Navarra Center for International Development

Working Paper nº 07/2016

The Geography of Linguistic Diversity and the Provision of Public Goods

Joseph Gomes

Navarra Center for International Development, University of Navarra

Klaus Desmet

Southern Methodist University

Ignacio Ortuño-Ortín

University Carlos III

Navarra Center for International Development
WP-07/2016

The Geography of Linguistic Diversity and the Provision of Public Goods*

Klaus Desmet
SMU

Joseph Gomes
Universidad de Navarra

Ignacio Ortuño-Ortín
UC3M

November 2016

Abstract

This paper theoretically analyzes and empirically investigates the importance of local interaction between individuals of different linguistic groups for the provision of public goods at the national level. Depending on whether local interaction mitigates or reinforces antagonism towards other groups, the micro-founded theory we develop predicts that a country's provision of public goods (i) decreases in its overall linguistic fractionalization, and (ii) either increases or decreases in how much individuals locally learn about other groups. After constructing a 5 km by 5 km geographic dataset on language use for 223 countries, we compute measures of overall fractionalization and local learning, and investigate their relation to public good provision at the country level. While overall fractionalization worsens outcomes, we find a positive causal relation between local learning and public goods. Local mixing therefore mitigates the negative impact of a country's overall linguistic fractionalization. An IV strategy shows that this result is not driven by the possible endogenous spatial distribution of language speakers within countries.

1 Introduction

Societies with a high degree of ethnolinguistic diversity often suffer from the underprovision of public goods (see, e.g., Alesina and La Ferrara, 2005). One reason may be that feelings of animosity make people value public goods less if they need to share them with others who are different. If so, more diverse societies would tend to underinvest in public goods and experience worse outcomes in health, schooling and infrastructure. The underlying assumption is that more diversity breeds more animosity.

The argument that diversity brings about animosity is consistent with *conflict theory*, which posits that more interaction with individuals of other groups is costly and generates greater antagonism. But not everyone agrees that more interaction should necessarily lead to more tension. In fact, *contact theory*, first developed by the psychologist Allport (1954), posits the opposite: repeated interpersonal contact between people tends to reduce prejudice and conflict between individuals of different groups. Important in these theories is that the change in prejudice extends to the entire out-group: if a Dutch-speaking Belgian has more contact with French-speaking Belgians, it affects his prejudice towards all French-speaking Belgians.

*Desmet: Department of Economics, Southern Methodist University, Dallas, TX, kdesmet@smu.edu, Gomes: Navarra Center for International Development, Universidad de Navarra, Pamplona (Navarra), Spain jgomes@unav.es, Ortuño-Ortín: Department of Economics, Universidad Carlos III, Getafe (Madrid), Spain, iortuno@eco.uc3m.es. Gomes gratefully acknowledges financial support from the Fundación Ramón Areces, and Ortuño-Ortín thanks the Spanish Ministry of Economics and Competitiveness (grant ECO2013-42710-P). We benefitted from discussions with Diego Puga and presentations at the University of Heidelberg, the 2016 10th CESIfo Workshop on Political Economy and the 2014 British Academy Workshop on "Interdisciplinary Perspectives on Peace Operations and Conflict Resolution".

Empirical evidence on the subject is divided. In a meta-analysis, Pettigrew and Tropp (2006) conclude that contact theory comes out on top. In contrast, Putnam's (2007) reading of the literature is that the balance is tilted in favor of conflict theory.

One reason for the lack of consensus on whether contact with individuals of other groups breeds or mitigates animosity is that most studies on the political economy of diversity do not directly speak to the question of contact. For example, the empirical cross-country literature that finds a negative association between ethnolinguistic diversity and public goods does not take into account the degree of contact between individuals of different groups. Put differently, a country's overall degree of diversity gives too noisy a signal of the effective contact between individuals of different groups. In Belgium, for example, many Dutch speakers rarely interact with French speakers in their daily lives. In contrast, in Mauritius, an individual may easily use English, Kreol Morisien, French and Bhojpuri, all on the same day.

Our premise is that how much potential there is for effective interaction between people of different groups depends on the spatial distribution of diversity. That is, we can use information on the geographic distribution of languages to get a proxy for the potential contact between individuals of different groups. The degree of potential contact will then either mitigate or reinforce the antagonism that individuals experience towards other groups in the society at large. In this paper we set out to theoretically analyze and empirically investigate the importance of the spatial distribution of diversity for the provision of public goods.

We start by proposing a theoretical model that explores how local interaction affects antagonism towards others. Society is geographically partitioned into many cells. Each individual speaks one language, and each cell has residents of one or multiple languages. Whenever an individual of one linguistic group meets someone from another linguistic group from outside his cell, the tension he feels will be either mitigated or reinforced by how many people of that other group are in his own cell. This captures the idea that frequent contact with others in his own cell may either weaken prejudice, as in contact theory, or worsen prejudice, as in conflict theory. Our micro-founded antagonism index will then be a linear combination of two terms: the first is society's overall ethnolinguistic fractionalization; the second is the probability that when an individual is randomly matched with someone from his own cell and someone from outside his cell, both are from the same group but different from his own group. The second term can be interpreted as the effect of local interaction with other groups on the antagonism towards those other groups in the society at large. This local interaction leads to learning which can improve or worsen one's view about others, and thus either mitigate or reinforce total antagonism. We refer to the first term as *global fractionalization* and to the second term as *local learning*.

A feature of our antagonism index is that for a society with a given level of global fractionalization, local learning is maximized when there is perfect geographic mixing of the linguistic groups. That is, if each cell is linguistically identical to the society as a whole, there is maximal learning about others. In the case of local interaction having a mitigating effect, this implies that total antagonism is minimized when there is perfect geographic mixing. Interestingly, under perfect mixing the learning term is equivalent to a polarization index, though the micro-foundations are different from those in the identification-alienation framework of Esteban and Ray (1994). In that paper, what generates polarization, rather than fractionalization, is that identification with the own group reinforces alienation towards others. Using the same identification-alienation idea, in our case we obtain a polarization index because local identification with other groups either mitigates or reinforces the alienation towards others in the society at large. So it is local identification with other groups, rather than with the own group, that in the aggregate yields the same polarization index as the one in Esteban and Ray (1994).

We then relate our theory of antagonism to the provision of public goods by postulating that an individual's valuation of public goods depends on how much antagonism he feels towards others in society. If the level of public goods depends on private contributions, then it is easy to show that the provision of public goods depends on average antagonism in society. The main empirical implication of the theory is that if one wants to understand the impact of diversity on the provision of public goods, one should distinguish between the effect of global diversity and the effect of local learning. Empirical studies that only control for global diversity suffer from omitted variable bias by not distinguishing between societies that have the same overall diversity, but very different levels of local learning.

Before empirically investigating the effect of the spatial distribution of diversity on the provision of public goods, we need detailed geographic information on the number of speakers of different languages. The Ethnologue has information on the number of speakers per country of 6,905 unique languages spoken across the globe. It also provides language maps that show the geographic distribution of languages in each country. Together with information on population at a fine geographic resolution from Landscan, we use an iterative proportional fitting algorithm to assign the number of speakers of different languages to each 5 km by 5 km cell in the world. Once we have information of the speakers of different languages at a 5 km by 5 km resolution, we compute our measures of global fractionalization and local learning for each country. The underlying assumption is that daily interaction occurs within cells of 5 km by 5 km.[1]

We then explore the effect of global fractionalization and local learning on a wide variety of public goods outcomes in health, education and infrastructure at the level of countries. We start with an OLS approach, and find that global fractionalization has a negative association with public good provision, whereas local learning has a positive association with public good provision. Consistent with contact theory, this implies that local learning mitigates the antagonism felt towards other groups. The magnitudes of these effects are considerable. For example, a one standard deviation increase in local learning lowers child mortality by 7.4 per thousand live births. To put this figure into perspective, in its effect on child mortality, a one standard deviation increase in local learning is equivalent to a 61 percent increase in GDP per capita. As another example, a one standard deviation increase in local learning lowers illiteracy by 5.1 percentage points. This corresponds to a standardized β of -29 percent, i.e., a one standard deviation increase in local learning decreases the illiteracy rate by 29 percent of its standard deviation.

Endogeneity is potentially a concern, as in societies with poor public goods provision individuals

[1] Note that we do not distinguish between contact and exposure. For a discussion on the difference between both concepts, see, for example, Finseraas et al. (2016).

from the same linguistic group may prefer to geographically cluster to support each other. In that sense, public goods outcomes may affect the spatial sorting of individuals of different linguistic groups. To address this issue, we develop an instrument that predicts a country's geographic distribution of linguistic groups based on the language use of its closest neighbor. More particularly, for each cell in a given country, we look at the languages spoken in the closest cell in a neighboring country, and use that information to predict language use in the original cell. We then rerun all our regressions using an IV approach. Reassuringly, the results are, if anything, stronger. For example, a one standard deviation increase in local learning lowers child mortality by 17.0 per thousand, compared to 7.4 per thousand in the OLS analysis. This allows us to conclude that there is a causal positive effect of local learning on the quality of public goods. Overall, contact theory trumps conflict theory.

We are not the first to make the point that the spatial distribution of diversity may matter for different political economy outcomes. In an unpublished working paper, Matuszeski and Schneider (2006) show that civil war is more likely in societies where ethnic groups are more clustered. However, they have no theory about why this should be the case, and their algorithm to assign speakers of different languages to geographic cells ignores the existence of widespread languages, thus introducing a bias in many countries of the New World. Also related is the work by Alesina and Zhuravskaya (2011) who show that geographic segregation has a negative effect on the quality of government. Their paper uses administrative regions, which are often quite large, making it an imprecise measure of the diversity people experience in their daily lives. For example, in their paper, Indonesia, a country with an area of 1.8 million square kilometers, consists of 34 provinces, whereas in our work it contains around 72,000 cells. Finally, in a recent study of sixteen African economies, Robinson (2013) shows that local level diversity increases interethnic trust.

Compared to the the existing literature, we are the first to provide a theoretical framework to think about how local interaction affects the antagonism deriving from a country's overall diversity, to analyze the impact on public goods, and to build a detailed 5 km by 5 km database that takes into account both geographically confined and widespread languages. As an additional theoretical contribution, we clarify that to adequately capture the underlying assumptions of contact theory and conflict theory, it is key to focus on the difference between global diversity and local learning, rather than on the difference between global diversity and local diversity, in contrast to Matuszeski and Schneider (2006) and Robinson (2013).

Our results may seem to contradict Alesina, Baqir and Easterly (1999) and other studies which find that the negative association between diversity and public goods provision does not only hold at the country-level but also at the city-level.[2] However, they do not. Our point is different: controlling for a country's total diversity, an increase in local learning improves the support for public goods provided at the country-level. The equivalent exercise in the context of Alesina, Baqir and Easterly (1999) would be to see whether more interaction in the neighborhoods increases public goods provision at the city-level.

How can our result that diversity might be good be reconciled with the standard negative association

[2]For other studies that look at the relation between diversity and different political economy outcomes at the local level, see Munshi and Rosenzweig (2015) and Montalvo and Reynal-Querol (2016).

between a country's diversity and its provision of public goods found in the literature? The answer is simple. On average, highly diverse countries are not sufficiently mixed geographically for the local learning effect to compensate the negative impact of global diversity. Hence, on average more diversity goes together with a worse provision of public goods. In only some countries that are both highly diverse and highly mixed spatially, do we find that diversity may outperform homogeneity. Although this does not overturn the average negative association between diversity and public goods, it shows that the relation between both is much more complex than previously thought.

The rest of the paper is organized as follows. Section 2 develops a model of antagonism and shows how it relates to the provision of public goods. Section 3 discusses the data, and provides details of the algorithm to assign languages and populations to all 5 km by 5 km cells in the world. Section 4 empirically establishes the positive effect of local learning on public goods provision. Section 5 concludes.

