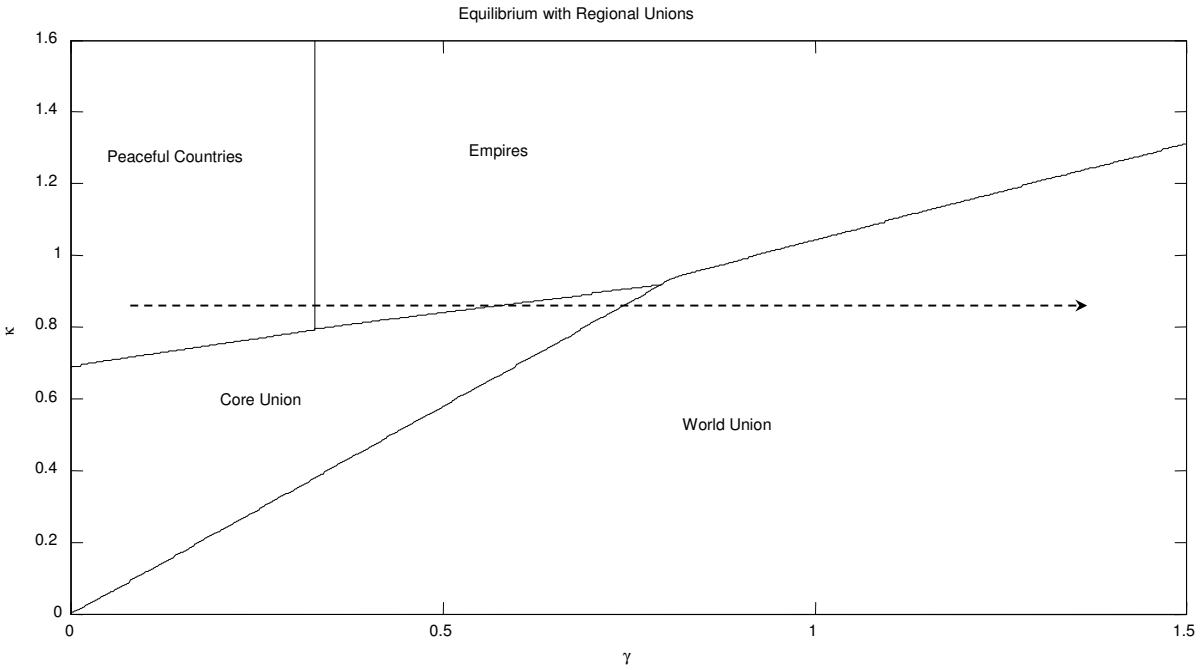

Figure 7: Gradualism and Regionalism. The figure shows how equilibrium political structure depends on economies of scope (κ) and trade opportunities (γ).

ble. Figure 7, which assumes a large enough value for ρ, shows a scenario in which a four-fold evolution takes place along the trajectory of expanding trade opportunities depicted as usual by the dashed arrow. The first stage is peaceful single-level government everywhere. The second is the creation of colonial empires through which the core conquers and rules distant localities in the periphery. As trade opportunities expand further, empires become overstretched and collapse, and a peaceful union replaces them. The novelty of this third stage is that, unlike in Figure 4, this is now a core union. Indeed, for the most part, in the real world core countries replaced empires with freer trade not with their former colonies, but with other core countries. Only as trade opportunities expand further a fourth and final stage is reached in which the whole world forms a union.

The model could be further extended in fruitful ways. One extension is a world with two, three or N peripheries that are located progressively farther away. In this world, there is a union that starts at the core and grows outwardly with expanding trade opportunities. When the first periphery joins the union, the size of its countries declines. When the second periphery joins the union, the size of its countries also declines, and so on. The union gradually advances outward and it keeps breaking up countries. The end point is the same as in our baseline model, but the world approaches it gradually.

A second extension is a world with two or more core-periphery structures, which we can think of as continents or regions. In this world, within each region there is one union that advances outward, breaking up countries. Across regions, however, there is no union initially. Eventually, trade opportunities may expand so much that a world union becomes cost-effective, and the regional unions merge. The world approaches the same end point again, but it now approaches it both gradually and regionally.

The bottom line is that asymmetric geography can explain the gradual appearance of international unions with a limited geographic scope. As distance becomes less and less important, the world gradually converges to the single international union we analyzed earlier.

6 Historical Evidence on Trade and Country Size

Our theory provides a new perspective on the connection between trade, country size and the emergence of international unions. Our motivation for developing this theory was to improve our understanding of global trends, as presented in Figure 1. But the mechanisms that connect trade and country size over time should also be at work when we compare the trajectories of different countries. The extension with regional unions in Section 5 considers trade costs that vary across countries. Moreover, localities could differ in their costs of government, of waging war and of forming unions. This would introduce heterogeneity in country size and also imply that different countries may join unions at different times.

We exploit such heterogeneity by examining cross-country historical data.[22] Our empirical evidence show that (i) increases in the volume of trade predict increases in country size in the absence of international unions, but (ii) the creation of international unions weakens or eliminates this pattern, and instead (iii) membership of international unions predicts decreases in country size. These findings, though far from conclusive, suggest that the mechanisms our theory highlights are empirically relevant.

6.1 Data

We draw our data from the Cross-National Time-Series (CNTS) Data Archive. This dataset provides information about land area and the volume of trade, measured as the sum of imports and exports per capita, for an unbalanced panel of countries with observations from

[22] In a more speculative vein, in Appendix A.5 we provide a brief qualitative narrative showing how our theory helps interpret historical events.

Table 1: Descriptive Statistics.

Period	Countries	Land Area	Expansion	Contraction	WTO	Δ Trade	
	Number	Mean	Mean	Mean	Mean	Mean	Std. Dev.
1870-1880	43	738,564	0.233	0.023	0.000	0.280	0.421
1880-1890	45	687,672	0.244	0.044	0.000	0.271	0.430
1890-1900	46	674,440	0.217	0.109	0.000	0.084	0.303
1900-1910	46	670,962	0.087	0.130	0.000	0.403	0.346
1910-1920	47	681,729	0.234	0.085	0.000	1.410	1.262
1920-1930	55	604,987	0.164	0.073	0.000	0.466	1.041
1950-1960	73	536,990	0.041	0.068	0.425	0.880	0.971
1960-1970	102	450,449	0.010	0.049	0.564	1.043	1.268
1970-1980	127	377,772	0.024	0.055	0.591	4.613	4.504
1980-1990	149	274,399	0.020	0.020	0.624	0.276	0.668
1990-2000	158	257,461	0.006	0.019	0.800	1.851	14.732
2000-2010	169	294,461	0.006	0.012	0.845	1.698	1.327

Notes: Countries and Land Area are measured at the beginning of each decade. Land Area is expressed in thousand square miles. Expansion and Contraction are dummies taking value 1 if the land area of an existing country expandend or contracted, respectively, over the decade. WTO is a dummy taking value 1 if a country is a member of GATT/WTO at the beginning of each decade.

1870 to 2010.[23]

Given that land area changes slowly and discontinuously, we focus on decades and build a dummy variable that takes value 1 if a country or empire has grown in size over the decade. We interpret averages over this variable as the "probability" of a territorial expansion. Similarly, we build dummy variables for territorial contraction that take value 1 if the land area of a country or empire has fallen over the decade. We interpret averages over this variable as the "probability" of a territorial contraction.

Table 1 contains descriptive statistics for the main variables of interest. For each decade, it reports the number of countries with non-missing observations, their average land area, the share of countries that expanded and that contracted their territories, the share of countries that are members of the WTO (GATT before 1995), the average change in the volume of trade and its standard deviation. Given that all years corresponding to the world wars have no observations in the dataset, the decades around 1940 are missing.

A quick look at Table 1 confirms the basic trends already discussed. Territorial expansions are much more frequent than contractions in the first part of the sample, but the opposite is

[23]Before 1870, trade data is missing for the majority of countries.

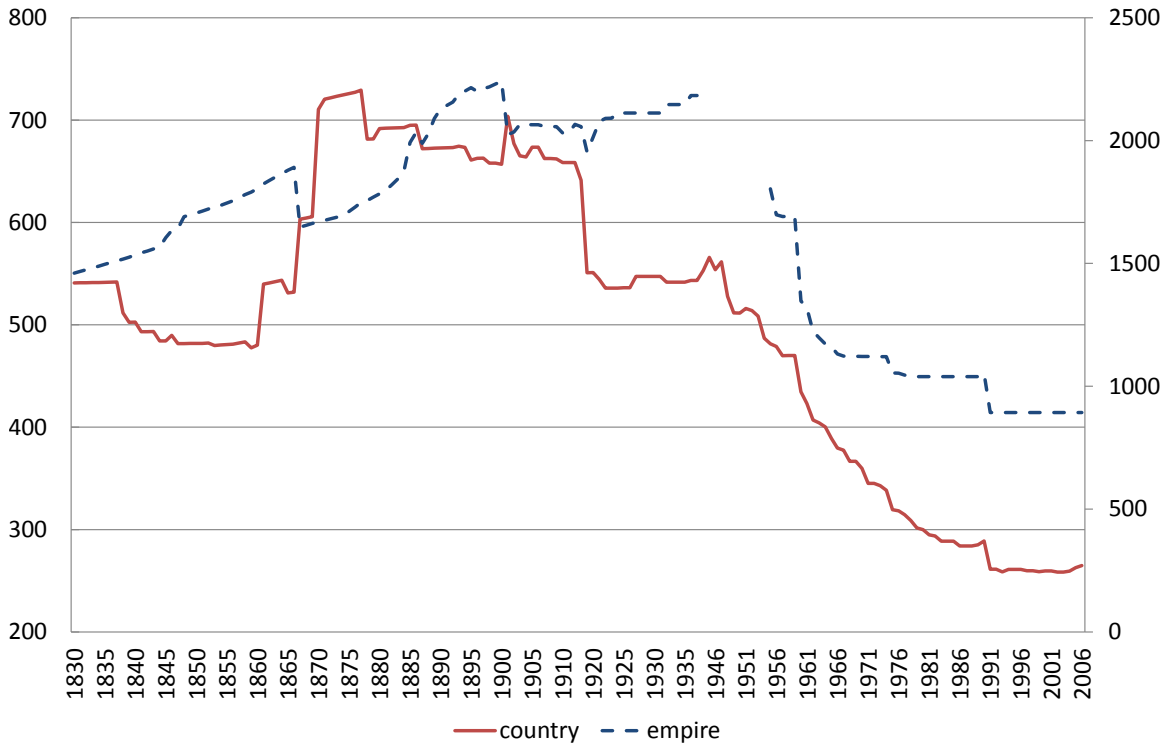

Figure 8: The Size of Countries and Empires. The figure plots the average size of countries (left axis) and of empires (right axis) in thousand squared miles. See Appendix A.1 for details on data.