2 Model

A country is geographically partitioned into K cells, indexed by ℓ or k. Each cell is small relative to the country. There are N individuals in the country, partitioned into M linguistic groups, indexed by i or j. Denote by $s_{\ell i}$ the share of people of location ℓ who belong to linguistic group i, by s_ℓ the share of individuals who live in cell ℓ, and by s_i the share of individuals who belong to ethnolinguistic group i. Each individual belongs to one linguistic group i and lives in one cell ℓ, so that $\sum_i s_{\ell i} = 1$, $\sum_\ell s_\ell = 1$ and $\sum_i s_i = 1$. An individual who speaks language i and resides in ℓ has preferences over private consumption c and public consumption G of the form

$$U_{\ell i}(c, G) = \ln c + v_{\ell i} \ln G, \tag{1}$$

where the valuation parameter $v_{\ell i}$ depends negatively on the antagonism the individual feels toward society. Public consumption G is common to the entire country, and is hence not cell-specific. Examples may include national education, nation-wide health policies and defense. Individuals only differ from each other by where they reside and which language they speak. In particular, all individuals have the same income y, which we normalize to 1. In what follows we start by proposing a model of antagonism, and then relate it to the valuation of public goods.[3]

2.1 Antagonism

Start with the simple case of each individual feeling antagonism of 1 towards people of other linguistic groups and of 0 towards people of his own group. For now, the fact that individuals may live in different cells is assumed to be inconsequential. The average antagonism experienced by an individual of linguistic group i is then $\sum_{j \neq i} s_j$. Summing up over all individuals of all linguistic groups yields an average antagonism in society of $\sum_i \sum_{j \neq i} s_i s_j$. This is of course nothing else than the well-known ethnolinguistic fractionalization

[3] For alternative models of the relation between ethnic diversity and public goods, see, e.g., Miguel and Gugerty (2005) which focus on local public goods and the role of social sanctions, and Alesina and La Ferrara (2000) who look at the relation between diversity and group formation.

index, which measures the probability that two randomly drawn individuals belong to different groups. Since this measures the overall fractionalization of a country, we will refer to this index as *global fractionalization* and will denote it by ELF_{glob}.

Local interaction and antagonism towards other groups. We now allow for the possibility that frequent interpersonal contact with other groups mitigates or reinforces the antagonism an individual experiences towards individuals of those groups. There are two opposing views in the literature on how interpersonal contact may affect prejudices people hold against others. *Contact theory* argues that knowing people from another group reduces prejudice against that group. *Conflict theory* also says that contact affects prejudice, but in this case the effect goes the other way, so that knowing people from other groups increases prejudice against that group.

Antagonism towards other groups: global fractionalization and local learning

If interpersonal contact affects antagonism, then the spatial distribution of the different groups becomes important for the degree of antagonism in society. We assume that interpersonal contact happens in the cell where an individual resides, but the frequency of this local contact affects the antagonism towards people of other groups in the society at large. To be more precise, the antagonism an individual of group i and cell ℓ feels towards an individual of group j in the society at large is given by

$$1 + \beta s_{\ell j}, \tag{2}$$

where $s_{\ell j}$ measures the frequency of contact with people of group j in his own cell ℓ. A positive value of β is consistent with conflict theory since interpersonal contact increases antagonism and thus reinforces prejudice. In that case an individual of group i and cell ℓ feels greater antagonism towards an individual of group j in society if the share of group j is bigger in his own cell. In contrast, a negative value of β is consistent with contact theory because more local contact with group j mitigates the prejudice of an individual of group i against an individual of group j in the society at large.

Starting from (2), we can now compute the share-weighted antagonism felt by an individual of group i and cell ℓ towards all individuals of group j in society:

$$a_{\ell ij} = s_j(1 + \beta s_{\ell j}) = s_j + \beta s_j s_{\ell j}, \tag{3}$$

where the first term can be interpreted as the antagonism in the absence of local interaction and the second term can be interpreted as the mitigating or reinforcing effect of local interaction on the antagonism experienced.

The average antagonism of all individuals of cell ℓ towards any other individuals in society is then:

$$a_\ell = \sum_i s_{\ell i} \left(\sum_{j \neq i} s_j (1 + \beta s_{\ell j}) \right).$$

Taking the population-weighted average across all cells yields a measure of the average antagonism in society,

$$A = \sum_{\ell} s_{\ell} \left(\sum_{i} s_{\ell i} \sum_{j \neq i} s_j (1 + \beta s_{\ell j}) \right),$$

which can be re-written as

$$\begin{aligned} A &= \sum_{i} \sum_{j \neq i} s_i s_j + \beta \sum_{\ell} s_{\ell} \sum_{i} \sum_{j \neq i} s_{\ell i} s_{\ell j} s_j \\ &= ELF_{glob} + \beta \sum_{\ell} s_{\ell} \sum_{i} \sum_{j \neq i} s_{\ell i} s_{\ell j} s_j \\ &= ELF_{glob} + \beta LL, \end{aligned} \tag{4}$$

where the first term is the country's overall linguistic fractionalization and the second term is the probability that when an individual is randomly matched with two other individuals, one from his own cell and another from the society at large, both are of the same group but different from his own group. The second term can be thought of as the average effect of local interaction on the antagonism experienced towards others. That is, local learning about others can either increase or decrease the tension towards others in the overall society. We refer to this second term LL as the *local learning* effect.

To further analyze the expression of average antagonism A, compare two societies with the same degree of ELF_{glob}, one with zero local mixing and the other with perfect local mixing. In society 1 all cells are linguistically homogeneous, so that individuals do not locally interact with people from other groups. There is no local learning, so that LL is zero and average antagonism is simply equal to ELF_{glob}:

$$A_1 = ELF_{glob} \tag{5}$$

In society 2, all cells are as diverse as society as a whole, so that $s_{\ell i} = s_i$ for all ℓ and all i. Average antagonism is then

$$A_2 = ELF_{glob} + \beta \sum_{i} \sum_{j \neq i} s_i s_j^2. \tag{6}$$

From comparing these two societies, we can derive three useful conclusions. First, for a given distribution of population across cells, and for a given distribution of languages, the *local learning* effect in (4) is minimized when each cell is linguistically homogeneous and it is maximized when $s_{\ell i} = s_i$ for all i and ℓ. That is, zero mixing, as represented by (5), yields no learning, and perfect mixing, as represented by (6), yields maximal learning. The worst way to learn about other groups in society is to have no other groups locally, and the best way to learn about other groups in society is for the local environment to be a copy of the country. Of course, the effect of learning on antagonism depends on the sign of β. If $\beta < 0$, average antagonism is minimized when there is perfect geographical mixing of linguistic groups, whereas if $\beta > 0$, average antagonism is minimized when there is no mixing. Second, if we impose a lower bound of -1 on β, the average antagonism in a diverse society is always greater than in a completely homogeneous society, even if there is perfect spatial mixing. However, if we allow β to drop below -1, then a diverse society with perfect spatial mixing may have a lower degree of antagonism than a completely homogeneous society. In our theoretical framework we do not impose any restriction on the value of β. Third, the learning effect under perfect mixing, $\sum_i \sum_{j\neq i} s_i s_j^2$, corresponds to an index of polarization. Understanding the link between our index and polarization requires some further discussion.

Relation with polarization

Esteban and Ray (1994) micro-found an index of polarization by arguing that the antagonism an individual experiences towards others increases in the size of his own group. Antagonism is greater if an individual feels more strongly identified with his own group, and this happens when the size of his group is bigger. As a result, instead of the standard fractionalization index, $\sum_i \sum_{j\neq i} s_i s_j$, which ignores the role of identification, they obtain a polarization index, $\sum_i \sum_{j\neq i} s_i^2 s_j$.[4] It is the identification with the own group that explains the quadratic term on the own-group share in their index. Although in our case individuals do not identify with their own group, we can still use the identification concept to interpret our index. Rather than local interaction with another group leading to learning, local interaction creates identification with that other group, which then affects the antagonism felt towards that group in the society at large. So we also get a quadratic term, but in the other-group share, rather than in the own-group share. That is, local learning in cell ℓ about group j increases in $s_{\ell j}$, and the effect of this learning on antagonism with group j in society gets multiplied by its share s_j. Hence, if $s_{\ell j} = s_j$, the effect of local learning on antagonism increases in s_j^2. Of course, since i and j are interchangeable in (6), whether the quadratic term appears in the own group or in the other group is irrelevant. This explains why the second term in (6) is identical to the standard polarization index, in spite of the micro-foundations being different.

Our index also says something useful about the relation between fractionalization and polarization. Take two societies, perfectly mixed at the local level, with the same level of ELF_{glob}, but with the first society having a higher degree of polarization than the second so that LL_{loc} is higher in the first society. For example, language shares might be $(1/3, 1/3, 1/3)$ in the first society and $(0.5, 1/6, 1/6, 1/6)$ in the second society. In that case, the society with the highest degree of polarization, i.e., society $(1/3, 1/3, 1/3)$, experiences the greatest local learning. Hence, if $\beta < 0$, then for a given level of global fractionalization, average antagonism is minimized when polarization is maximized. That is, in a perfectly mixed society, the local learning effect is maximized when polarization is maximized. The intuition for this result is straightforward: learning about fewer, but bigger, groups is more useful than learning about more, but smaller, groups. Although our local learning is not about the languages per se, an example based on languages may help. Learning two words of a language that accounts for 1/3 of the population is more useful than learning one word of each of two languages that make up 1/6 of the population. Although in both cases we have learned two words, it is better to be able to use two words with 1/3 of the population than to be able to use one word with 1/6 of the population and another word with 1/6 of the population.

[4]Esteban and Ray (1994) consider a more general index where identification with the own group i is a function of s_i^α, with $\alpha \in [1, 1.6]$. This yields a polarization index which is a function of $s_i^{1+\alpha}$. In our discussion we take $\alpha = 1$. For further reference, see also the work of Montalvo and Reynal-Querol (2005).

Relation with local fractionalization

Our measure of average antagonism is a linear combination of global fractionalization and average local learning. Other papers have taken the view that what might matter is *global fractionalization* and average *local fractionalization.* For example, Matuszeski and Schneider (2006) analyze the effect of the difference between global and local fractionalization on conflict, and Robinson (2013) studies the relation between global fractionalization and local fractionalization on trust. Although these papers do not provide a model to rationalize their choice of focusing on global and local diversity, it would be easy to do so by slightly changing our model. However, as we will see, those changes would make it incompatible with the underpinnings of contact theory or conflict theory.

A possible way to rationalize the approach taken by those authors is to suppose that the antagonism an individual of group i and cell ℓ feels towards people of group j in the society at large is given by

$$a_{\ell ij} = s_j + \beta s_{\ell j}, \tag{7}$$

so that the average antagonism felt by individuals of cell ℓ towards any other individuals in society would be

$$a_\ell = \sum_i s_{\ell i} \left(\sum_{j \neq i} s_j + \beta s_{\ell j} \right).$$

Average antagonism in society by individuals of any cell can then be written as

$$\begin{aligned} A &= \left(1 - \sum_i s_i^2\right) + \beta \sum_\ell s_\ell \left(1 - \sum_i s_{\ell i}^2\right) \\ &= ELF_{glob} + \beta \sum_\ell s_\ell ELF_\ell \\ &= ELF_{glob} + \beta ELF_{loc}, \end{aligned} \tag{8}$$

where ELF_{loc} is the population-weighted average of local fractionalization across cells.

In spite of its intuitive interpretation and its appealing simplicity, this index does not appropriately capture the basic insights of contact theory and conflict theory which argue that local interaction may reduce or increase the prejudice individuals feel towards the out-group. The point is that the impact of local interaction should only operate if there is antagonism in society. That is, there can only be a reduction in antagonism if there is antagonism in the first place. The alternative index based on global and local fractionalization does not satisfy this premise. To see this, consider a situation where $s_j \approx 0$ and $s_{\ell j} \approx 1$, and assume $\beta < 0$. Since no one outside cell ℓ is from group j, the antagonism experienced by individuals of group i and cell ℓ towards individuals from group j must be zero.[5] However, expression (7) would imply that the antagonism towards individuals of group j is $-\beta$. That is, local learning reduces antagonism even if there was no antagonism to start with. This happens because rather than local learning affecting the existing antagonism toward the out-group, local learning simply enters additively into the overall expression

[5] Recall that each cell is very small relative to the country, so that $s_{\ell j}$ does not affect average antagonism directly, but only through the *local learning* effect.

of antagonism. In contrast, in our corresponding measure of antagonism (3) local learning enters multiplicatively in the antagonism felt toward the out-group, so that the change in antagonism is always a fraction of the pre-existing antagonism. For example, in the above example, using (3), the tension felt towards the non-existent out-group would be zero.

2.2 Public Goods Provision

As already mentioned, we assume that the valuation $v_{\ell i}$ that an individual from linguistic group i and cell ℓ attaches to the public good G is a negative function of the antagonism he feels. The average antagonism experienced by an individual of group i living in cell ℓ towards the rest of society is

$$a_{\ell i} = \sum_{j \neq i} s_j (1 + \beta s_j s_{\ell j}).$$

Note that since β can be negative, $a_{\ell i}$ need not be positive. Hence, we postulate

$$v_{\ell i} = \frac{\kappa_1}{\kappa_2 + a_{\ell i}}, \tag{9}$$

where $\kappa_1 \geq 1$ and $\kappa_2 \geq 0$ is large enough so that $v_{\ell i} > 0$.

Suppose that G is determined by private contributions.[6] In particular, consider a simultaneous private contribution game where the equilibrium concept is Nash. We denote by $g_{\ell i}$ the contribution of an agent living in ℓ of group i. The total level of the public good is then

$$G = \sum_{\ell} \sum_{i} g_{\ell i}.$$

An agent with valuation $v_{\ell i}$ chooses his contribution $g_{\ell i}$ by solving

$$\begin{aligned} &\max_{g_{\ell i}} \ln(1 - g_{\ell i}) + v_{\ell i} \ln(G_{-\ell i} + g_{\ell i}) \\ &s.t. \quad 1 \geq g_{\ell i} \geq 0 \end{aligned}$$

where $G_{-\ell i}$ is the contribution of the rest of agents.

Proposition 1 *Suppose that in the Nash equilibrium of the contribution game all agents contribute a strictly positive amount. Then we have that G is a decreasing function of total average antogonism A.*

Proof. We use the "replacement function" approach of Cornes and Hartley (2007) to analyze the Nash equilibrium of a private contribution game. Given that agents have Cobb-Douglas utility functions and all of them have the same income, the proof is simple. In our case the replacement function for an agent i from cell ℓ with valuation of the public good $v_{\ell i}$ is

$$\max\{y - \frac{G}{v_{\ell i}}, 0\}$$

where $y = 1$ denotes the individual's income and G denotes the total provision of public good (Karaivanov, 2009, example 3.1.1, page 780). The replacement function $\max\{y - \frac{G}{v_{\ell i}}, 0\}$ gives "the quantity that if

[6] Later we will briefly discuss that a similar result can be derived in a model where G is determined by a democratic vote.

subtracted from the total provision G, the player's best reply response to the remaining quantity would exactly replace the quantity removed" (Cornes and Hartley, 2007, page 205). Assuming all individuals contribute a strictly positive amount, Karaivanov (2009) shows that the equilibrium total contribution G^* solves

$$\sum_{\ell}\sum_{i} N\ s_\ell\ s_{\ell i}\ (1-\frac{G}{v_{\ell i}})=G \tag{10}$$

which can be written as

$$\sum_{\ell}\sum_{i} N\ s_\ell\ s_{\ell i}-\sum_{\ell}\sum_{i} N\ s_\ell\ s_{\ell i}\left(\frac{\kappa_2+a_{\ell i}}{\kappa_1}\right)G=G \tag{11}$$

or

$$\begin{aligned} N-NG\frac{\kappa_2}{\kappa_1}-NG\sum_{\ell}\sum_{i}\ s_\ell\ s_{\ell i}\frac{a_{\ell i}}{\kappa_1} &= G \\ N-\frac{NG}{\kappa_1}(\kappa_2-\sum_{\ell}\ s_\ell a_\ell) &= G. \end{aligned}$$

Recall that average antagonism is

$$A=\sum_{\ell} s_\ell a_\ell$$

so the solution to (11) can be written as

$$G^*=\frac{\kappa_1 N}{\kappa_1+N(\kappa_2+A)} \tag{12}$$

Expression (12) shows that the equilibrium total contribution G^* is a decreasing function of average antagonism A. ■

For our empirical estimation it will be useful to linearly approximate (12). After dividing numerator and denominator by $\kappa_1 N$ and assuming that $1/N \approx 0$, we can write

$$G^*\approx\frac{\kappa_1}{\kappa_2+A} \tag{13}$$

Assuming that κ_2 is sufficiently large and that A is sufficiently small, we can take a first-order Taylor approximation of (13) which yields

$$G^*\approx\frac{\kappa_1}{\kappa_2}-\frac{\kappa_1}{\kappa_2^2}A=\frac{\kappa_1}{\kappa_2}-\frac{\kappa_1}{\kappa_2^2}ELF_{glob}-\frac{\kappa_1}{\kappa_2^2}\beta LL. \tag{14}$$

This will serve as our estimating equation in the empirical part. From (14) we can conclude that the provision of public goods depends negatively on global fractionalization and either positively or negatively on local learning (positively if $\beta < 0$ and negatively if $\beta > 0$). The theory implies that one should distinguish between global fractionalization and local learning when empirically exploring the relation between diversity and public goods provision.[7]

[7]If instead of a private contributions mechanism, a democratic vote decides the public good, the exact same result would hold if the mean agent coincides with the median agent. The result would still hold qualitatively if the median agent and the mean agent are not too different. In this context the mean agent refers to the agent with the mean valuation of the public good.

3 Data

In this section we describe the different data we use in our empirical analysis. After briefly describing the data on public goods, we mainly focus on how we construct a database of language use at the local level for the entire globe.

3.1 Public Goods

When analyzing the effect of diversity on the provision of public goods, we do not focus on a particular public good. Instead, we look at many different measures, related to health, education and infrastructure. In doing so, we build on previous work by La Porta et al. (1999), Alesina et al. (2003) and Desmet et al. (2012). These include child mortality, measles immunization, illiteracy, school attainment, access to clean water and road infrastructure. The exact variables and their sources are given in Appendix A. Our analysis is done at the country level.

3.2 Spatial Distribution of Languages

To compute the local learning index for all countries of the world, we need to know how many people speak each language at the local level. To that end, we start by splitting up the world in grid cells. Since local learning has to do with the degree of personal interaction in people's daily lives, we need grid cells of a size that captures this daily interaction. In our baseline analysis we take a grid with a resolution of 5 km by 5 km to be a reasonable size. In what follows we explain how we allocate the speakers of the world's different languages to the individual grid cells.

We use two main data sources. The first data source is the digitized version of the 16$^{\text{th}}$ edition of the Ethnologue. This gives us a polygon shapefile, where the 6,905 languages are represented as polygons across space. Since in some areas more than one language is spoken, these polygons may overlap. In addition, when certain widely spoken languages in a country cannot be assigned to any particular geographic region, Ethnologue classifies them as widespread languages. This is equivalent to having a polygon that consists of the entire country.[8] A few languages are assigned to specific points, rather than to polygons, and a few others have unknown locations. For the point languages, we create circular polygons around the points, with a radius that is proportional to the number of speakers of those languages.[9] As for the languages with unknown locations, we treat them as widespread. Some areas, such as the sparsely populated Sahara Desert, have no information on languages. In those cases we assign the language of the nearest cell that has information on language. Since we use grid cells of 5 km by 5 km, we rasterize the data using a resolution of $2'30''$ by $2'30''$. In addition to the linguistic polygons, the Ethnologue also provides the number of people that speak the different languages by country.

The second data source is Landscan which provides population at the same resolution of $30''$ by $30''$.

[8]Excluding widespread languages, there is a maximum of seven overlapping polygons.

[9]We ignore point languages that account for less than 0.5% of the population.

Here as well, we rasterize the data using a resolution of $2'30''$ by $2'30''$. To make the language data from the Ethnologue consistent with the population data from Landscan, we normalize the language data for each country so that the sum of a country's different language speakers equals the country's total population.

These two data sources yield three pieces of information: the number of people per grid cell; the number of speakers of each language per country; and whether a language is spoken or not in a given grid cell. What it does not tell us is how many people speak each language in each cell. Hence, using these three pieces of information, we need to allocate language speakers to grid cells. To that end, we use an iterative proportional fitting algorithm, commonly used in statistics, which we now describe in further detail.[10]

Consider a country that has M linguistic groups and is split up into K cells. Using the two data sources, we construct three matrices that correspond to the three pieces of information we referred to above: $\mathcal{N}$ is a K x 1 matrix of which the elements give the total population of each cell; $\mathcal{L}$ is a 1 x M matrix of which the elements give the number of speakers of each language in the country; $\mathcal{B}$ is a K x M binary matrix of which the elements take a value 1 if the language corresponding to the column is spoken in the cell corresponding to the row (and a value 0 otherwise). The iterative proportional fitting algorithm is a way of allocating language speakers to cells, such that the total population per cell and the total population per language correspond to their actual values. It goes through the following steps.

1. *Step 0.* Define $\mathcal{T}^{(0)} = \mathcal{B}$.

2. *Step 1.* For each location ℓ, assign a share $\mathcal{T}^{(2n-2)}(\ell, i)/\sum_j \mathcal{T}^{(2n-2)}(\ell, j)$ to language i. Hence,

$$\mathcal{T}^{(2n-1)}(\ell, i) = \frac{\mathcal{T}^{(2n-2)}(\ell, i)}{\sum_j \mathcal{T}^{(2n-2)}(\ell, j)} \mathcal{N}(\ell, 1), \tag{15}$$

   where $n = 1, 2, \ldots$ refers to the times the algorithm has iterated through Step 1 and Step 2. To provide some intuition, the first time the algorithm goes through Step 1, the cell's population gets divided equally between the different languages that are spoken there. If, for example, 5 languages are spoken in a cell, then each language gets assigned 20 percent of that cell's population. This allocation always ensures that $\sum_j \mathcal{T}^{(2n-1)}(\ell, j) = \mathcal{N}(\ell, 1)$, i.e., the sum of people allocated to each cell is equal to the actual population of each cell. That is, the allocation satisfies the marginals on the cell populations.

3. *Step 2.* For each language i, assign a share $\mathcal{T}^{(2n-1)}(\ell, i)/\sum_k \mathcal{T}^{(2n-1)}(k, i)$ to cell ℓ. Hence,

$$\mathcal{T}^{(2n)}(\ell, i) = \frac{\mathcal{T}^{(2n-1)}(\ell, i)}{\sum_k \mathcal{T}^{(2n-1)}(k, i)} \mathcal{L}(1, i) \tag{16}$$

   This allocation always ensures that $\sum_k \mathcal{T}^{(2n)}(k, i) = \mathcal{L}(1, i)$, i.e., the sum of population allocated to a language is equal to the actual total number of speakers of that language. That is, the allocation satisfies the marginals on the language populations.

4. *Step 3.* Go through Step 1 and Step 2 until $\mathcal{T}^{(2n-1)}(\ell, i)$ converge to $\mathcal{T}^{(2n)}(\ell, i)$ for all ℓ and i.

[10] See, e.g., Deming and Stephan (1940) and Bishop, Fienberg and Holland (1975).

This iterative proportional fitting algorithm therefore provides us with an allocation of language speakers by cell, $\mathcal{T}^{(2n)}(\ell, i)$. If the three matrices $\mathcal{L}$, $\mathcal{N}$ and $\mathcal{B}$ are fully consistent with each other, then the iterative proportional fitting algorithm is guaranteed to converge. However, there may be small inconsistencies between the data sources. As a simple example, it is possible that the polygon assigned to language i has a population that is smaller than the total population that speaks language i. This inconsistency could in principle be due to three reasons: the local population data from Landscan may contain imprecisions; the country-level language shares from the Ethnologue might have errors; or the language polygons from the Ethnologue may not be a completely accurate reflection of where the different languages are spoken. How we deal with these minor inconsistencies requires taking a stance on the most likely source of error. We take the view that local population and language shares are relatively easy to estimate, whereas language polygons are unlikely to be completely precise. For example, although most Catalan speakers in Spain live in the East of the country, a small percentage of them live in other parts of the country. However, since Catalan is not a "widespread" language in the sense that is not widely spoken across the entire territory, the Ethnologue assigns a specific polygon to Catalan. Given the binary nature of such a geographic polygon, it hence assumes that all Catalan speakers reside within the polygon. Since this is obviously an approximation, we replace the 0 values in the binary matrix $\mathcal{B}$ by 0.000001. This amounts to allowing some speakers of language i to live outside their corresponding language polygon. With this correction, the iterative proportional fitting algorithm is once again guaranteed to converge (Fienberg, 1970).

3.3 Global Diversity and Local Learning Measures

With the spatial distribution of languages at a resolution of 5 km by 5 km in hand, we can now compute, for each country, our measures of global fractionalization and local learning. For comparison purposes, we will also compute a measure of local fractionalization.

When computing measures of linguistic fractionalization, it is not always obvious which linguistic groups should be used as primitives. For example, should Alemannisch and Bavarian, two variations of German, be considered as two separate language groups or should they be aggregated into German? To address this issue, Desmet et al. (2012) use the language tree of the Ethnologue compute measures of linguistic fractionalization at different levels of aggregation. There are 15 possible levels of aggregation, going from the most aggregate at level 1, where only the big language families, such as Indo-European and Niger-Congo are considered to be different groups, to the most disaggregate at level 15, where Alemannisch and Bavarian are taken to be different groups. As Desmet et al. (2012) argue, coarse divisions, obtained at high levels of aggregation, can be thought of as cleavages that go back far in history, whereas finer divisions, obtained at low levels of aggregation, are due to more recent cleavages. They show that certain political economy outcomes, such as conflict, are better explained by measures of linguistic fractionalization at high levels of aggregation, indicating that they have to do with deep cleavages. In contrast, other outcomes, such as economic growth, are better explained by measures of linguistic fractionalization at low levels of aggregation, suggesting that they depend on more shallow cleavages.