true after 1950. The table also reports the average growth in the volume of trade per capita. It shows that trade grows throughout the entire period, but at different speeds both over the decades and across countries.[24]

Figure 8 shows the average land area of "internationally recognized" countries and of thirteen major empires since 1830. The size of empires increases until World War II and falls thereafter. Empires, besides being larger than countries, start reducing their size somewhat later than countries. This picture is consistent with the data on the number of countries shown in Figure 1. Despite the different data sources, both figures tell a remarkably similar story.[25] During the nineteenth century there was a phase of political concentration in which countries and empires expanded their territories. But this trend reversed during the twentieth

[24] Caution should be taken when comparing some of these variables across decades since the composition of the sample is changing.

[25] Figure 8 is based on the "International System" (Singer and Small 1966), which includes countries with international recognition and reliable data. See Appendix A.1 for more details. This classification, however, underestimates the number of independent political entities in the nineteenth-century developing world. Figure 1 is instead based on Butcher and Griffiths (2013), who provide a more comprehensive record.

century, and especially after World War II.

Finally, our theory also implies that territorial changes should be associated with conflict during the age of empires, but be more peaceful in era of international unions. Data from the Correlates of War project shows that indeed before 1950 more than one third of all territorial disputes involved military conflict, while after that date border changes were peaceful in almost 90% of cases.

6.2 Regression Analysis

We start by studying how the probability of observing a territorial expansion depends on changes in the volume of trade, by running a series of binomial regressions. To alleviate simultaneity, we compute the change in the volume of trade over the previous decade. Furthermore, to check if the correlation between trade and territorial expansion changes after World War II, we include the interaction between changes in the volume of trade and a post-1945 dummy.

Table 2 reports the main results of estimating a logit model. Standard errors are corrected for clustering by country, to accommodate autocorrelated shocks at the country level. In column 1, we start by including as regressors the lagged change in the volume of trade and its interaction with the post-1945 dummy. As predicted by our theory, the coefficient for the lagged change in trade is positive, the one for the interaction term is negative, and both are highly significant. In other words, trade predicts territorial expansion before World War II, but not after it. In column 2, we also include the post-1945 dummy. The inclusion of the constant and the post-war dummy mean that the identifying variation is deviations from the global trends visible in Figures 1 and 8. Nevertheless, the two coefficients of interest remain significant. In column 3, we add year fixed effects and we keep them in the rest of the table. While these time dummies control for shocks affecting all countries equally, including some common effects of globalization, they do not affect our main results. In fact, the coefficients for the trade variables become even more significant.

We then add some control variables inspired by our model. In column 4, we add the log of population and the urbanization rate at the beginning of each decade.[26] The first variable controls for the effect of size while the second is a proxy for economic development, and both are likely to be correlated with the military strength of a country.[27] Both coefficients are

[26] These variables are also sourced from the CNTS Data Archive. The urbanization rate is the fraction of population living in urban centers of at least 100,000 inhabitants.

[27] More direct measures, such as GDP, are not available for the entire period of analysis.

Table 2: Trade and Territorial Expansion.

Dependent variable: Expansion dummy								
	All	All	All	All	All	All	Pre1945	Post1945
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Trade	0.818***	0.285*	0.463*	0.607**	0.545**	0.650*	0.577**	-0.179
	[0.189]	[0.176]	[0.257]	[0.239]	[0.269]	[0.335]	[0.290]	[0.142]
Δ Trade × Post1945	-1.294***	-0.287*	-0.580*	-0.896***	-0.776**	-1.636**		
	[0.314]	[0.177]	[0.300]	[0.314]	[0.318]	[0.819]		
Post1945		-2.724***	-5.262***	-5.626***	-4.850***	-2.859		
		[0.472]	[1.242]	[1.566]	[1.353]	[2.405]		
Log Population				0.595***	0.599***	-2.053	0.640***	0.460*
				[0.141]	[0.141]	[1.702]	[0.161]	[0.271]
Urbanization Rate				0.002	0.002*	0.003	0.003	0.001
				[0.001]	[0.001]	[0.004]	[0.002]	[0.002]
Δ Democracy					-0.166**	-0.159	-0.267**	0.113
					[0.076]	[0.132]	[0.111]	[0.103]
Country FE	No	No	No	No	No	Yes	No	No
Time FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	822	822	822	799	651	227	212	439
R^2	0.090	0.218	0.260	0.362	0.326	0.386	0.214	0.102

Notes: All observations refer to 10-year periods. The dependent variable is a dummy taking value 1 if the country's land area expanded over the decade and 0 otherwise. Δ Trade and Δ Democracy are changes over the previous decade. Post1945 is a dummy for decades after 1945. All other variables are measured at the beginning of each decade. Constant always included and Pseudo-R^2 reported. Standard errors, clustered by country, are in brackets. *, ** and *** denote significance at 10%, 5% and 1% respectively.

positive and significant, but their inclusion does not affect the trade variables. In column 5 we add the change in the level of democracy over the previous decade.[28] Consistent with our model and other existing theories, we find that democratization lowers the probability of territorial expansion. Yet, its inclusion has almost no effect on the coefficient for trade. In column 6 we add country fixed effects. This specification is quite demanding: all countries with no changes in size are dropped and the coefficients are identified only from within-country deviations from country-specific trends. As a result, sample size falls markedly. Nonetheless, trade and its interaction with the post-1945 dummy are the only two coefficients

[28] We use the polity2 index sourced from the Polity IV database. Unfortunately, this index is missing for some of the country-year cells in our dataset.

that remain significant.

Finally, we allow for heterogeneity in all coefficients before and after World War II. To this end, in columns 7 and 8 we split the sample and re-estimate the specifications without country fixed effects separately for the decades before and after 1945. The results confirm that trade predicts territorial expansion in the first part of the sample, but not in the postwar era. As to the remaining coefficients, population has a larger and more significant coefficients in the pre-1945 period, while democracy becomes insignificant after 1945. A possible interpretation of these results is that variables capturing military strength or tolerance for aggression become less relevant when territorial changes are more peaceful. However, the lack of significance may also indicate low statistical power due to the very few territorial expansions observed in the post World War II period. To study how trade correlates with border changes after 1945, we need to turn to territorial contractions, which become relatively more frequent.

In Table 3, we run similar binomial regressions to study how the probability of territorial contraction depends on changes in the volume of trade and on being part of an international economic union, measured by WTO membership. As in the previous table, the change in the volume of trade is computed over the previous ten years, the WTO dummy is measured at the beginning of each decade, and standard errors are corrected for clustering by country. We restrict the analysis to the period after 1945, when the WTO dummy starts to have positive values, but results are similar if earlier decades are added.[29] In column 1, we include the lagged change in the volume of trade and the WTO dummy. Only the coefficient for the latter is positive and precisely estimated, indicating that countries joining the WTO have a higher probability of a subsequent territorial contraction. In column 2, we add the interaction between changes in trade and the WTO dummy. The coefficient for trade turns positive and become statistically significant. When we add year fixed effects, in column 3, the negative coefficient of the interaction term also become significant. Growth in trade is followed by a higher probability of territorial contraction, but not after a country joins the WTO. This finding suggests that the negative effect of trade on country size may be mediated by membership of international unions like the WTO, as our theory predicts.

In column 4 we add the log of population and the log of GDP per capita at the beginning of each decade. These proxies for size and economic development, which where associated with territorial expansion before World War II, are now correlated with territorial contraction.

[29] Before 1950, territorial contractions are infrequent and occur in periphery countries, for which data are scarce and less reliable. If we perform the analysis over the entire sample, we obtain similar results, but these are driven by the observations after 1950. Hence, adding more decades is not very informative.

Table 3: Trade, Unions and Territorial Contraction.

Dependent variable: Contraction dummy						
	Post1945 (1)	Post1945 (2)	Post1945 (3)	Post1945 (4)	Post1945 (5)	Post1945 (6)
Δ Trade	-0.003	0.097*	0.146***	0.105**	0.104**	0.080
	[0.011]	[0.055]	[0.051]	[0.052]	[0.052]	[0.058]
WTO	1.555**	1.594**	2.120***	1.744**	1.785**	2.423**
	[0.724]	[0.760]	[0.754]	[0.884]	[0.886]	[1.139]
Δ Trade × WTO		-0.113	-0.146**	-0.234	-0.251	-0.468
		[0.072]	[0.057]	[0.284]	[0.287]	[0.355]
Log Population				0.503***	0.490***	0.558***
				[0.143]	[0.150]	[0.210]
Log GDP per capita				0.549***	0.557***	0.140
				[0.188]	[0.205]	[0.412]
Δ Democracy					-0.106	-0.166
					[0.139]	[0.220]
Region FE	No	No	No	No	No	Yes
Time FE	No	No	Yes	Yes	Yes	Yes
Observations	588	532	532	530	486	355
R^2	0.038	0.032	0.155	0.248	0.239	0.255

Notes: All observations refer to 10-year periods. The dependent variable is a dummy equal to 1 if the country's land area contracted over the decade and 0 otherwise. WTO is a dummy for WTO/GATT membership. Δ Trade and Δ Democracy are changes over the previous decade. All other variables are measured at the beginning of each decade. Constant always included and Pseudo-R^2 reported. Standard errors, clustered by country, in brackets. *, ** and *** denote significance at 10%, 5% and 1% respectively.