Figure 1. Global ELF by Country – Level 15

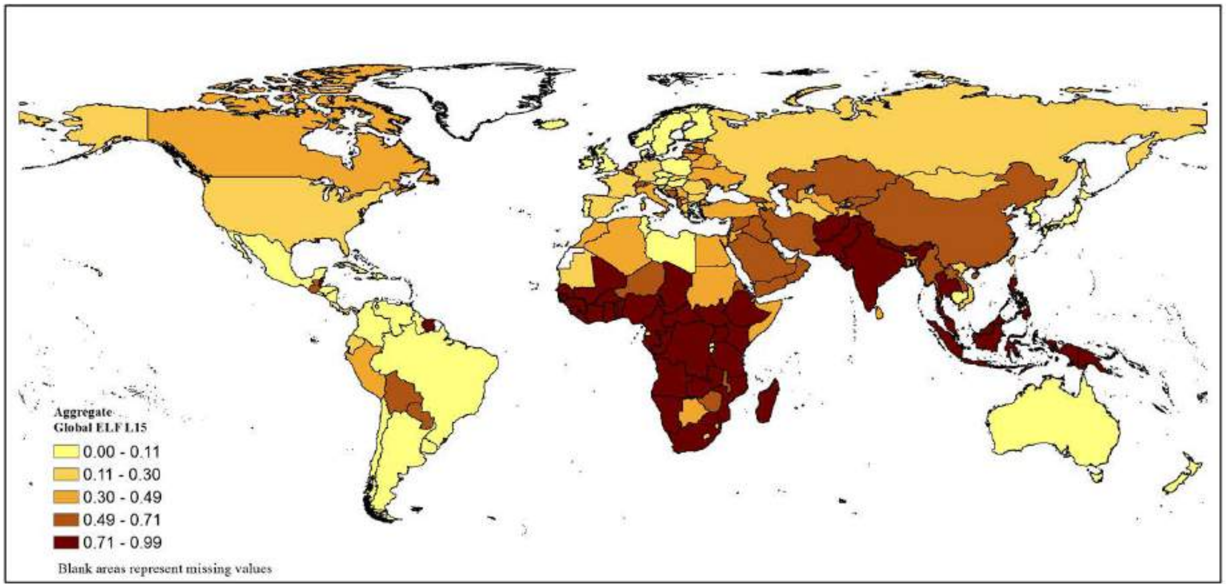


Figure 2. Local Learning by Country – Level 15

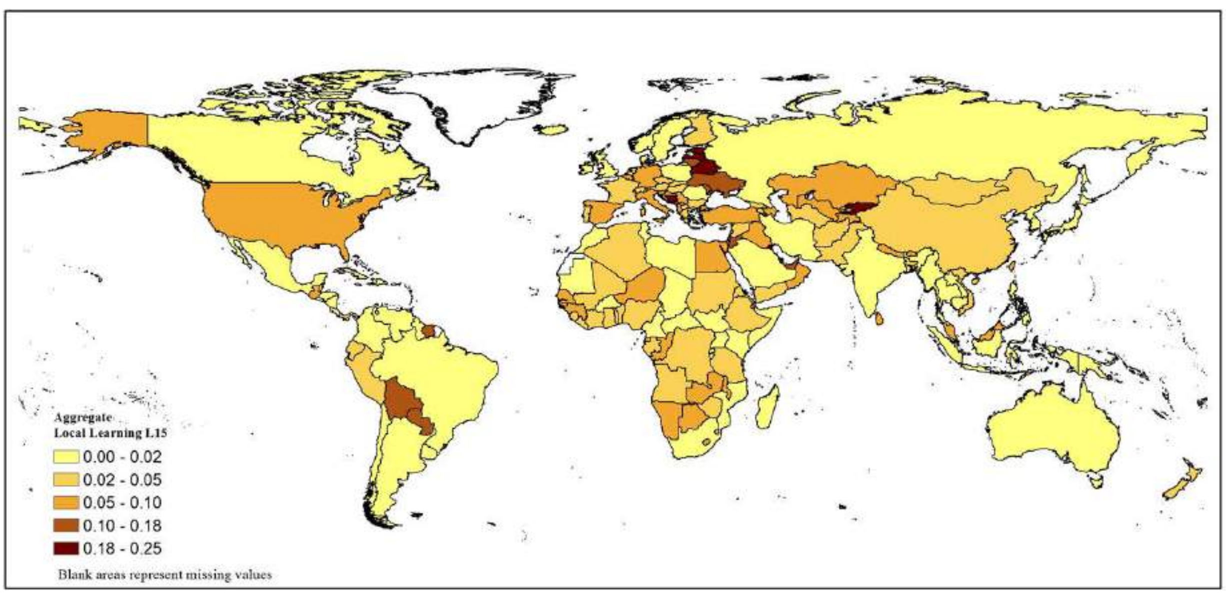
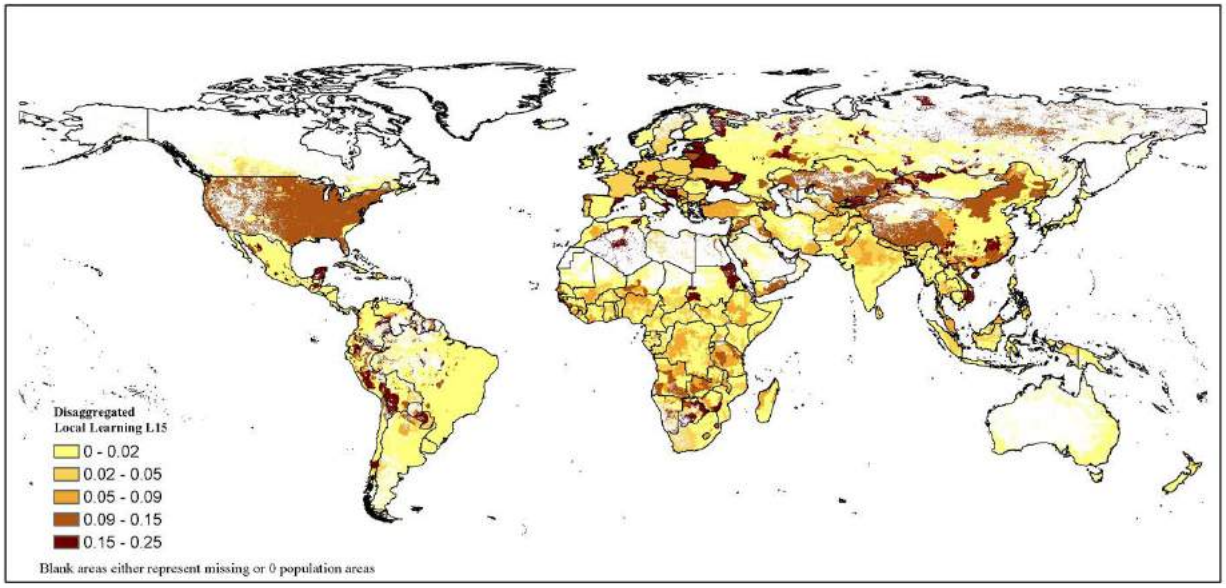


Figure 3. Local Learning at 5 km by 5 km Resolution – Level 15



Using the most disaggregated data, Figure 1 and Figure 2 show the global fractionalization and the local learning index by country for level 15 of disaggregation. As is immediately obvious, there are many differences between both indices. Compare, for example, Botswana and Zimbabwe. Global fractionalization is slightly higher in Zimbabwe, but the degree of local mixing is much higher in Botswana. This example is not an exception: the correlation between both measures is only 0.26. Figure 3 also displays the information on local learning, but at a spatial resolution of 5 km by 5 km. Going back to our previous example, there are more areas of high local learning in Botswana than in Zimbabwe.

Another comparison we highlighted in the introduction is Belgium and Mauritius. In spite of overall diversity being higher in Belgium (0.66) than in Mauritius (0.53), local interaction is lower. The reason is that in Belgium linguistic regions are more or less well defined. Belgians who live in Bruges mainly interact with Dutch-speaking people in their daily lives, whereas Belgians who live in Namur are almost exclusively exposed to French in their day-to-day interaction with others. In contrast, according to Chiba (2006), Mauritians "switch languages according to the occasion in the way other people change clothes". He goes on to argue that "over the course of a day a typical Mauritian might use English to write a school essay, Kreol Morisien to chat with friends, French to read a novel and Bhojpuri to spend a quiet evening with the family." Consistent with this, our index of local learning is much higher in Mauritius (0.20) than in Belgium (0.06).

Appendix Figures B.1 through B.3 display the same maps, but for level 2 of disaggregation. Many of the countries in the Sahel and in central Africa continue to be highly diverse, whereas some of the countries in the southern part of Africa now display lower levels of diversity. For example, Zambia, where nearly everyone speaks a language of the Niger-Congo family, no longer shows up as being highly diverse.

In the case of public goods, Desmet et al. (2012) find that intermediate levels of aggregation are most relevant. Hence, in our empirical analysis of public goods, we aggregate languages to level 5. Appendix Figures B.4 through B.6 show maps at this level of aggregation. To give a sense of what this level of aggregation means, it implies that Spanish, Catalan and Portuguese are aggregated into the same group, but French and Italian are not. Similarly, Hindi and Urdu are considered to be in the same group. As another example, in Tanzania 104 out of the 129 languages are aggregated into the same group (Niger-Congo/Atlantic-Congo/Volta-Congo/Benue-Congo/Bantoid/Southern). For our benchmark analysis we therefore focus on diversity measures computed at that level of aggregation.

To assess the quality of our iterative proportional fitting algorithm that allocates language speakers to grid cells, we compare our local diversity measures to the ones obtained by Gershman and Rivera (2016). Rather than using maps, they rely on national population censuses and regional household surveys to infer the linguistic composition of almost 400 first-level administrative regions in 36 sub-Saharan African countries. They then use the language tree from the Ethnologue to compute local fractionalization measures for these regions at different levels of linguistic aggregation. To compare our measures to the ones in Gershman and Rivera (2016), we start by aggregating our 5 km by 5 km language allocation up to the same first-level administrative regions, and then calculate for each of these regions a measure of local fractionalization. At linguistic aggregation level 5, we find a correlation between our measure of local fractionalization and the one in Gershman and Rivera (2016) of 0.70. This gives us confidence that our spatial allocation of language speakers is reasonable.

4 Empirical Analysis

In this section we test our theory by exploring the impact of global fractionalization and local learning on a country's provision of public goods. We use the following econometric specification:

$$g_c = \beta X_c + \gamma_1 ELF_{glob,c} + \gamma_2 LL_c + \varepsilon_c, \tag{17}$$

where g_c is the level of public goods in country c, $ELF_{glob,c}$ is the global fractionalization measure of country c, LL_c is the local learning measure of country c, X_c are a set of additional controls, and ε_c is an error term. As mentioned in the previous section, in the benchmark we use grid cells of 5 km by 5 km to compute LL_c and a linguistic aggregation level of 5 to compute both $ELF_{glob,c}$ and LL_c. We start by ordinary least square analysis, and then turn to instrumental variable analysis.

4.1 Local Learning and Public Goods (OLS Analysis)

In our baseline analysis we focus on child mortality. It measures the mortality rate under age 5 (per 1,000 live births), and captures well the effectiveness of public goods provision. We then extend our analysis to include a variety of additional measures. In particular, we focus on two more health outcomes (hospital beds per 1000 people, rate of measles immunization), two education outcomes (illiteracy rate, log of average years of schooling) and two infrastructure measures (percentage of households with access to improved sanitation and km of roads per 1,000 people).

Child mortality. Table 1 analyzes the relation between global fractionalization, local learning and child mortality, using OLS. The first two columns follow the standard approach of the literature and only control for a country's global fractionalization. The first specification includes regional dummies and latitude as covariates, whereas the second specification is identical to the one in Desmet, Ortuño-Ortín and Wacziarg (2012) and also controls for legal origin and GDP per capita. Consistent with previous papers, we find that an increase in a country's level of fractionalization is associated with worse outcomes. The coefficients on global fractionalization are statistically significant at the 1% level. The next two columns replace global fractionalization by local learning. The coefficients switch signs, suggesting that local learning is associated with less child mortality, but the coefficients are not statistically significant at the 10% level.

Table 1. Child Mortality: Global ELF and Local Learning

	Child mortality					
	(1)	(2)	(3)	(4)	(5)	(6)
Global ELF	41.130***	25.302***			71.145***	43.568***
	(10.876)	(9.006)			(13.605)	(11.020)
Local Learning			-34.484	-15.777	-216.315***	-125.284***
			(39.988)	(31.693)	(49.881)	(39.920)
Absolute Latitude	-0.707***	-0.143	-0.874***	-0.134	-0.644***	-0.123
	(0.180)	(0.203)	(0.169)	(0.204)	(0.172)	(0.197)
Latin America & Carib.	-11.674*	-5.001	-17.831**	-7.603	-9.614	-3.935
	(6.977)	(5.406)	(6.895)	(5.348)	(6.553)	(5.539)
Sub-Saharan Africa	78.140***	59.661***	78.569***	59.372***	74.665***	59.159***
	(9.349)	(8.322)	(9.751)	(8.581)	(8.700)	(7.915)
East and S.E. Asia	-0.850	-3.414	-1.636	-3.330	-7.029	-7.063
	(9.985)	(8.117)	(10.134)	(8.271)	(9.515)	(8.090)
Log GDP per Capita		-16.660***		-17.925***		-15.512***
		(1.960)		(2.002)		(1.852)
French Legal Origin		-18.165***		-14.038**		-13.596**
		(5.996)		(5.596)		(5.231)
German Legal Origin		-24.016***		-22.807***		-19.678***
		(4.712)		(4.769)		(4.183)
UK Legal Origin		-20.963***		-15.933***		-16.702***
		(5.925)		(5.744)		(5.481)
Constant	46.187***	190.291***	64.352***	204.386***	48.402***	178.007***
	(9.094)	(19.072)	(8.847)	(18.804)	(8.583)	(17.567)
Observations	178	171	178	171	178	171
R^2	0.688	0.815	0.655	0.804	0.720	0.826

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the child mortality rate per 1000 live births. The Global ELF and the Local Learning variables are measured at level 5 of aggregation and are based on the authors' calculations. The variable definitions and data sources for each of the variables are provided in Appendix A.

Motivated by the theory, in the last two columns we include simultaneously global fractionalization and local learning. As before, countries with higher levels of global fractionalization continue to have greater rates of child mortality. In fact, the coefficients are greater in magnitude. Focusing on column (6), we find that a one standard deviation increase in global fractionalization raises child mortality by 11.4 per thousand live births. More interestingly, the local learning coefficients are now statistically significant at the 1% level. Their sign is negative, indicating that local learning reduces child mortality. The economic magnitude of the effect is substantive. Using once again column (6) as our preferred specification, a one standard deviation increase in local learning lowers child mortality by 7.4 per thousand live births. To put this figure into perspective, in its effect on child mortality, a one standard deviation increase in local learning has the same effect on child mortality as a 61% increase in GDP per capita.

These findings are consistent with contact theory. A higher degree of linguistic diversity at the country level implies greater antagonism, so that people value public goods less, but that negative effect is mitigated if people experience diversity in their daily lives. Next, we analyze whether this result generalizes to a wider variety of public goods.

Other public good outcomes. Table 2 reports the results for an additional six public goods outcomes; the specification is identical to the one in column (6) of Table 1. As in the case of child mortality, an increase in a country's overall linguistic fractionalization tends to be associated with worse outcomes, whereas an increase in local learning tends to be associated with better outcomes. The coefficients on global fractionalization are statistically significant at the 5% level in four out of the six outcomes, whereas the coefficients on local learning are always statistically significant at the 5% level, with the exception of road density. In terms of magnitude, a one standard deviation increase in local learning increases the rate of measles immunization by 3.9 percentage points. The corresponding standardized β is 26%, meaning that a one standard deviation increase in local learning increases the measles immunization rate by 26% of its standard deviation. In the case of illiteracy, the standardized β is -29%. These results are further evidence in favor of contact theory.

Decentralization. In our theory the geography consists of two levels: the local grid cells, where individuals interact in their daily lives, and the country, where individuals from the different local cells decide the level of public goods everyone in the country has access to. In reality, in some countries there may be heterogeneity in the access to certain public goods, and the financing of public goods may be decentralized. To see whether this changes our results, we use data on decentralization from Treisman (2008). Table 3, column (1), takes the benchmark specification and controls for whether a country has a federal structure. Being a federal state has no direct effect on the provision of public goods. The rest of the results are largely unchanged. In particular, global fractionalization continues to worsen child mortality, whereas local learning continues to have a benign effect, with both coefficients being statistically significant at the 1% level.

As an alternative measure of decentralization, column (2) focuses on whether subnational governments have autonomy in certain areas and/or have residual powers (i.e., they can legislate in areas that are not explicitly assigned to other levels). As before, this measure of decentralization has no direct effect, and local learning continues to lower child mortality (with a coefficient that is statistically significant at the 5% level). Columns (3) and (4) introduce interaction terms between our two measures of decentralization and

Table 2. Other Public Goods Outcomes: Global ELF and Local Learning

	(1) Measles Immunization	(2) Hospital Beds	(3) Illiteracy Rate	(4) Schooling	(5) Improved Sanitation	(6) Road Density
Global ELF	-25.834***	-1.335**	32.146***	-0.262*	-25.284***	1.049
	(4.593)	(0.625)	(6.061)	(0.140)	(6.337)	(1.775)
Local Learning	66.935***	7.743**	-102.537***	1.383***	82.703***	6.588
	(15.497)	(3.135)	(22.537)	(0.500)	(25.857)	(12.840)
Absolute Latitude	0.201**	0.101***	-0.307***	0.008***	0.140	0.173***
	(0.087)	(0.019)	(0.117)	(0.003)	(0.162)	(0.065)
Latin America & Carib.	4.291	0.308	-6.097*	0.209**	-4.875	2.350
	(2.665)	(0.451)	(3.430)	(0.090)	(4.808)	(1.777)
Sub-Saharan Africa	-9.341***	0.135	7.995*	-0.122	-25.252***	2.989*
	(2.954)	(0.465)	(4.171)	(0.109)	(5.235)	(1.782)
East and S.E. Asia	2.388	1.003	-14.245***	0.085	-6.531	-1.256
	(3.752)	(0.905)	(4.781)	(0.103)	(6.492)	(2.127)
Log GDP per Capita	1.326	0.252	-4.344***	0.116***	10.166***	1.797***
	(0.820)	(0.155)	(1.139)	(0.023)	(1.280)	(0.452)
French Legal Origin	2.416	0.478	3.683	0.082	13.565***	-14.049*
	(2.940)	(1.446)	(3.492)	(0.080)	(4.099)	(7.886)
German Legal Origin	3.312	2.302		0.236***	13.600***	-11.650
	(2.751)	(1.483)		(0.064)	(3.594)	(7.912)
UK Legal Origin	7.766**	0.517	-0.175	0.256***	10.784***	-10.989
	(3.290)	(1.452)	(4.169)	(0.083)	(4.083)	(8.096)
Constant	66.986***	-2.006	54.103***	0.654**	-17.278	-0.602
	(8.729)	(1.953)	(10.967)	(0.257)	(14.230)	(9.531)
Observations	171	173	138	136	171	172
R^2	0.576	0.599	0.665	0.676	0.783	0.459

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The column headings give the dependent variables for each of the columns. The Global ELF and the Local Learning variables are measured at level 5 of aggregation and are based on the authors' calculations. The variable definitions and data sources for each of the variables are provided in Appendix A.

global fractionalization and local learning. This does not change our findings. From this we conclude that our results do not depend on a country's degree of decentralization.

Local fractionalization. As argued in the theory, local learning is different, but related to local fractionalization. Whereas local learning captures the average effect of local interaction on the antagonism experienced towards others, local fractionalization is the population-weighted average of fractionalization at the cell level. The former is micro-founded in contact (or conflict) theory, whereas the latter is not. Column (5) in Table 3 displays what happens when we substitute local learning by local fractionalization, whereas column (6) introduces both measures jointly. Given the high correlation of 0.92 between both measures, the results have to be interpreted with some caution. Not surprisingly, we find that local fractionalization, when introduced by itself, reduces child mortality. However, when we control for both measures, local learning trumps local fractionalization.

Table 3. Child Mortality: Decentralization and Local ELF

	(1) Federal State	(2) Subnational	(3) Federal State	(4) Subnational	(5) Local ELF	(6) Local ELF
Global ELF	42.930***	22.838	43.208***	25.054	49.785***	36.499**
	(12.745)	(15.674)	(14.703)	(18.607)	(13.633)	(14.741)
Local Learning	-145.751***	-119.115**	-152.854***	-136.965**		-183.127*
	(51.096)	(55.775)	(57.304)	(65.442)		(95.270)
Local ELF					-48.771**	30.881
					(19.363)	(47.253)
Decentralization$^{(1)}$	-0.152	0.815	-3.214	3.730		
	(5.283)	(4.985)	(5.570)	(6.727)		
Decentralization$^{(1)}$			-0.171	6.225		
x Global ELF			(28.363)	(30.432)		
Decentralization$^{(1)}$			81.381	130.122		
x Local Learning			(141.141)	(133.635)		
Absolute Latitude	-0.119	-0.184	-0.089	-0.159	-0.145	-0.113
	(0.201)	(0.197)	(0.213)	(0.209)	(0.199)	(0.199)
Latin America & Carib.	-3.750	-5.646	-2.809	-4.030	-4.078	-4.027
	(6.473)	(7.050)	(6.642)	(7.326)	(5.497)	(5.527)
Sub-Saharan Africa	61.260***	57.260***	61.743***	57.243***	59.412***	59.084***
	(9.202)	(9.870)	(9.250)	(9.916)	(8.108)	(7.873)
East and S.E. Asia	-7.302	-6.687	-7.751	-8.034	-5.238	-7.592
	(8.957)	(8.759)	(9.096)	(9.095)	(8.022)	(8.189)
Log GDP per Capita	-15.827***	-16.842***	-15.954***	-17.094***	-15.851***	-15.494***
	(1.799)	(1.853)	(1.832)	(1.869)	(1.847)	(1.871)
French Legal Origin	-12.687**	-13.910**	-11.987*	-13.822**	-16.196***	-12.733**
	(6.021)	(6.173)	(6.124)	(6.207)	(5.454)	(5.495)
German Legal Origin	-19.481***	-20.980***	-18.701***	-19.981***	-21.968***	-18.971***
	(4.727)	(4.640)	(4.775)	(4.771)	(4.198)	(4.437)
UK Legal Origin	-16.269**	-14.461**	-15.897**	-14.416**	-19.161***	-15.875***
	(6.272)	(6.070)	(6.396)	(6.161)	(5.623)	(5.531)
Constant	181.425***	196.579***	181.050***	197.906***	182.370***	177.351***
	(20.608)	(20.441)	(20.769)	(20.419)	(17.724)	(17.695)
Observations	149	122	149	122	171	171
R^2	0.849	0.848	0.850	0.849	0.822	0.826

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the child mortality rate per 1000 live births. The Global ELF and the Local Learning variables are measured at level 5 of aggregation and are based on the authors' calculations. In columns (1) and (3) decentralization is equal to 1 if country is a federal state, and 0 otherwise, whereas in columns (2) and (4) it is equal to 1 if subnational governments have autonomy or residual powers, and 0 otherwise. The variable definitions and data sources for each of the variables are provided in Appendix A.

Different levels of linguistic and geographic aggregation. Desmet et al. (2012) find that diversity measured at an intermediate level of linguistic aggregation is most significant for the provision of public goods. This is why our baseline analysis aggregates languages up to level 5 (out of a maximum 15). To see whether our results change for lower and higher levels of linguistic aggregation, we recompute our measures of global fractionalization and local learning for levels 15 and 2. Recall that at level 15 closely-related dialects, such as Alemannisch and Bavarian, are considered to be different languages, whereas at level 2 all Germanic languages, such as English and German, pertain to the same group. Columns (1) and (2) in Table 4 report our baseline regression, using level 15 and level 2. The results are virtually identical; we only see a small drop in the statistical significance of local learning, from 1% to 5%.

Table 4. Child Mortality: Linguistic Aggregation, Income Inequality and Ethnic Inequality

	(1) Level 15	(2) Level 2	(3) 10km x 10km	(4) Gini	(5) Linguistic Inequality
Global ELF	30.154***	37.885**	43.074***	48.187***	41.348***
	(9.238)	(15.410)	(10.595)	(11.379)	(12.390)
Local Learning	-83.219**	-90.960**	-128.141***	-163.046***	-112.423***
	(38.757)	(45.239)	(38.250)	(54.205)	(41.863)
Gini				-0.079	
				(0.434)	
Linguistic Inequality					10.424
					(8.329)
Absolute Latitude	-0.015	-0.137	-0.118	0.067	0.022
	(0.214)	(0.204)	(0.196)	(0.254)	(0.207)
Latin America & Carib.	0.292	-7.280	-4.198	0.460	-1.666
	(6.117)	(5.461)	(5.442)	(9.400)	(5.936)
Sub-Saharan Africa	56.759***	59.321***	59.105***	65.777***	63.025***
	(8.247)	(8.412)	(7.847)	(9.967)	(7.997)
East and S.E. Asia	-5.380	-4.927	-7.057	-3.560	-8.128
	(8.750)	(7.876)	(8.171)	(9.073)	(8.260)
Log GDP per Capita	-15.674***	-16.810***	-15.513***	-17.001***	-14.868***
	(1.820)	(1.913)	(1.850)	(2.327)	(1.900)
French Legal Origin	-12.739**	-14.424**	-13.091**	-15.050**	-8.791
	(5.461)	(5.668)	(5.201)	(6.054)	(6.090)
German Legal Origin	-18.853***	-22.418***	-19.670***	-19.629***	-14.570***
	(4.744)	(4.474)	(4.169)	(4.709)	(5.262)
UK Legal Origin	-15.993***	-16.821***	-16.166***	-15.444**	-11.271*
	(5.956)	(5.878)	(5.397)	(6.892)	(6.606)
Constant	172.810***	192.504***	177.830***	184.755***	158.504***
	(18.404)	(18.255)	(17.404)	(22.094)	(19.697)
Observations	171	171	171	124	165
R^2	0.819	0.812	0.826	0.843	0.827

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is child mortality rate per 1000 live births. The column headings refer to the robustness exercises carried out. The Global ELF and the Local Learning variables are measured at level 5 of aggregation, except in columns 1 and 2 where they are measured at levels 15 and 2 of aggregation. In column 3, we use the Local Learning variable calculated at the 10 km x 10 km spatial resolution, instead of 5 km x 5 km. In column 4, we control for the income Gini coefficient, while in column 5 we control for income inequality between linguistic groups. The variable definitions and data sources for each of the variables are provided in Appendix A.