This suggest that the countries or empires that where growing before 1940 may be precisely those that start to break up after it. In column 5 we also add the change in the democracy index. While democratization is followed by a lower probability of territorial expansion in the era of empires, it has no statistically significant correlation with territorial contraction after 1950. Adding these controls barely affects the coefficient for the WTO dummy, while its interaction with trade becomes statistically insignificant. Country fixed effects are not feasible because they would shrink our sample size to a mere 63 observations. Instead, in column 6 we add region fixed effects. While the estimated coefficient for the change in the volume of trade loses precision, the WTO dummy remains positive and statistically significant. In sum, throughout all specifications, the most robust result is the positive correlation between WTO membership and subsequent territorial contractions.

We have shown three significant patterns in historical data since 1870. First, increases in trade predict territorial expansion before international unions. Second, this correlation is reversed after international unions are created. Third, membership of international unions predicts territorial contraction. To the best of our knowledge, these empirical findings are novel in the literature. While not conclusive, they are remarkably consistent with our theory. The predictions of our model are also consistent with other empirical findings already documented elsewhere. Besides the well-documented trade-creating effect of international unions, of particular interest is Martin, Mayer and Thoenig's (2012) finding that regional trade agreements promote peaceful relations between member states.

7 Conclusions

In this paper we have studied the interaction between globalization and political structure. We have shown that the expansion of trade opportunities can help explain two salient aspects of the evolution of political structure over the last couple of centuries: (i) the rise and subsequent fall in the size of countries observed during the nineteenth and twentieth century, and (ii) the seemingly contradictory trends towards more political fragmentation at the national level and more political integration at the supranational level in the second half of the twentieth century. We have also provided some historical evidence in support of this claim. Yet, we have deliberately left aside several important factors. We now briefly mention three that seem particularly promising for future research.

First, we have modeled international unions as mere economic agreements to facilitate trade. Although this simplification provides a useful and realistic starting point, it does

not do justice to the full political scope of unions, which in reality also provide public goods and coordinate policy choices with cross-border consequences.[30] Hence, supranational policymaking involves distributive conflict that we have abstracted from, and which can help explain gradual changes in union membership.

Second, our concept of locality abstracts from internal heterogeneity in both preferences and economic attributes. Yet, historical experience suggests that internal conflict has played a role in the process of country formation and break-up (Bolton and Roland 1996, 1997). It would be interesting to explore how expanding trade opportunities also affect political structure through their effect on domestic heterogeneity and conflict. We know, for instance, that international unions may exacerbate domestic conflict and shape political preferences (Gancia, Ponzetto and Ventura 2020).

Finally, we have focused on the economic aspects of globalization: an expansion of trade opportunities and how political structure adjusts to take advantage of them. Yet, there are also cultural aspects of globalization, which may entail for instance a reduction in preference heterogeneity. Exploring the simultaneous effects of both economic and cultural drivers of globalization on political structure seems a fruitful avenue for further research.

References

[1] Alesina, Alberto, Ignazio Angeloni and Federico Etro (2005). "International Unions," *American Economic Review* 95(3): 602–15.

[2] Alesina, Alberto, and Enrico Spolaore (1997). "On the Number and Size of Nations," *Quarterly Journal of Economics* 112(4): 1027–1056.

[3] —— (2003). *The Size of Nations*. Cambridge, MA: MIT Press.

[4] Alesina, Alberto, Enrico Spolaore and Romain Wacziarg (2000). "Economic Integration and Political Disintegration," *American Economic Review* 90(5): 1276–1296.

[5] Boffa, Federico, Amedeo Piolatto, and Giacomo A. M. Ponzetto (2016). "Political Centralization and Government Accountability," *Quarterly Journal of Economics* 131(1): 381–422.

[30] Globalization is likely to make the coordination role of unions more important because it exacerbates international policy externalities (Epifani and Gancia 2009; Broner and Ventura 2011).

[6] Bolton, Patrick, and Gerard Roland (1996). "Distributional Conflicts, Factor Mobility and Political Integration," *American Economic Review* 86(2): 99–104.

[7] —— (1997). "The Breakup of Nations: A Political Economy Analysis," *Quarterly Journal of Economics* 112(4): 1057–1090.

[8] Bonfatti, Roberto and Kevin H. O'Rourke (2018). "Growth, Import Dependence and War," *Economic Journal*, 128(614): 2222–2257.

[9] Broner, Fernando and Jaume Ventura (2011). "Globalization and Risk Sharing," *Review of Economic Studies* 78(1): 49-82.

[10] Butcher, Charles, and Ryan Griffiths (2013). "Introducing the International System(s) Dataset (ISD), 1816–2011," *International Interactions* 39(5): 748–768.

[11] Casella, Alessandra (2001). "The Role of Market Size in the Formation of Jurisdictions," *Review of Economic Studies* 68(1): 83–108.

[12] Casella, Alessandra, and Feinstein, Jonathan S. (2002). "Public Goods in Trade: On the Formation of Markets and Jurisdictions," *International Economic Review* 43(2): 437–462.

[13] Coase, Ronald H. (1960). "The Problem of Social Cost," *Journal of Law and Economics* 3: 1–44.

[14] Desmet, Klaus, Michel Le Breton, Ignacio Ortuño-Ortín and Shlomo Weber (2011). "The Stability and Breakup of Nations: A Quantitative Analysis," *Journal of Economic Growth* 16(3): 183–213.

[15] Epifani Paolo and Gino Gancia (2009)."Openness, Government Size and the Terms of Trade," *Review of Economic Studies* 76: 629–668.

[16] Gancia, Gino, Giacomo Ponzetto and Jaume Ventura (2020). "A Theory of Economic Unions," *Journal of Monetary Economics* 109: 107–127.

[17] Gennaioli, Nicola and Joachim Voth (2015). "State Capacity and Military Conflict," *Review of Economic Studies* 82(4): 1409–1448.

[18] Griffiths, Ryan (2014). "Secession and the Invisible Hand of the International System," *Review of International Studies* 40(3): 559–581.

[19] Hooghe, Liesbet, and Gary Marks (2003). "Unraveling the Central State, but How? Types of Multi-Level Governance," *American Political Science Review* 97(2): 233–243.

[20] Kahler, Miles, and David A. Lake (2004). "Governance in a Global Economy: Political Authority in Transition," *PS: Political Science & Politics* 37(3): 409–414.

[21] Lake, David A., and Angela O'Mahony (2004). "The Incredible Shrinking State: Explaining Change in the Territorial Size of Countries," *Journal of Conflict Resolution*, 48(5): 699–722.

[22] Lake, David A., and Angela O'Mahony (2006). "Territory and War: State Size and Patterns of Interstate Conflict," in *Territoriality and Conflict in an Era of Globalization*, edited by Miles Kahler and Barbara F. Walter, 133–55. Cambridge: Cambridge University Press.

[23] Lockwood, Benjamin (2006). "The Political Economy of Decentralization." In *Handbook of Fiscal Federalism*, edited by Ehtisham Ahmad and Giorgio Brosio, 33–60. Cheltenham: Edward Elgar.

[24] Maddison, Angus (2001). *The World Economy: A Millennial Perspective*. Paris: OECD.

[25] Martin, Philippe, Thierry Mayer and Mathias Thoenig (2008). "Make Trade Not War?" *Review of Economic Studies* 75(3): 865–900.

[26] —— (2012). "The Geography of Conflicts and Regional Trade Agreements," *American Economic Journal: Macroeconomics* 4(4): 1–35.

[27] Oates, Wallace E. (1972). *Fiscal Federalism*. New York, NY: Harcourt Brace Jovanovich.

[28] O'Rourke, Kevin H., and Jeffrey G. Williamson. (1999). *Globalization and History: The Evolution of a Nineteenth-Century Atlantic Economy*. Cambridge, MA: MIT Press.

[29] Polachek, Solomon W. (1980). "Conflict and Trade," *Journal of Conflict Resolution* 24(1): 55–78.

[30] Posner, Richard A. (2014). *Economic Analysis of Law*, 9th edition [1st edition 1973]. New York, NY: Wolters Kluwer Law & Business.

[31] Rohner, Dominic, Mathias Thoenig and Fabrizio Zilibotti (2013). "War Signals: A Theory of Trade, Trust, and Conflict," *Review of Economic Studies*, 80(3): 1114–1147.

[32] Ruta, Michele (2005). "Economic Theories of Political (Dis)integration," *Journal of Economic Surveys* 19(1): 1–21.

[33] Singer, David J., and Melvin Small (1966). "The Composition and Status Ordering of the International System: 1815–1940," *World Politics* 18(2): 236–282.

[34] Treisman, Daniel (2007). *The Architecture of Government: Rethinking Political Decentralization.* Cambridge: Cambridge University Press.

A Appendix – For Online Publication

A.1 Data Sources

The trade share reported in Figure 1 is merchandise exports as percent of GDP in 1990 prices, from Maddison (2001). Maddison provides trade data for selected countries in the years 1820, 1870, 1913, 1929, 1950, 1973 and 1998. To avoid compositional effects, we report the value of merchandise export as a share of GDP for the set of countries with data for all those years (Austria, Belgium, France, Italy, Spain, Switzerland, the United Kingdom and the United States). The trade share computed using the data for all available countries in every year is very similar to the one displayed in Figure 1.