Our regressions so far are based on local learning that take 5 km by 5 km cells to be a reasonable area for an individual's daily interaction. We next analyze whether our results are sensitive to that particular choice. Column (3) reports our findings when computing local learning based on 10 km by 10 km cells. Essentially nothing changes: when in the benchmark a one standard deviation increase in local learning lowered child mortality by 7.4 per thousand live births, it now lowers it by 7.8 per thousand live births.

Income and ethnic inequality. The degree of inequality in a country may be related to its linguistic diversity, and may affect public goods outcomes. Table 4, column (4), introduces the income Gini coefficient as an additional control. We do not find a direct effect of income inequality on child mortality, and the coefficients on global fractionalization and local learning hardly change.

Another concern is that income inequality between linguistic groups may be driving the results. To explore whether this is the case, we use the ethnic inequality variable of Alesina, Michalopoulos and Papaioannou (2016). They construct several measures of ethnic inequality. To be consistent with our measures of diversity, we use the one based on linguistic groups from the Ethnologue, aggregated to level 5. Column (5) in Table 4 reports the results when we control for this measures of linguistic inequality. Once again, the coefficients on global fractionalization and local learning are very similar to our baseline regression in Table 1.

Further robustness. Table 5 further analyzes the robustness of our findings in a variety of ways. We first establish that the results are not driven by particular regions. To that end, we drop, one at a time, sub-Saharan Africa, East and Southeast Asia and Latin America & the Caribbean. Columns (1), (2) and (3) show that the results are unchanged: local learning is associated with lower child mortality, while the opposite is true for global fractionalization. Next we replace legal origins by colonial origins. Column (4) shows that this does not affect our main findings. As another robustness check we replace our three regional dummies by the full set of six World Bank regional dummies. The results can be seen in column (5); nothing changes. In column (6) we add a number of additional geographic dummies, such as roughness of terrain, mean elevation, soil fertility and being landlocked, while dropping endogenous variables, such as GDP per capita. We continue to find a benign effect of local learning on child mortality. Lastly, column (7) reports the most comprehensive specification, including GDP per capita, population and a full set of geographic controls. Our coefficients of interest do not change qualitatively.

Table 6 takes this most comprehensive specification and replicates it for all other public goods outcomes. As before, higher global fractionalization is associated with worse outcomes, whereas more local learning is associated with improved outcomes. The coefficients on global fractionalization are statistically significant at the 5% level in four out of the six outcomes, whereas the coefficients on local learning are always statistically significant at the 5% level, with the exception of road density. This confirms our findings in Table 2. From these different empirical exercises we can conclude that the evidence in favor of contact theory holds across a broad variety of specifications and for a wide spectrum of public goods.

Table 5. Child Mortality: Further Robustness

	(1) Drop sub-Sahara	(2) Drop East S.E. Asia	(3) Drop Latin America	(4) Colonial Origin	(5) WB Regions	(6) Geography	(7) All
Global ELF	29.510***	47.486***	49.685***	41.469***	41.293***	60.671***	40.384***
	(9.783)	(12.279)	(11.929)	(11.057)	(11.825)	(13.562)	(10.813)
Local Learning	-71.217**	-127.499***	-161.519***	-107.828**	-111.595**	-197.822***	-139.469***
	(28.085)	(41.789)	(47.176)	(45.585)	(43.227)	(50.101)	(38.865)
Absolute Latitude	-0.040	-0.166	-0.173	-0.085	-0.082	-0.796***	-0.039
	(0.173)	(0.201)	(0.222)	(0.200)	(0.236)	(0.296)	(0.264)
Log GDP per Capita	-14.289***	-15.291***	-15.041***	-15.449***	-15.235***		-13.975***
	(1.612)	(1.923)	(1.926)	(1.996)	(2.069)		(2.269)
Soil Fertility						1.665	-16.077**
						(7.421)	(6.879)
Roughness						-42.014	10.822
						(28.484)	(30.302)
Elevation						15.516*	-2.899
						(7.948)	(9.301)
Island						2.892	2.882
						(10.050)	(8.629)
Landlocked						18.447***	14.109**
						(6.612)	(6.189)
Log Population							1.201
							(1.366)
Constant	159.573***	178.174***	176.374***	162.511***	182.144***	63.754***	123.883***
	(16.735)	(18.307)	(18.367)	(15.229)	(25.268)	(24.382)	(31.384)
Baseline Regional Dummies	Yes	Yes	Yes	Yes	No	Yes	Yes
WB Regional Dummies	No	No	No	No	Yes	No	No
Legal Origins	Yes	Yes	Yes	No	Yes	Yes	Yes
Colonial Origins	No	No	No	Yes	No	No	No
Observations	125	159	139	175	171	149	147
R^2	0.690	0.832	0.827	0.827	0.827	0.811	0.869

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is child mortality rate per 1000 live births. The Global ELF and the Local Learning variables are measured at level 5 of aggregation and are based on the authors' calculations. The column headings refer to the robustness exercises carried out. Column (1) drops countries from sub-Saharan Africa, column (2) drops countries from East and South East Asia, column (3) drops countries from Latin American and the Caribbean, column (4) replaces legal origin by colonial origin, column (5) replaces the three regional dummies of the baseline by six regional dummies commonly used by the World Bank, column (6) controls for additional geographic features, and column (7) is the most comprehensive specification that includes both the additional geographic features and population. The variable definitions and data sources for each of the variables are provided in Appendix A.

Table 6. Other Outcomes: Comprehensive Specification

	(1) Measles Immunization	(2) Hospital Beds	(3) Illiteracy Rate	(4) Schooling	(5) Improved Sanitation	(6) Road Density
Global ELF	-26.974***	-1.155*	31.452***	-0.289**	-29.997***	1.790
	(5.056)	(0.660)	(6.857)	(0.130)	(6.905)	(2.053)
Local Learning	59.778***	8.862**	-114.863***	1.448***	97.611***	1.010
	(21.005)	(3.984)	(27.079)	(0.545)	(33.832)	(16.806)
Log GDP per Capita	0.780	0.241	-6.412***	0.153***	9.185***	1.979***
	(1.022)	(0.177)	(1.549)	(0.023)	(1.497)	(0.621)
Absolute Latitude	0.083	0.082***	-0.101	0.000	-0.031	0.137
	(0.115)	(0.025)	(0.165)	(0.003)	(0.171)	(0.107)
Soil Fertility	3.581	0.932	-9.469**	0.311***	2.164	2.109
	(2.944)	(0.594)	(4.602)	(0.077)	(4.176)	(2.439)
Roughness	-13.251	-0.018	2.422	-0.587**	26.250	-24.182***
	(12.354)	(2.041)	(16.136)	(0.290)	(18.752)	(7.728)
Elevation	1.560	-0.595	-5.089	0.206**	-6.418	5.449***
	(3.406)	(0.608)	(4.727)	(0.089)	(5.616)	(1.779)
Island	-7.682*	1.033	-2.273	0.042	-7.469	3.462
	(4.059)	(0.790)	(5.996)	(0.083)	(5.580)	(2.253)
Landlocked	-1.405	1.039**	-1.500	-0.084	2.595	-2.662*
	(2.563)	(0.473)	(3.769)	(0.070)	(3.412)	(1.422)
Log Population	-0.964	-0.057	0.140	-0.023	-0.451	-1.220***
	(0.646)	(0.124)	(0.871)	(0.016)	(1.044)	(0.414)
Constant	100.563***	0.857	60.133***	1.382***	12.096	20.778*
	(16.010)	(3.760)	(20.063)	(0.378)	(24.384)	(12.103)
Legal Origins	Yes	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	147	147	121	126	145	147
R^2	0.638	0.678	0.707	0.739	0.821	0.518

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The column headings give the dependent variables for each of the columns. The Global ELF and the Local Learning variables are measured at level 5 of aggregation and are based on the authors' calculations. The specification is the same as the one of column (7) in Table 5. The variable definitions and data sources for each of the variables are provided in Appendix A.

4.2 Local Learning and Public Goods (IV Analysis)

In spite of the evidence consistent with contact theory, our findings so far fall short in establishing a causal relation between local learning and public goods provision. A potentially important endogeneity concern is whether worse provision of public goods may give individuals of the same group an incentive to geographically cluster. For example, evidence by Greif (1993) and others have suggested that in the absence of an effective state, ethnic groups may be able to provide security and contract enforcement to individuals of their own groups. If so, in states with relatively poor public goods provision, people of the same linguistic or ethnic group may choose to live together. This would imply a positive effect of public goods provision on the degree of spatial mixing, and hence on local learning.

To see whether this is a serious concern, we start by considering the empirical evidence on whether spatial sorting by ethnolinguistic group is likely to be affected by public goods provision. Gershman and Rivera (2016) use census data for a number of African countries, and find that subnational diversity is very stable over time. In particular, the correlation in subnational diversity over two to three decades exceeds 0.95. More importantly for our purpose, they find that the small changes in subnational diversity are not related to regional economic performance. While this somewhat alleviates the endogeneity concern, their results are limited to a subset of African countries. In what follows we therefore design an instrumental variable strategy, with the aim of causally establishing the relation between local learning and public goods provision.

Instrument for local learning. Our endogeneity concern relates to the spatial distribution of a country's language groups across its territory. That is, it does not pertain to the number and the population shares of the groups, but to their geographic distribution. More specifically, we want to develop an instrument for $\mathcal{B}$, our K x M binary matrix of which the elements take a value 1 if the language corresponding to the column is spoken in the cell corresponding to the row (and a value 0 otherwise).

To that end, we follow an approach similar to the one in Alesina and Zhuravskaya (2011), and create a predicted measure of local language use by relying on language use in neighboring countries. In particular, for each cell ℓ in country c, we determine the closest cell k in any of the neighboring countries of c. Any language that is spoken in k and that is also spoken in c is then assigned to ℓ. For languages that are spoken in c and that are not spoken in any of the closest cells in the neighboring countries, we assume that they are spoken in all cells of c. This methodology yields a K x M binary matrix $\hat{\mathcal{B}}$ with predicted values of language use. We then use the same algorithm as the one described in Section 3.2, but using $\hat{\mathcal{B}}$ instead of $\mathcal{B}$. This yields an instrument for local learning. Appendix C provides a simple example of how neighboring countries are used to predict a country's geographic distribution of languages.

IV results. Table 7 reports the IV results for our baseline regression. For five of the seven public goods, local learning has the expected sign and is statistically significant at the 5% level. (For the other two outcomes, the effect is not statistically significant.) The F-statistics of the first stage are all larger than the Stock-Yogo critical values, so we can reject the hypothesis that the instruments are weak. In general we find the effect of local learning to be between 50% and 150% larger than in the OLS regressions. For example, while a one standard deviation increase in local learning lowered child mortality by 7.4 per thousand in the OLS regression, it now lowers child mortality by 17.0 per thousand. The corresponding standardized β is -30%.

Finally, Table 8 reports the IV regressions for all outcomes, using the most comprehensive specification of Table 5, which includes a host of geographic controls, as well as legal origins, regional dummies, population and GDP per capita. With the exception of road density, for which we get the wrong sign, the results for local learning become even stronger. A one-standard deviation increase in local learning lowers child mortality by an estimated 19.9 per thousand, with a corresponding standardized β of -34%. From this we can conclude that there is a causal effect between higher local learning and improved public goods outcomes. The evidence hence overwhelmingly supports contact theory.