The number of countries is reported for the same years. Data on the number of countries in the twentieth century is not very controversial. For the nineteenth century, however, some leading conventions grossly underestimate the number of countries. For example, the "International System," developed by Singer and Small (1966) and adopted in the Correlates of War project or in the Cross-National Time-Series Data Archive, only includes countries with international recognition. In particular, prior to 1920, the criteria to be recognized as an independent country were to have population greater than 500,000 and to have had diplomatic missions at or above the rank of chargé d'affaires with Britain and France. Clearly, this definition is too strict for our purposes, which require the identification of even relatively small political units living in economic and political autarky. We follow Butcher and Griffiths (2013), who recognize the problem and offer alternative criteria to identify the number of countries between 1816 and 2011.

The number of WTO members is from the WTO website.

The size of countries and empires displayed in Figure 8 is from the Cross-National Time-Series (CNTS) Data Archive. It provides data on contiguous territorial area in thousand square miles for all countries existing in a given year according to the International System. In a few instances, missing data have been imputed by cross-checking major territorial changes from other sources (China and Persia before 1860). Area of empire is provided for a consistent sample of 13 countries: Austria (formerly Austria-Hungary), Belgium, France, Germany (formerly Prussia), Italy (formerly Sardinia), Japan, the Netherlands, Portugal, Russia, Spain, Turkey (formerly the Ottoman Empire), the United Kingdom, and the United States. For these countries, empire area includes "overseas" territories (i.e., colonies). Data for the two World War periods, 1914–1918 and 1940–1945 (1938–1954 for empires) are missing.

For the analysis in Section 6, the volume of trade is computed as the value of imports plus exports per capita. Population is in millions. Urbanization is aggregate population in cities of 100,000 and over divided by total population. For empires, trade, population and urbanization refer to the metropolis only. Land area includes overseas territories.

A.2 Discussion of Assumptions

The starting point of our analysis is the idea that improvements in transport technology are a major driver of globalization, which we model as a fall in trade costs between localities. This view is uncontroversial: there is clear evidence that the secular rise in trade volumes was made possible by better transport technologies. For instance, some of the major drivers of the first wave of globalization are the adoption of the steamship (Pascali 2017), the telegraph (Steinwender 2018) and the spread of railroads (Donaldson 2018). Likewise, the main drivers of the second wave of globalization include containerization in ocean shipping (Levinson 2006), the development of jet aircraft engines (Hummels 2007) and more recently the ICT revolution. All these innovations promoted trade both between and within countries.

We also assume that borders obstruct trade. The large negative effect of political borders on trade volumes is well know at least since the work of McCallum (1995), who showed that, controlling for distance and income, trade between two Canadian provinces is 20 times larger than trade between a Canadian province and a U.S. state. While the exact magnitude of the border effect is still subject to debate, all existing studies coincide in finding large effects. For instance, in a recent survey of the empirical literature on gravity equations, Head and Mayer (2014) report that countries are typically found to trade 5 to 7 times more with themselves than with any other country.

There is equally strong evidence that sharing economic regulations and signing economic agreements promotes trade and reduces the border effect. For instance, Head and Mayer (2014) also report that sharing a common currency or being part of a free trade area is associated on average with a doubling of the volume of trade. Likewise, Helpman, Melitz and Rubinstein (2008) find that having a similar legal system increases the bilateral volume of trade by more than 60 percent. Using a simple model, Anderson and van Wincoop (2004) attempt a rough decomposition of the border effect. They argue that the overall cost of borders is equivalent to an ad valorem tariff of 44 percent, which can be broken down into 8 percent policy related barriers (including non-tariff barriers), 7 percent language barriers, 14 percent currency barriers, 6 percent information cost barriers, and 3 percent security barriers.

Our modeling assumption that technological barriers reduce trade along the intensive margin while policy-induced barriers affect the extensive margin is also grounded in empirical evidence. There is a wide consensus that transport costs affect significantly the intensive margin of trade. On the other hand, Helpman, Melitz and Rubinstein (2008) and Dutt, Mihov and Van Zandt (2013) find that free-trade agreements and WTO membership predominantly affect the extensive margin.

Turning next to governments, our model follows the standard assumptions that underpin the literature on federalism and the architecture of government since Oates (1972). Having separate local governments enables better preference matching, but sharing a common government enables beneficial policy coordination and reaps economies of scale. While these assumptions originated as simple observations of real-world patterns, models of political economy have provided them with rigorous micro-foundations. Majority rule makes centralization costly when localities have different preferences (Lockwood 2002; Besley and Coate 2003). Frictions in bargaining between political leaders (Harstad 2007) and in their agency relationship to their constituents (Boffa, Piolatto and Ponzetto 2016) explain both why multiple local government cannot fully coordinate their policies and why a single central government cannot fully tailor public services to local preferences. Accordingly, Strumpf and Oberholzer-Gee (2002) find empirically that U.S. states with more heterogeneous preferences are more likely to decentralize policy-making, while Lassen and Serritzlew (2011) show that municipal amalgamations cause citizens to experience a worse fit between what they want and what they get from local government. Evidence from sub-national governments also confirms the existence of economies of scale in government administration, albeit—consistent with our model—not in the provision of public goods more broadly (Reingewertz 2012; Blom-Hansen, Houlberg and Serritzlew 2014; Blom-Hansen et al. 2016; Blesse and Baskaran 2016).

Our assumption of economies of scope in government is equally classic (Musgrave 1971; Dahl and Tufte 1973; Alesina and Spolaore 2003). Marks and Hooghe (2004, p. 18) "emphasize the costs of decomposing authority" as a paramount concern in the analysis of multi-level governance, especially in the international arena with its prevalence of intersecting task-specific jurisdictions. Empirical evidence shows that multiplying administrative tiers reduces their efficiency, and is particularly associated with lower labor productivity and excess government employment (Le Galès and John 1997; Andrews and Boyne 2009). This cost is particularly pronounced for special-purpose governments in charge of a single task (Berry 2009). These efficiency losses reflect both the costs of administrative duplication and economies of scope in political accountability. Boffa, Piolatto and Ponzetto (2016) show the-

oretically that dividing policy-making responsibilities across multiple levels of government increases overall rent extraction by government officials. Fan, Lin and Treisman (2009) report that across countries corruption increases with the number of administrative tiers: as they rise from two to six, the probability of a firm reporting that it is never expected to pay bribes falls by 32 percentage points.[31]

A.3 Analytical Derivations

A.3.1 Computing Equilibrium Consumption

Locality l maximizes the objective function:

$$W_l^M = \int_0^1 \int_0^1 \ln c_l(m,i)\, dmdi, \tag{A1}$$

subject to the budget constraint:

$$\int_0^1 \int_0^1 p_l(m,i)\left[c_l(m,i) - q_l(m,i)\right] dmdi \leq 0, \tag{A2}$$

where $q_l(m,i)$ and $p_l(m,i)$ are the production and price of input m of industry i in locality l. The productions $q_l(m,i)$ must be consistent with available technology as described in Section 2. Since individuals are atomistic, they take prices as given in their maximization problems.

Equilibrium prices are as follows:

$$p_l(m,i) = \begin{cases} 1 & \text{if } l = m \\ e^{\eta} & \text{if } i \in [0,\beta] \text{ and } I^R_{l=m} = 0 \\ e^{\tau} & \text{if } i \in (\beta,1] \text{ or } I^R_{l=m} = 1 \text{ but } l \neq m. \end{cases} \tag{A3}$$

To prove this claim, normalize world income to unity ($Y = 1$). Note first that each locality has unit density of expenditure on each input in each industry. We next examine production. Consider first industries that require contract enforcement, $i \in [0,\beta]$. Locality l employs unit density of labor to produce each non-traded input m for which $I^R_{l=m} = 0$. Thus, output of

[31] The evidence about the importance of economies of scope has been gathered mostly at the sub-national level. Casual observation suggests, however, that economies of scope also apply at the supra-national level. An example is the European Central Bank, which aims at reducing the border effect by eliminating currency barriers. Its creation does not seem to have reduced the size or costs of national central banks, but instead it seems to have just added to these costs.

each non-traded input has density $e^{-\eta}$, so the value of output has unit density given price e^{η}. The remaining mass $\int_0^1 I_{l=m}^R dm$ of industry-i labor is employed to produce an identical mass of output. Unit density of it is sold domestically at a unit price. The remainder is shipped in identical amounts to other localities with $I_{l=m}^R = 1$, each of which receives a density $e^{-\tau}$ of imports, hence import value of unit density given price e^{τ}. In industries that do not require contract enforcement, $i \in (\beta, 1]$, the whole unit mass of industry-i labor is employed to produce the locality's own input variety, which is sold in identical amounts to all other localities in the world. Thus, the value of sales in each locality of each input in each industry has unit density, just like expenditure. This proves our claim.[32]

With these prices at hand, we can compute the equilibrium productions and consumptions described in Section 2.

A.3.2 Dealing with Integer Constraints

A world without unions consists of $N^* \in \mathbb{N}$ countries. Country n consists of measure $S_n > 0$ of localities, such that $\sum_{n=1}^{N^*} S_n = 1$. Utilitarian social welfare is $W = \sum_{n=1}^{N^*} S_n W^F(S_n, S_n)$.

The welfare function:

$$W^F(S_n, S_n) = -\eta + \gamma(1 - \beta + \beta S_n) - \delta S_n - \frac{\phi}{S_n} \tag{A4}$$

is concave in S_n and increasing at $S_n = 0$. Whenever $\delta > \phi + \beta\gamma$ it has a unique maximum at $S_n = S_1^*$.