Table 7. Global ELF and Local Learning (IV)

	(1) Child Mortality	(2) Measles Immunization	(3) Hospital Beds	(4) Illiteracy Rate	(5) Schooling	(6) Improved Sanitation	(7) Road Density
Global ELF	67.303***	-31.891***	-0.414	42.760***	-0.404**	-31.685***	4.460**
	(14.104)	(5.375)	(0.798)	(7.433)	(0.168)	(7.540)	(2.049)
Local Learning	-288.082***	108.475***	1.448	-175.505***	2.309**	126.256***	-16.790
	(77.394)	(28.235)	(4.235)	(44.429)	(0.970)	(42.452)	(12.676)
Absolute Latitude	-0.097	0.195**	0.102***	-0.250**	0.008***	0.134	0.177***
	(0.207)	(0.087)	(0.019)	(0.122)	(0.003)	(0.160)	(0.062)
Latin America & Carib.	-2.550	3.937	0.343	-4.975	0.210**	-5.265	2.534
	(6.308)	(2.771)	(0.430)	(3.605)	(0.089)	(4.887)	(1.697)
Sub-Saharan Africa	58.505***	-9.175***	0.098	8.819**	-0.125	-25.074***	2.891*
	(7.490)	(2.843)	(0.457)	(4.084)	(0.103)	(5.081)	(1.702)
East and S.E. Asia	-11.803	3.598	0.811	-15.298***	0.113	-5.248	-1.939
	(7.947)	(3.621)	(0.859)	(4.961)	(0.102)	(6.392)	(2.097)
Log GDP per Capita	-14.019***	0.945	0.306**	-3.441***	0.109***	9.750***	2.008***
	(1.929)	(0.819)	(0.154)	(1.140)	(0.022)	(1.255)	(0.447)
French Legal Origin	-7.658	0.900	0.693	4.078	0.049	11.975***	-13.205*
	(6.094)	(2.970)	(1.415)	(2.735)	(0.087)	(4.206)	(7.598)
German Legal Origin	-14.041***	1.873	2.511*	0.386	0.206***	12.083***	-10.846
	(5.177)	(2.806)	(1.437)	(4.366)	(0.074)	(3.802)	(7.741)
UK Legal Origin	-11.164*	6.353**	0.713	0.000	0.233***	9.313**	-10.195
	(6.056)	(3.188)	(1.425)	(.)	(0.084)	(4.160)	(7.786)
Constant	162.045***	71.059***	-2.572	45.969***	0.731***	-12.903	-2.862
	(19.162)	(8.855)	(1.953)	(11.464)	(0.258)	(14.199)	(9.155)
First-Stage F-Statistic	59.531	59.531	60.314	39.262	36.197	62.082	59.596
Observations	171	171	173	138	136	171	172
R^2	0.808	0.561	0.589	0.640	0.667	0.779	0.443

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table gives the second stage of 2SLS IV regressions. The column headings give the dependent variables for each of the columns. The Global ELF and the Local Learning variables are measured at level 5 of aggregation and are based on the authors' calculations. The Local Learning variable has been instrumented using a predicted Local Learning variable based on languages spoken in neighboring countries. The variable definitions and data sources for each of the variables are provided in Appendix A.

5 Concluding Remarks

In this paper we have developed a model of antagonism which suggests that both a society's overall linguistic diversity and local learning about other groups should affect the provision of public goods. Empirically, we have found that overall linguistic diversity worsens the provision of public goods, whereas the opposite is true for local learning. These findings are consistent with contact theory which hypothesizes that prejudice against other groups is mitigated by frequent interaction with those groups. More generally, our results indicate that the geography of diversity is key for understanding its impact on public good provision.

An important concern when analyzing the effect of the spatial distribution of diversity on different outcomes is reverse causality. Countries that are unsuccessful in providing public goods may give individuals an incentive to geographically cluster with others of their own group. To address this concern, we developed an IV, by using the neighbors' languages to predict a country's spatial distribution of diversity. Doing so allowed us to causally establish the positive effect of local learning on public goods outcomes.

Table 8. Global ELF and Local Learning, Comprehensive Specification (IV)

	(1) Child Mortality	(2) Measles Immunization	(3) Hospital Beds	(4) Illiteracy Rate	(5) Schooling	(6) Improved Sanitation	(7) Road Density
Global ELF	69.395***	-35.918***	0.002	49.004***	-0.393**	-40.629***	8.301***
	(14.193)	(6.143)	(1.000)	(9.468)	(0.171)	(9.464)	(3.006)
Local Learning	-379.314***	133.714***	-0.702	-254.244***	2.226*	185.798***	-52.823**
	(95.097)	(43.140)	(7.514)	(73.119)	(1.315)	(70.870)	(22.518)
Log GDP per Capita	-12.844***	0.432	0.286*	-5.379***	0.148***	8.817***	2.233***
	(2.392)	(1.025)	(0.166)	(1.616)	(0.023)	(1.482)	(0.626)
Absolute Latitude	0.043	0.058	0.086***	-0.006	0.000	-0.062	0.155
	(0.267)	(0.112)	(0.024)	(0.183)	(0.003)	(0.169)	(0.100)
Soil Fertility	-15.722**	3.471	0.946*	-9.031**	0.311***	2.115	2.189
	(6.827)	(2.863)	(0.563)	(4.586)	(0.072)	(4.149)	(2.260)
Roughness	-13.491	-5.756	-0.988	-13.987	-0.503	35.584*	-29.639***
	(31.833)	(13.404)	(2.111)	(17.916)	(0.308)	(19.120)	(8.154)
Elevation	3.083	-0.284	-0.357	-1.133	0.187**	-8.683	6.792***
	(9.426)	(3.491)	(0.647)	(4.969)	(0.087)	(5.283)	(1.897)
Island	2.898	-7.687**	1.034	-0.255	0.043	-7.506	3.466*
	(7.896)	(3.767)	(0.746)	(6.235)	(0.077)	(5.162)	(2.030)
Landlocked	13.037**	-1.075	0.996**	-1.573	-0.089	2.999	-2.903*
	(6.359)	(2.555)	(0.459)	(3.637)	(0.066)	(3.448)	(1.547)
Log Population	-0.920	-0.311	-0.141	-0.988	-0.017	0.353	-1.696***
	(1.632)	(0.730)	(0.136)	(0.926)	(0.016)	(1.143)	(0.528)
Constant	142.799***	94.732***	1.611	69.451***	1.330***	4.331	25.024**
	(32.700)	(16.208)	(3.645)	(22.783)	(0.346)	(24.216)	(11.934)
Legal Origins	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-Stage F-Statistic	30.109	30.109	30.109	19.573	21.763	29.455	30.109
Observations	147	147	147	121	126	145	147
R^2	0.844	0.606	0.664	0.643	0.734	0.810	0.466

Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table gives the second stage of 2SLS IV regressions. The column headings give the dependent variables for each of the columns. The Global ELF and the Local Learning variables are measured at level 5 of aggregation and are based on the authors' calculations. The Local Learning variable has been instrumented using a predicted Local Learning variable based on languages spoken in neighboring countries. The variable definitions and data sources for each of the variables are provided in Appendix A.

Another key contribution of this paper is the construction of a worldwide dataset on local language use. To that end, we combined data on local population, country-level language use, and local language maps. We then applied an iterative proportional fitting algorithm to allocate the speakers of 6,905 different languages to all 5 km by 5 km cells in the world. This database should be useful for researchers interested in analyzing the effect of local diversity on a variety of political economy outcomes, such as development and conflict.

References

[1] Alesina, A., Baqir, R., and Easterly, W., 1999. "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 114, 1243-1284.

[2] Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R., 2003. “Fractionalization,” *Journal of Economic Growth*, 8, 15594.

[3] Alesina, A. and La Ferrara, E., 2000. “Participation in Heterogeneous Communities,” *Quarterly Journal of Economics*, 115, 847-904.

[4] Alesina, A. and La Ferrara, E., 2005. “Ethnic Diversity and Economic Performance,” *Journal of Economic Literature*, 43, 762-800.

[5] Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). “Ethnic Inequality,” *Journal of Political Economy*, 124, 428-488.

[6] Alesina, A. and Zhuravskaya, E., 2011. “Segregation and the Quality of Government in a Cross-Section of Countries,” *American Economic Review*, 101, 1872-1911.

[7] Ashraf, Q. and Galor, O., 2013. “The ‘Out of Africa’ Hypothesis, Human Genetic Diversity, and Comparative Economic Development,” *American Economic Review*, 103, 1-46.

[8] Barro, R. and Lee, J. (2010). “A New Data Set of Educational Attainment in the World, 1950-2010,” NBER Working Paper #15902.

[9] Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W., 1975. *Discrete Multivariate Analysis: Theory and Practice*, MIT Press.

[10] Allport, G., 1954. *The Nature of Prejudice*, Reading, MA: Addison-Wesley.

[11] Central Intelligence Agency (2001). *The World Factbook 2001*, Washington, DC: Central Intelligence Agency.

[12] Chiba, E., 2006. “English Usage in Mauritius,” `http://homes.chass.utoronto.ca/~cpercy/courses/6362-chiba.htm` (accessed September 1, 2014)

[13] Cornes, R. and Hartley, R., 2007. “Aggregative Public Good Games,” *Journal of Public Economic Theory*, 9, 201-219.

[14] Deming, W. E. and Stephan, F. F., 1940. “On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known,” *Annals of Mathematical Statistics,* 11, 427-444.

[15] Desmet, K., Ortuño-Ortín, I., and Wacziarg, R., 2012. “The Political Economy of Linguistic Cleavages,” *Journal of Development Economics*, 97, 322-338.

[16] Esteban, J. M. and Ray, D., 1994. “On the Measurement of Polarization,” *Econometrica*, 62, 819-851.

[17] Fienberg, S. E., 1970. “An Iterative Procedure for Estimation in Contingency Tables,” *Annals of Mathematical Statistics*, 41, 907-917.

[18] Finseraas, H., Hanson, T., Johnsen, A. A., Kotsadam, A., and Torsvik, G., 2016. "Trust, Ethnic Diversity, and Personal Contact: Experimental Field Evidence," unpublished manuscript.

[19] Gershman, B. and Rivera, D., 2016. "Subnational Diversity in Sub-Saharan Africa: Insights from a New Dataset," mimeo.

[20] Greif, A., 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *American Economic Review*, 83, 525-48.

[21] Karaivanov, A., 2009. "Heterogeneity, Returns to Scale, and Collective Action," *Canadian Journal of Economics*, 42, 771-807.

[22] La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R., 1999. "The Quality of Government," *Journal of Law, Economics, and Organization*, 15, 22279.

[23] La Porta, R., Lopez de Silanes, F., and A. Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature*, 46, 285-332.

[24] Matuszeski, J. and Schneider, F., 2006. "Patterns of Ethnic Group Segregation and Civil Conflict," Harvard University, unpublished working paper.

[25] Miguel, E. and Gugerty, M.K., 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, 89, 2325-2368.

[26] Montalvo, J.G. and Reynal-Querol, M., 2005. "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 95, 796-816.

[27] Montalvo, J.G. and Reynal-Querol, M., 2016. "Ethnic Diversity and Growth: Revisiting the Evidence," UPF, mimeo.

[28] Munshi, K. and Rosenzweig, M., 2015. "Insiders and Outsiders: Local Ethnic Politics and Public Goods Provision," NBER Working Paper # 21720.

[29] Pettigrew, T.F. and Tropp, L.R., 2006. "A Meta-Analytic Test of Intergroup Contact Theory," *Journal of Personality and Social Psychology*, 90, 751-783.

[30] Putnam, R.D., 2007. "*E Pluribus Unum*: Diversity and Community in the Twenty-First Century," *Scandinavian Political Studies*, 30, 137-174.

[31] Robinson, A.L., 2013. "Ethnic Diversity, Segregation, and Ethnocentric Trust in Africa," Stanford University, unpublished working paper.

[32] Treisman, D., 2008. Decentralization Dataset. Downloaded on August 10, 2016 at `http://www.sscnet.ucla.edu/polisci/faculty/treisman/`.