Pareto efficiency then requires that either $S_n \leq S_1^*$ for all $n = 1, 2, ..., N$ or $S_n \geq S_1^*$ for all n. Otherwise some localities could leave a country with excessive size $S_n > S_1^*$ and join another with insufficient size $S_n < S_1^*$, raising the welfare of every locality in both countries.

Utilitarian welfare maximization requires all countries to have the same size. If there are two countries m and n such that $S_m > S_n > S_1^*$, then transferring the marginal locality from m to n not only raises its welfare, but it also raises the welfare of S_m localities by more than it lowers the welfare of $S_n < S_m$ localities. Likewise if $S_m < S_n < S_1^*$.

Therefore, once integer constraints are taken into account, the utilitarian welfare optimum

[32] It is straightforward to show that this equilibrium is unique. First, rule out variation in the prices of traded inputs since this would generate excess demand (supply) of cheap (expensive) varieties. Second, rule out that the relative prices of traded and nontraded varieties be above (below) τ/η since this would lead to an excess demand (supply) of nontraded inputs.

without unions is a partition of the world into a number:

$$N_1^* = \arg\max_{N\in\mathbb{N}} \left\{ -\eta + \gamma\left(1 - \beta + \frac{\beta}{N}\right) - \frac{\delta}{N} - \phi N \right\} \tag{A5}$$

of identical countries. The objective function W has strictly decreasing differences in (N, γ) because for any $\gamma_H > \gamma_L$ and and $N_H > N_L$:

$$W(N_H, \gamma_H) - W(N_H, \gamma_L) = \beta\frac{\gamma_H - \gamma_L}{N_H} < W(N_L, \gamma_H) - W(N_L, \gamma_L) = \beta\frac{\gamma_H - \gamma_L}{N_L}. \tag{A6}$$

Thus, the welfare-maximizing number of countries N_1^* is decreasing in γ in the sense of monotone comparative statics. It is likewise decreasing in β, and increasing in δ and ϕ.

By the same reasoning, the utilitarian welfare optimum with unions is a world union composed of a number:

$$N_2^* = \arg\max_{N\in\mathbb{N}} \left\{ -\frac{\delta}{N} - \phi N \right\} \tag{A7}$$

of identical countries.

A.3.3 Equilibrium Conditions for the General Model

The most general model in this paper is discussed in Section 5 when we assume that $\pi \geq 0$ and $\rho \geq 0$. The model of war and conquest of Section 4 applies in the limit as $\rho \to 0$, while the model of diplomacy in Section 3 applies in the limit as $\pi \to 0$ and $\rho \to 0$. In this Appendix, we discuss the different possible equilibria of the general model.

Law and Diplomacy Assume first that the core decides to forego warfare and join the free world, so $F = 1$. Then we can define the welfare of core localities as:

$$W^C(S_C, U_C) = -\eta + (\gamma + \rho\pi)(1 - \beta) + (\gamma U_C + \rho \min\{U_C, \pi\})\beta - \delta S_C - \frac{\phi}{S_C} - \kappa I_C^U, \tag{A8}$$

and the welfare of periphery localities as:

$$W^P(S_P, U_P) = -\eta + \gamma(1 - \beta + \beta U_P) - \delta S_P - \frac{\phi}{S_P} - \kappa I_P^U, \tag{A9}$$

where I_C^U and I_P^U are indicator functions that take value 1 if $S_C = U_C$ and $S_P = U_P$, respectively, and zero otherwise. Utilitarian social welfare for the entire world equals:

$$W\left(S_C, S_P, U_C, U_P\right) = \pi W^C\left(S_C, U_C\right) + \left(1-\pi\right) W^P\left(S_P, U_P\right). \tag{A10}$$

There are three possible equilibrium political structures. The first is a single-level government structure with countries of optimal sizes:

$$S_{1C}^* = \sqrt{\frac{\phi}{\delta - \beta\left(\gamma+\rho\right)}} \text{ and } S_{1P}^* = \sqrt{\frac{\phi}{\delta - \beta\gamma}}, \tag{A11}$$

and without any unions.[33] Utilitarian world welfare under this first structure is:

$$\begin{aligned} W^1 &\equiv W\left(S_{1C}^*, S_{1P}^*, S_{1C}^*, S_{1P}^*\right) \\ &= -\eta + \left(\gamma + \rho\pi^2\right)\left(1-\beta\right) - 2\left\{\pi\sqrt{\phi\left[\delta - \beta\left(\gamma+\rho\right)\right]} + \left(1-\pi\right)\sqrt{\phi\left(\delta-\beta\gamma\right)}\right\}. \end{aligned} \tag{A12}$$

The second potential equilibrium is a two-level government structure with countries of optimal size:

$$S_{2C}^* = S_{2P}^* = \sqrt{\frac{\phi}{\delta}}, \tag{A13}$$

and with a world union: $U_C = U_P = 1$. Utilitarian world welfare under this second structure is:

$$W^2 \equiv W\left(S_{2C}^*, S_{2P}^*, 1, 1\right) = -\eta + \left(\gamma + \rho\pi^2\right) - 2\sqrt{\phi\delta} - \kappa. \tag{A14}$$

The third potential equilibrium features a two-level government structure for the core with countries of size $S_{3C}^* = S_{2C}^*$ and a core union $U_C = \pi$, but a single-level government structure for the periphery with countries of size $S_{3P}^* = S_{1P}^*$ and no union ($U_P = S_{3P}^*$). Utilitarian world welfare under this third structure is:

$$\begin{aligned} W^2 &\equiv W\left(S_{2C}^*, S_{2P}^*, 1, 1\right) \\ &= -\eta + \left(\gamma + \rho\pi^2\right)\left(1-\beta\right) + \left(\gamma+\rho\right)\pi^2\beta - 2\left[\pi\sqrt{\phi\delta} + \left(1-\pi\right)\sqrt{\phi\left(\delta-\beta\gamma\right)}\right] - \pi\kappa. \end{aligned} \tag{A15}$$

The equilibrium political structure under diplomacy is the one that delivers the highest

[33] Equation (A11) assumes that $\delta > \phi\pi^{-2} + \beta\left(\eta+\rho\right)$, so that there is enough preference heterogeneity to ensure that the optimal core country size is always smaller than the entire core.

welfare:

$$\arg\max W \ni \begin{cases} (S_{1C}^*, S_{1P}^*, S_{1C}^*, S_{1P}^*) & \text{if } W^1 \geq \max\{W^2, W^3\} \\ (S_{2C}^*, S_{2P}^*, 1, 1) & \text{if } W^2 \geq \max\{W^1, W^3\} \\ (S_{3C}^*, S_{3P}^*, \pi, S_{3P}^*) & \text{if } W^3 \geq \max\{W^1, W^2\}. \end{cases} \tag{A16}$$

The core union yields higher welfare than single-level government ($W^3 > W^1$) if:

$$\kappa < \bar{\kappa}_1(\gamma) \equiv \beta(\gamma + \rho)\pi - 2\sqrt{\phi}\left[\sqrt{\delta} - \sqrt{\delta - \beta(\gamma + \rho)}\right], \tag{A17}$$

for an increasing and concave threshold ($\partial\bar{\kappa}_1/\partial\gamma > 0 > \partial^2\bar{\kappa}_1/\partial\gamma^2$) such that $\partial\bar{\kappa}_1/\partial\beta > 0$, $\partial\bar{\kappa}_1/\partial\delta > 0$, $\partial\bar{\kappa}_1/\partial\phi < 0$, $\partial\bar{\kappa}_1/\partial\pi > 0$ and $\partial\bar{\kappa}_1/\partial\rho > 0$.

The core union yields higher welfare than the world union ($W^3 > W^2$) if:

$$\kappa > \bar{\kappa}_2(\gamma) \equiv \beta\gamma(1 + \pi) - 2\sqrt{\phi}\left(\sqrt{\delta} - \sqrt{\delta - \beta\gamma}\right), \tag{A18}$$

for an increasing and concave threshold ($\partial\bar{\kappa}_2/\partial\gamma > 0 > \partial^2\bar{\kappa}_2/\partial\gamma^2$) such that $\partial\bar{\kappa}_2/\partial\beta > 0$, $\partial\bar{\kappa}_2/\partial\delta > 0$, $\partial\bar{\kappa}_2/\partial\phi < 0$, $\partial\bar{\kappa}_1/\partial\pi > 0$ and $\partial\bar{\kappa}_2/\partial\rho = 0$.

Single-level government yields higher welfare than the world union ($W^1 > W^2$) if:

$$\kappa > \pi\bar{\kappa}_1 + (1 - \pi)\bar{\kappa}_2. \tag{A19}$$

The two functions $\bar{\kappa}_1(\gamma)$ and $\bar{\kappa}_2(\gamma)$ have a single crossing because:

$$\bar{\kappa}_1(0) > \bar{\kappa}_2(0) = 0 \text{ and } \frac{\partial\bar{\kappa}_2}{\partial\gamma} > \frac{\partial\bar{\kappa}_1}{\partial\gamma}. \tag{A20}$$

In other words, the core union can follow but not precede single-level government and precede but not follow the world union because:

$$\frac{\partial}{\partial\gamma}\left(W^2 - W^3\right) = \beta\left[1 - \pi^2 - (1 - \pi)S_{3P}^*\right] > 0 \tag{A21}$$

and

$$\frac{\partial}{\partial\gamma}\left(W^3 - W^1\right) = \beta\pi(\pi - S_{1C}^*) > 0. \tag{A22}$$

War and Conquest If there are empires, the analysis is as essentially as it was in Section 4 for $\rho = 0$. The welfare of core localities that form an imperial metropolis is given by:

$$W_l = W^E(E) = -\eta + (1-\beta)\left[\gamma + \rho\pi + \ln\left(1 + \lambda\frac{1-\mu}{\mu}\right)\right] + \beta(\gamma + \rho\mu)E - \delta\mu E - \frac{\phi}{E} - \omega. \tag{A23}$$

The optimal size of empires is larger because so are gains from trade within the metropolis:

$$E^* = \sqrt{\frac{\phi}{\delta\mu - \beta(\gamma + \rho\mu)}}. \tag{A24}$$

Thus, if core localities build empires their welfare is given by:

$$W^E(E^*) = -\eta + (1-\beta)\left[\gamma + \rho\pi + \ln\left(1 + \lambda\frac{1-\mu}{\mu}\right)\right] - 2\sqrt{\phi\left[\delta\mu - \beta(\gamma + \rho\mu)\right]} - \omega. \tag{A25}$$

The free world contains a measure $F < 1 - \pi$ of localities in the periphery. In this case, all the analysis in Section 4 applies and, in particular, Condition (18) still determines whether the free world has a single or two-level government structure.

When are empires formed? In the absence of empires, the welfare of core localities in the welfare-maximizing political structure is given by:

$$W_C^F = \begin{cases} W^C(S_{1C}^*, S_{1C}^*) & \text{if} \quad W^1 > \max\{W^2, W^3\} \\ W^C(S_2^*, \pi) & \text{if} \quad W^2 > \max\{W^1, W^3\} \\ W^C(S_2^*, 1) & \text{if} \quad W^3 \geq \max\{W^1, W^2\}. \end{cases} \tag{A26}$$

If $W^E(E^*) < W_C^F$, there are no empires, diplomacy prevails and the size of the free world is $F = 1$. If instead $W^E(E^*) \geq W_C^F$, there are $\pi/\mu E^*$ empires of size E^*, and the size of the free world is reduced to $F = 1 - \pi/\mu$.

Core localities prefer empires to peaceful countries ($W^E(E^*) > W^C(S_{1C}^*, S_{1C}^*)$) if $\gamma > \bar{\gamma}_E$, for a threshold $\bar{\gamma}_E$ implicitly defined by:

$$2\sqrt{\phi}\left[\sqrt{\delta - \beta(\bar{\gamma}_E + \rho)} - \sqrt{\delta\mu - \beta(\bar{\gamma}_E + \rho\mu)}\right] = \omega - (1-\beta)\ln\left(1 + \lambda\frac{1-\mu}{\mu}\right), \tag{A27}$$

such that (recalling the definition of μ) $\partial\bar{\gamma}_E/\partial\beta < 0$, $\partial\bar{\gamma}_E/\partial\phi < 0$, $\partial\bar{\gamma}_E/\partial\pi \geq 0$, $\partial\bar{\gamma}_E/\partial\theta \leq 0$, $\partial\bar{\gamma}_E/\partial\omega > 0$ and $\partial\bar{\gamma}_E/\partial\lambda < 0$.

Core localities prefer empires to the peaceful core union ($W^E(E^*) > W^C(S_2^*, \pi)$) if:

$$\kappa > \bar{\kappa}_{E\pi}(\gamma) \equiv \omega - (1-\beta)\ln\left(1+\lambda\frac{1-\mu}{\mu}\right) + \beta(\gamma+\rho)\pi - 2\sqrt{\phi}\left[\sqrt{\delta} - \sqrt{\delta\mu - \beta(\gamma+\rho\mu)}\right], \tag{A28}$$

for a concave threshold ($\partial^2\bar{\kappa}_{E\pi}/\partial\gamma^2 < 0$) with:

$$\bar{\kappa}_{E\pi}(\bar{\gamma}_E) = \bar{\kappa}_1(\bar{\gamma}_E) \text{ and } \frac{\partial\bar{\kappa}_{E\pi}}{\partial\gamma} < \frac{\partial\bar{\kappa}_1}{\partial\gamma}, \tag{A29}$$

such that (recalling the definition of μ) $\partial\bar{\kappa}_{E\pi}/\partial\phi < 0$, $\partial\bar{\kappa}_{E\pi}/\partial\pi > 0$, $\partial\bar{\kappa}_{E\pi}/\partial\theta \leq 0$, $\partial\bar{\kappa}_{E\pi}/\partial\omega = 1$, $\partial\bar{\kappa}_{E\pi}/\partial\lambda < 0$ and $\partial\bar{\kappa}_{E\pi}/\partial\rho > 0$.

Core localities prefer empires to the peaceful world union ($W^E(E^*) > W^C(S_2^*, 1)$) if:

$$\kappa > \bar{\kappa}_{E1}(\gamma) \equiv \omega - (1-\beta)\ln\left(1+\lambda\frac{1-\mu}{\mu}\right) + \beta(\gamma+\rho\pi) - 2\sqrt{\phi}\left[\sqrt{\delta} - \sqrt{\delta\mu - \beta(\gamma+\rho\mu)}\right], \tag{A30}$$

for an increasing and concave threshold ($\partial\bar{\kappa}_{E1}/\partial\gamma > 0 > \partial^2\bar{\kappa}_{E1}/\partial\gamma^2$) with:

$$\bar{\kappa}_{E1}(\gamma) > \bar{\kappa}_{E\pi}(\gamma) \text{ and } \frac{\partial\bar{\kappa}_{E\pi}}{\partial\gamma} < \frac{\partial\bar{\kappa}_{E1}}{\partial\gamma} < \pi\frac{\partial\bar{\kappa}_1}{\partial\gamma} + (1-\pi)\frac{\partial\bar{\kappa}_2}{\partial\gamma}, \tag{A31}$$

such that (recalling the definition of μ) $\partial\bar{\kappa}_{E1}/\partial\phi < 0$, $\partial\bar{\kappa}_{E1}/\partial\pi > 0$, $\partial\bar{\kappa}_{E1}/\partial\theta \leq 0$, $\partial\bar{\kappa}_{E1}/\partial\omega = 1$, $\partial\bar{\kappa}_{E1}/\partial\lambda < 0$ and $\partial\bar{\kappa}_{E1}/\partial\rho > 0$.

A.4 Extensions

A.4.1 Looting in All Industries

In Section 4 we assumed for simplicity that imperial powers loot from their colonies a fraction $\lambda \in [0,1]$ of market goods in the industries that do not rely on contract enforcement (a fraction $1-\beta$ of the total). Analogous results obtain, at the cost of a greater algebraic burden, if we assume looting can occur in all industries.

In contract-intensive industries, the metropolis then loots not only traded output, but also output that would otherwise be non-traded. Such looting inefficiently incurs at the same time both the efficiency costs of autarky and the technological costs of trade. As a consequence, In a share $\beta(1-E)$ of industries, consumption in the metropolis is $c_l(m,i) = e^{-\eta}\left[1+\lambda e^{-\tau}(E-M)/M\right]$.

As a result, the welfare of a core locality that builds an empire is:

$$W_l = W^E(E) = -\eta + \beta(1-E)\ln\left(1+\lambda\frac{1-\mu}{\mu}e^{\gamma-\eta}\right) + (1-\beta+\beta E)\left[\gamma+\ln\left(1+\lambda\frac{1-\mu}{\mu}\right)\right] - \delta\mu E - \frac{\phi}{E} - \omega. \quad \text{(A32)}$$

The optimal size of an empire, from the perspective of the metropolis, is then:

$$E^* = \sqrt{\frac{\phi}{\delta\mu - \beta\left[\gamma+\ln\left(1+\lambda\frac{1-\mu}{\mu}\right) - \ln\left(1+\lambda\frac{1-\mu}{\mu}e^{\gamma-\eta}\right)\right]}}. \quad \text{(A33)}$$

The difference relative to our baseline is that the equilibrium size of empires is increasing in their ability to loot. If an empire is larger, fewer goods need to be produced with autarky technology. Thus, more looted goods would have incurred transport costs regardless of looting. These economies of scale in looting result from the same transportation technology that enables global trade. Thus they strengthen our baseline findings, making the size of empires increase even more sharply in response to globalization.

A.4.2 Imperial Unions

In Section 4 we ruled out by a simplifying assumption the creation of extractive empires within trade-promoting international unions. This alternative political structure would be desirable if and only if the extractive benefits of colonialism were sufficiently high.

To consider such an extension, suppose that core localities can impose their preferred political structure, including economic unions, onto the rest of the world. If they choose to conquer empires but also create a union, they obtain welfare:

$$W_l = W_2^E(E,U) = -\eta + (1-\beta)\left[\gamma+\ln\left(1+\lambda\frac{1-\mu}{\mu}\right)\right] + \beta\gamma U - \delta\mu E - \frac{\phi}{E} - \omega - \kappa, \quad \text{(A34)}$$

which is maximized by a world union and empires of size:

$$E_2^* = \sqrt{\frac{\phi}{\delta\mu}}. \quad \text{(A35)}$$

Intuitively, empires are smaller with a world union than without it ($E_2^* < E^*$) because their size no longer yields gains from market access. Yet they are larger than peaceful countries

within the union ($E_2^* > S_2^*$) because their size still yields extractive gains.

There are then two possibilities. If Equations (23)–(25) are satisfied, aggressive imperialism is an intermediate stage. Colonialism emerges only when the opportunities for colonial trade are large enough ($\gamma_L > 0$), and it endogenously disappears when economic unions are created. If instead Equations (24) and (25) fail to hold, core localities always resort to war and conquest. They never form peaceful countries ($\gamma_L = 0$), nor do they cease waging war after creating unions. Results are otherwise analogous to our baseline scenario. An economic union is created when trade opportunities cross a threshold value (γ_U^E).[34] Imperialism delays the creation of international unions ($\gamma_U^E > \gamma_U$), especially when democratic values are less widespread ($\partial\gamma_U^E/\partial\theta \geq 0$) and great powers are less likely to clash with each other in their colonial conquests ($\partial\gamma_U^E/\partial\pi \leq 0$).

A.4.3 Shallow Unions

In Sections 3 and 4 we assumed that an international union provides all market regulation. In reality, unions often differ in their depth. They may range from a roundtable for negotiating tariffs to a complex set of treaties regulating disparate issues such as product standards and intellectual property rights.

To capture such differences in the depth of international integration, we can extend our model to include the possibility of forming a "shallow union:" that is, a union that enables trade in a fraction $(1-\alpha)$ of the β industries subject to the border effects. A shallow union may be easier to implement, so that its cost is reduced to $\kappa_\alpha < \kappa$. Then, the utility of a locality in a shallow union is:

$$W^F\left(S_{2\alpha}^*, 1\right) = -\eta + \gamma\left(1 - \alpha\beta + \alpha\beta S_{2\alpha}^*\right) - \delta S_{2\alpha}^* - \frac{\phi}{S_{2\alpha}^*} - \kappa_\alpha, \tag{A36}$$

where optimal country size becomes:

$$S_{2\alpha}^* = \sqrt{\frac{\phi}{\delta - \alpha\beta\gamma}}. \tag{A37}$$

A shallow union leaves the border effect in $\alpha\beta$ industries. As a result, countries in a shallow union are larger than countries in a deep union, but smaller than in the case of no union.

This simple extension yields a rich set of possible outcomes. Expanding trade opportu-

[34] The threshold is formally defined by: $\beta\gamma_U^E + 2\sqrt{\phi\left(\delta\mu - \beta\gamma_U^E\right)} = \kappa + 2\sqrt{\phi\delta\mu}$.

nities make deep unions more attractive than shallow unions. However, shallow unions are cheaper and may therefore be chosen first. Hence, a deep union may follow a shallow union, suggesting that the creation of supranational institutions may be gradual (as in Section 5, but along a different margin). Expanding trade opportunities still make deep unions more attractive than empires. But whether they make shallow unions more attractive than empires depends on which political structure generates a higher volume of trade. By completely removing internal borders, empires promote a high volume of trade between their localities. A shallow union, instead, promotes trade between all world localities, but removes the border effect only partially. To compare them, we compute how an expansion of trade opportunities affects welfare under the two alternative political structures:

$$\frac{\partial}{\partial \gamma}\left[W^{F}\left(S_{2\alpha}^{*},1\right)-W^{E}\left(E\right)\right]=\beta\left[1-E^{*}-\alpha\left(1-S_{2\alpha}^{*}\right)\right]. \tag{A38}$$

The expression in (A38) shows that expanding trade opportunities make a shallow union more attractive than empires when the equilibrium size of empires is sufficiently small: $E^{*}<1-\alpha+\alpha S_{2\alpha}^{*}$. However, the equilibrium size of empires grows with expanding trade opportunities. This makes a non-monotonic evolution possible. At low levels of γ, expanding trade opportunities may trigger a switch from empires to shallow unions. But shallow unions may then be followed by another era of even larger empires for high levels of γ. This possibility may help explain why shallow forms of international cooperation appeared before World War II, such the extensive European network of bilateral trade agreements built around the 1860 French-British treaty (Irwin 1993) or the League of Nations. It may also help explain the decline in international cooperation during the interwar period, and to some extent the backlash against globalization and the revived nationalism observed in recent times.

A.4.4 Federalism

In Section 3 we did not consider the possibility of creating a sub-national level of government in addition to, or instead of, a supranational one. Sub-national federalism is a classic strategy to improve preference-matching in the provision of public services, at the cost of sacrificing economies of scale and scope (Oates 1972).

To consider this possibility, we can extend our model and assume it is possible to create a federal structure that subdivides a country into regional jurisdictions that provide some public services. The cost of creating this new level of government is to forego economies of scope equal to κ_F. The benefit is that the regional governments can provide a share $(1+\psi)/2$

of public services while incurring only a share $(1-\psi)/2$ of fixed costs, for an index $\psi \in [0,1]$ that defines the benefits of federalism. Each region in a country provides a bundle of public services that contains equal amounts of the ideal varieties of the localities that form the region, but not of the localities that form other regions in the same country. Under this assumption, we can re-write Equation (11), describing the utility of a representative locality l, as follows:

$$W_l = -\eta + \gamma\left(1-\beta+\beta U\right) - \delta\left(\frac{1-\psi}{2}S + \frac{1+\psi}{2}F\right) - \phi\left(\frac{1+\psi}{2S} + \frac{1-\psi}{2F}\right) - \kappa_F I_l^F - \kappa_U I_l^U, \tag{A39}$$

where F if the size of regions and I_l^F is an indicator variable that takes value 1 if the country of locality l has a federal structure ($F < S$), and zero otherwise ($F = S$).

In Section 3 there were two possible political structures: (i) centralized countries without an international union ($I_l^U = I_l^F = 0$); and (ii) centralized countries within an international union ($I_l^U = 1$, $I_l^F = 0$). The country sizes and welfares that these options deliver are still those described in Section 3. The present extension adds two additional possibilities: (iii) federal countries without an international union ($I_l^U = 0$, $I_l^F = 1$), which implies the following sizes for countries and their regions:

$$S_f^* = \sqrt{\frac{\phi\left(1+\psi\right)}{\delta\left(1-\psi\right)-2\beta\gamma}} > F^* = \sqrt{\frac{\phi\left(1-\psi\right)}{\delta\left(1+\psi\right)}}, \tag{A40}$$

and welfare:

$$W_l\left(F^*, S_f^*, S_f^*\right) = -\eta + \gamma\left(1-\beta\right) - \sqrt{\phi\left(1+\psi\right)\left[\delta\left(1-\psi\right)-2\beta\gamma\right]} - \sqrt{\delta\phi\left(1-\psi^2\right)} - \kappa_F; \tag{A41}$$

and (iv) federal countries with an international union ($I_l^U = I_l^F = 1$), which implies the following sizes for countries and their regions:

$$S_3^* = \sqrt{\frac{\phi\left(1+\psi\right)}{\delta\left(1-\psi\right)}} > F^* = \sqrt{\frac{\phi\left(1-\psi\right)}{\delta\left(1+\psi\right)}}, \tag{A42}$$

and welfare

$$W_l\left(F^*, S_3^*, 1\right) = -\eta + \gamma - 2\sqrt{\delta\phi\left(1-\psi^2\right)} - \kappa_F - \kappa_U. \tag{A43}$$

The creation of a federal structure increases country size, both in the absence of an international union ($S_f^* > S_1^*$) and in its presence ($S_3^* \geq S_2^*$). The creation of an international

union reduces the size of federal countries just as it does the size of unitary countries ($S_3^* < S_f^*$). However, it does not change the size of federal regions (F^*), since their optimal design responds purely to considerations related to public-service provision, and not to market regulation.

The benefit of creating a federal structure is a reduction in the cost of preference mismatch (δ). Its cost is a sacrifice of economies of scale (ϕ) and scope (κ_F). Overall, a federal structure is desirable for a country that does not belong to an international union if:

$$\sqrt{(1+\psi)\left[\delta\left(1-\psi\right)-2\beta\gamma\right]}+\sqrt{\delta\left(1-\psi^2\right)}+\frac{\kappa_F}{\sqrt{\phi}}\leq 2\sqrt{\delta-\beta\gamma}. \tag{A44}$$

For a country that belongs to an international union, a federal structure is desirable if:

$$\sqrt{1-\psi^2}+\frac{\kappa_F}{2\sqrt{\delta\phi}}\leq 1. \tag{A45}$$

These conditions show that, if ψ is large enough and κ_F small enough, adopting a federal structure can be an equilibrium outcome both with and without international unions.

This finding does not change the main results of our model. As trade opportunities expand, the increase in welfare is: $\partial W_l/\partial\gamma = \beta U$. Thus, it is smallest for a unitary country outside a union ($\partial W_l/\partial\gamma = \beta S_1^*$); it is intermediate for a federal country outside a union ($\partial W_l/\partial\gamma = \beta S_f^*$); it is greatest for countries inside a union, irrespective of their sub-national structure ($\partial W_l/\partial\gamma = \beta$). At the supranational level, the evolution of political structure remains as explained in Section 3. The difference is that the creation of international unions may be delayed if countries adopt a federal structure, which makes them larger. At the sub-national level, a federal structure may be an intermediate stage, becoming obsolete when an international union is created, enabling a decline in country size.

A.5 Historical Narratives

A.5.1 European History

Since the late Middle Ages, European sovereign states on average grew in size until the end of the nineteenth century, when this trend was dramatically reversed. For example, Kitamura and Lagerlöf (2020) show that borders declined monotonically from 1500 to 1900, and then started to increase. Medieval Europe was fragmented into hundreds of small states at a time when trade was costly, insecure and limited to few commodities. The early modern period saw important changes in both the economic and the political organization of the continent.

With the Commercial Revolution, trade began to flourish and the feudal system started to be replaced by a smaller number of countries of growing size. While in 1600 there were 112 sovereign states in Europe and the Near East, at the beginning of 1800 the number had fallen to 79.

The Industrial Revolution gave trade an even more prominent role and triggered major changes in socioeconomic conditions that ultimately made the rise of the nation state possible. Trade expansion was enabled by the introduction of canals, improved roads and railways. At the same time, the high degree of political fragmentation at the time of the Congress of Vienna (1815) was followed by the unification of Germany and Italy (1871) and the further consolidation of other nation states. The economic rationale of building large internal markets was especially evident in the case of German unification, which started with the formation of a customs union (*Zollverein*). This process of political centralization culminated at the beginning of the twentieth century, when Europe was divided into just 28 independent states.

Yet, the twentieth century marks a turning point. Europe entered a stage of political fragmentation, with the number of independent states growing to 58 in 2000.[35] Simultaneously, the process of creating the European Union (EU) started. Our model in Section 5 shows that international unions start from countries that have closer economic ties and expands outward as globalization increases the value of trade with more remote locations too. This prediction is consistent with the history of the EU. In 1957, Belgium, France, Italy, Luxembourg, the Netherlands and West Germany signed the Treaty of Rome establishing the European Economic Community (EEC) as a customs union. The EEC gradually expanded to include nearby countries: Denmark, Ireland and the United Kingdom in 1973, Greece in 1981, Portugal and Spain in 1986. In 1992, the Maastricht Treaty converted the EEC into the EU and, in 1995, Austria, Finland and Sweden joined. In 2002, the Euro was launched. Then the EU grew to include 28 member states in 2013. Over this period, the union has grown in size and scope.

In June 2016, however, the United Kingdom voted in a referendum to leave the union and there is a concern that other countries might follow in the future. While rising nationalism may have played a role, it is instructive to look at these events from the lens of our model. Three lessons can be learned. First, the value of union membership is proportional to the economic ties between countries. These are stronger for countries located in the core of

[35] The number of independent states in Europe and Near East is taken from Euratlas-Nüssli (http://www.euratlas.com).

continental Europe. For instance, while EU members account for almost 80% of Belgium's exports, the same figure is around 50% for the United Kingdom. Second, our model provides a rationale for trade-promoting unions. As more power is shifted to the union in other areas such as migration, tensions may arise, especially in countries with a strong national identity. These two observations may explain why some British politicians have advanced the idea of replacing the EU single market with a Commonwealth free-trade zone. Third, our model suggests that the value of joining the union is proportional to its size. As a country exits, the economic foundations of the union become more fragile.

A.5.2 The Rise and Fall of Colonial Empires

The pattern of an initial decline and subsequent increase in the number of countries is not confined to Europe only. For instance, the number of African countries fell from 36 in 1816 to 4 in 1914, to rise again to 51 in 2000. Similarly, in South-East Asia, these numbers changed from 37 to 4 and then 20 in the years 1816, 1914 and 2000, respectively.[36] However, to better interpret the political evolutions in these regions, it is important to bring conflict into the picture.

In our model of war and conquest, empires are built to extract trade surplus from the colonies and disappear when a world union is formed to foster free markets. According to historians and in line with this view, one of the key driving forces behind colonial expansion was the desire to secure trade and access to scarce resources in an era of revived commerce, but when mercantilist practices where common. For much of the second millennium, states deployed force to create markets (Findlay and O'Rourke, 2007). Due to the scarcity of land and the desire to avoid powerful rivals, European great powers expanded by conquering territories overseas. The role of colonial powers in enforcing trade within the empire but not outside was very clear in the case of maritime commerce. On the one hand, large navies were built to control and protect trading routes; on the other hand, privateers were often authorized to capture merchant ships belonging to enemy nations. Despite some notable setbacks, colonialism continued to grow prior to World War I and finally collapsed after World War II.

The sharp decline of empires started after the creation of international agreements aimed at promoting economic cooperation. This is no coincidence. In the words of Rosecrance (1986) and Spruyt (2005), empires dissolved because the gains through commerce displaced

[36] The number of countries is taken from Butcher and Griffiths (2013).

gains through territorial acquisition.[37] There is also evidence that international organizations played a direct role in the process of decolonization. For example, in 1960 the UN General Assembly voted the Declaration on the Granting of Independence to Colonial Countries and Peoples. Our model of regional unions is consistent with the very different patterns of geographic expansions of empires and unions. As long as war is the dominant means of territorial expansion, the European Great Powers try to avoid conflict with each other by conquering far-away colonies and building global empires. Yet, once trade is enforced by peaceful international unions, countries seek economic integration with proximate partners, with whom the gains from trade are higher. The switch from global empires to regional unions is also consistent with the increased regionalization of world trade patterns observed in the data (e.g., Fouquin and Hugot 2016).

References

[1] Anderson, James E. and Eric van Wincoop (2004). "Trade Costs," *Journal of Economic Literature* 42(3): 691–751.

[2] Andrews, Rhys, and George Boyne (2009). "Size, Structure and Administrative Overheads: An Empirical Analysis of English Local Authorities," *Urban Studies* 46(4): 739–759.

[3] Berry, Christopher B. (2009). *Imperfect Union: Representation and Taxation in Multilevel Governments*. Cambridge: Cambridge University Press.

[4] Besley, Timothy, and Stephen Coate (2003). "Centralized versus Decentralized Provision of Local Public Goods: A Political Economy Approach," *Journal of Public Economic* 87(12): 2611–2637.

[5] Blom-Hansen, Jens, Kurt Houlberg and Søren Serritzlew (2014). "Size, Democracy, and the Economic Costs of Running the Political System," *American Journal of Political Science* 58(4): 790–803.

[6] Blom-Hansen, Jens, Kurt Houlberg, Søren Serritzlew and Daniel Treisman (2016). "Jurisdiction Size and Local Government Policy Expenditure: Assessing the Effect of Municipal Amalgamation," *American Political Science Review* 110(4): 812–831.

[37] Bonfatti (2012) also attributes the fall of empires to the growing importance of trade between industrial countries relative to trade with colonies.

[7] Blesse, Sebastian, and Thushyanthan Baskaran (2016). "Do Municipal Mergers Reduce Costs? Evidence from a German Federal State," *Regional Science and Urban Economics* 59: 54–74.

[8] Bonfatti, Roberto (2012)."Trade and the Pattern of European Imperialism, 1492–2000," Oxford University Department of Economics Discussion Paper No. 618.

[9] Dahl, Robert A., and Edward R. Tufte (1973). *Size and Democracy*. Stanford, CA: Stanford University Press.

[10] Donaldson, Dave (2018). "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review* 108(4–5): 899–934.

[11] Dutt, Pushan, Ilian Mihov and Timothy Van Zandt (2013). "The Effect of WTO on the Extensive and the Intensive Margins of Trade." *Journal of International Economics* 91(2): 204–219.

[12] Fan, C. Simon, Chen Lin, and Daniel Treisman (2009). "Political Decentralization and Corruption: Evidence from Around the World," *Journal of Public Economics* 93(1–2): 14–34.

[13] Findlay, Ronald and Kevin H. O'Rourke (2007). *Power and Plenty: Trade, War and the World Economy in the Second Millennium*. Princeton, NJ: Princeton University Press.

[14] Fouquin, Michel and Jules Hugot (2016). "Back to the Future: Trade Costs and the Two Globalizations," CEPII Working Paper No. 2016-13.

[15] Harstad, Bård (2007). "Harmonization and Side Payments in Political Cooperation," *American Economic Review* 97(3): 871–889.

[16] Head, Keith and Thierry Mayer (2014). "Gravity Equations: Workhorse,Toolkit, and Cookbook," in *Handbook of International Economics*, Vol. 4, edited by Gita Gopinath, Elhanan Helpman and Kenneth Rogoff, 131–195. Amsterdam: North Holland.

[17] Helpman, Elhanan, Marc Melitz and Yona Rubinstein (2008). "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 123(2): 441–487.

[18] Hummels, David (2007). "Transportation Costs and International Trade in the Second Era of Globalization", *Journal of Economic Perspectives* 21(3): 131–154.

[19] Irwin, Douglas A. (1993). "Multilateral and bilateral trade policies in the world trading system: an historical perspective," in *New Dimensions in Regional Integration*, edited by Jaime De Melo and Arvind Panagariya, 90-119. Cambridge: Cambridge University Press.

[20] Kitamura, Shuhei, and Nils-Petter Lagerlöf (2020). "Geography and State Fragmentation," *Journal of the European Economic Association* 18(4): 1726–1769.

[21] Lassen, David D., and Søren Serritzlew (2011). "Jurisdiction Size and Local Democracy: Evidence on Internal Political Efficacy from Large-Scale Municipal Reform," *American Political Science Review* 105(2): 238–258.

[22] Le Galès, Patrick, and Peter John (1997). "Is the Grass Greener on the Other Side? What Went Wrong with French Regions, and the Implications for England," *Policy & Politics* 25(1): 51–60.

[23] Levinson, Marc (2006). *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton, NJ: Princeton University Press.

[24] Lockwood, Benjamin (2002). "Distributive Politics and the Costs of Centralization," *Review of Economic Studies* 69(2): 313–337.

[25] McCallum, John (1995). "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review* 85(3): 615–623

[26] Musgrave, Richard A. (1971). "Economics of Fiscal Federalism," *Nebraska Journal of Economics and Business* 10(4): 3–13.

[27] Pascali, Luigi (2017) "The Wind of Change: Maritime Technology, Trade and Economic Development," *American Economic Review* 107(9): 2821–2854.

[28] Reingewertz, Yaniv (2012). "Do Municipal Amalgamations Work? Evidence from Municipalities in Israel," *Journal of Urban Economics* 72(2–3): 240–251.

[29] Rosecrance, Richard (1986). *The Rise of the Trading State: Commerce and Conquest in the Modern World*. New York, NY: Basic Books.

[30] Spruyt, Hendrik (2005). *Ending Empire: Contested Sovereignty and Territorial Partition*. Ithaca, NY: Cornell University Press.

[31] Steinwender, Claudia (2018). "Real Effects of Information Frictions: When the States and the Kingdom became United," *American Economic Review* 108(3): 657–696.

[32] Strumpf, Koleman S., and Felix Oberholzer-Gee (2002). "Endogenous Policy Decentralization: Testing the Central Tenet of Economic Federalism," *Journal of Political Economy* 110(1): 1–36.