A. Data Appendix

Child mortality. Child mortality rate per 1,000 live births, 1990-2010 average. *Source: World Development Indicators, World Bank.*

Hospital beds. Hospital beds per 1,000 people, 1990-2010 average. *Source: World Development Indicators, World Bank.*

Measles Immunization. Percentage of children between the age of 12 and 23 months that have been immunized against measles, 1990-2010 average. *Source: World Development Indicators, World Bank.*

Improved sanitation. Percentage of population with access to improved sanitation facilities, 1990-2010 average. *Source: World Development Indicators, World Bank.*

Roads. Road network density, km per 1,000 people, 1990-2010 average. *Source: World Development Indicators, World Bank.*

Illiteracy. Percentage of people aged 15 and above who are illiterate, 1990-2010 average. *Source: World Development Indicators, World Bank.*

School attainment. Log of 1 + average years of schooling for people 25 years of age and above, 1990-2010 average. *Source: Barro R. and J.W. Lee v. 1.3, 04/13.*

Decentralization. Two measures of decentralization: countries classified as federal states, and countries of which subnational legislatures have either autonomy in certain specified areas or residual powers to legislate in areas not explicitly assigned to other levels of government. *Source: Treisman (2008).*

Log GDP per capita and log population. Both variables are the average for the period 1990-2010. *Source: World Development Indicators, World Bank.*

Legal origin. French, German or UK legal origin *Source: La Porta, Lopez-de-Silanes and Shleifer (2008).*

Colonial origin. Country from which a country became independent. *Source: CIA World Factbook (2001).*

Geographic controls. Absolute latitude, log of soil fertility, roughness of terrain and mean elevation. *Source: Ashraf and Galor (2013).*

Gini coefficient. Income Gini coefficient, 1990-2010 average. *Source: World Development Indicators, World Bank.*

Linguistic inequality. Income inequality between ethnic groups, based on linguistic groups of Ethnologue aggregated to level 5. *Source: Alesina, Michalopoulos and Papaioannou (2016).*

Country boundary shapefile. ArcGIS shapefile with the political boundaries of all countries in the world. *Source: Seamless Digital Chart of the World Base Map Version 10.0, World GeoDatasets.*

Ethno-linguistic maps. For information on linguistic groups we use the digitized version of the 16^{th} edition of Ethnologue which maps over 6,905 linguistic groups for the whole world. The data on different languages come from a variety of censuses and years and approximately correspond to the 1990s. The digitized version of the Ethnologue is a polygon shape file where 6,905 languages spoken in the world are represented as polygons across space, where each polygon represents the homeland of a particular linguistic group. Areas where multiple languages are spoken are represented by overlapping polygons. Ethnologue also provides the total population pertaining to each particular linguistic polygon within the political boundaries of each country. When certain widely spoken languages cannot be assigned to any particular homeland in a given country, Ethnologue classifies such languages as widespread languages, which are represented as random points within the geographic boundary of the country where the language is spoken. The respective populations of each of these languages are also provided. Since these languages are widespread, there are speakers of these languages randomly distributed across the country rather than being restricted to particular linguistic homelands. *Source: World Language Mapping System Version 16, World GeoDatasets.*

Cell Level Population data. The cell level population data comes from LandScan who provide global population distribution data at the resolution of approximately 1km X 1km (30" X 30") which represents an ambient population (average over 24 hours). *Source:* `http://web.ornl.gov/sci/landscan/`

B. Appendix Figures

Figure B.1. Global ELF by Country – Level 2

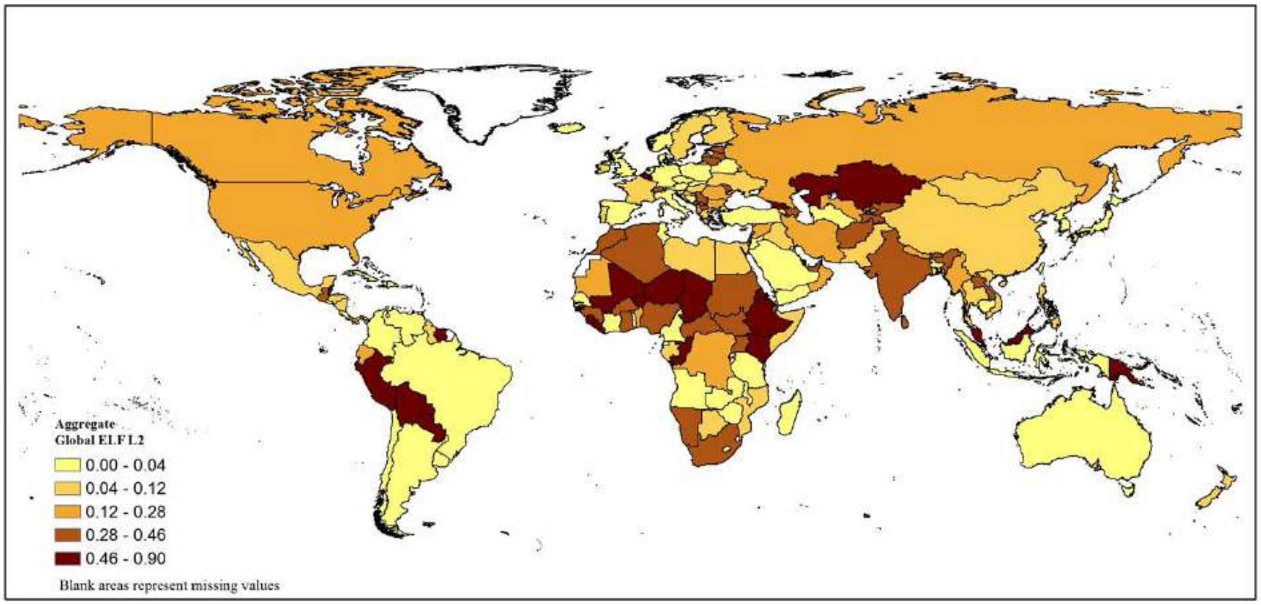


Figure B.2. Local Learning by Country – Level 2

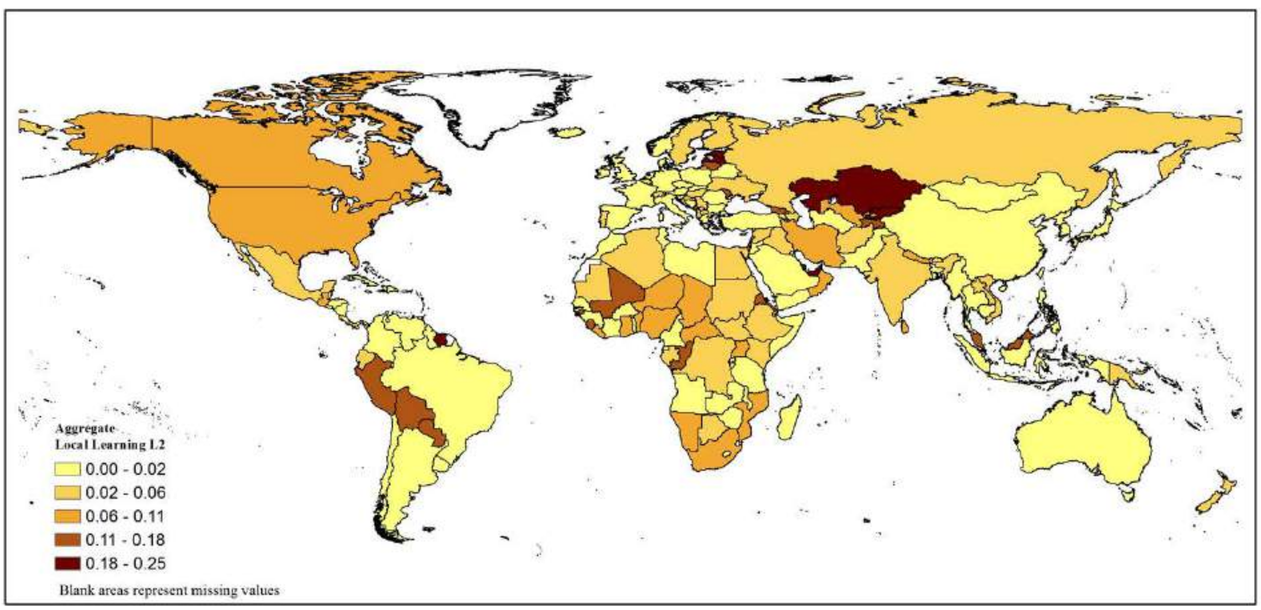


Figure B.3. Local Learning at 5 km by 5 km Resolution – Level 2

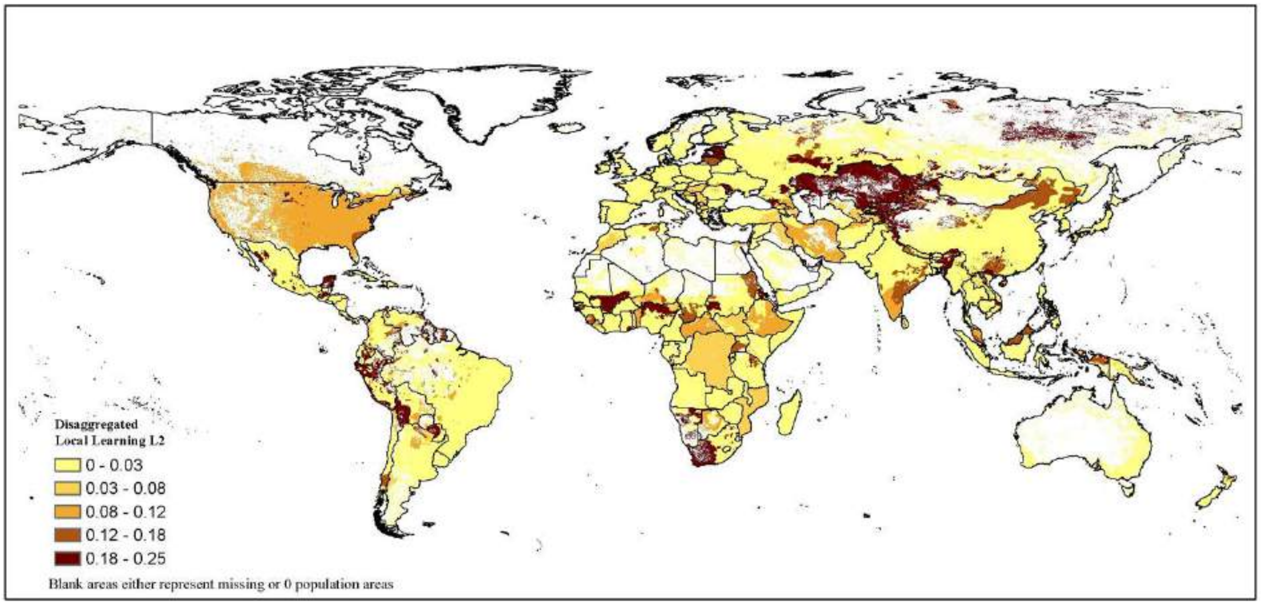


Figure B.4. Global ELF by Country – Level 5

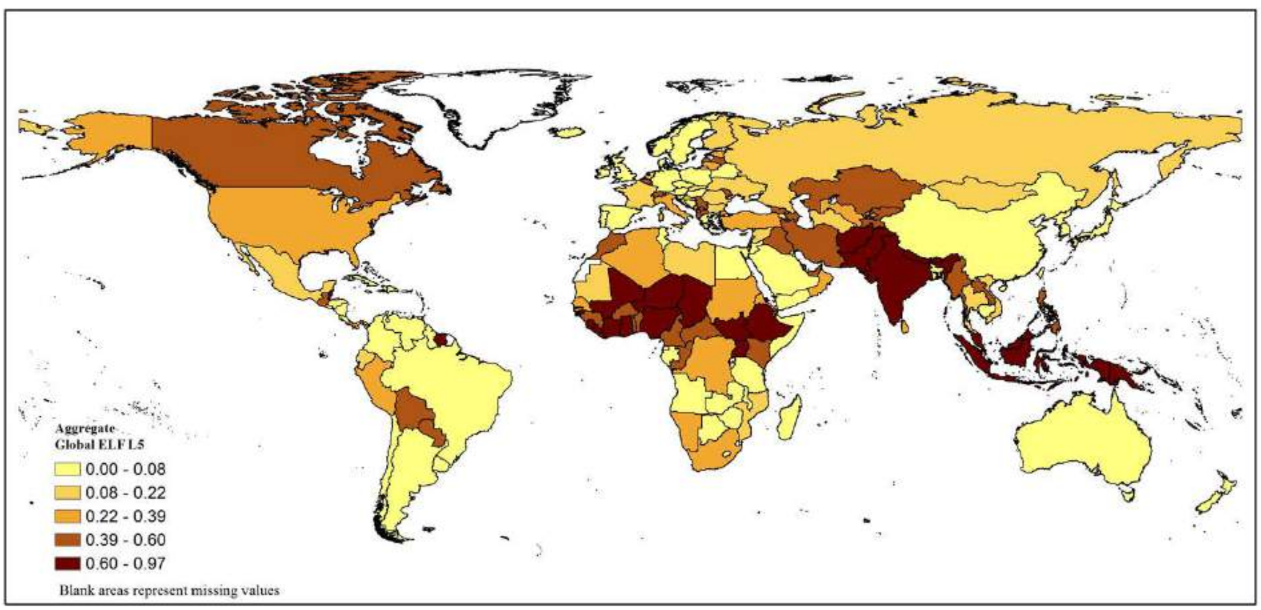


Figure B.5. Local Learning by Country – Level 5

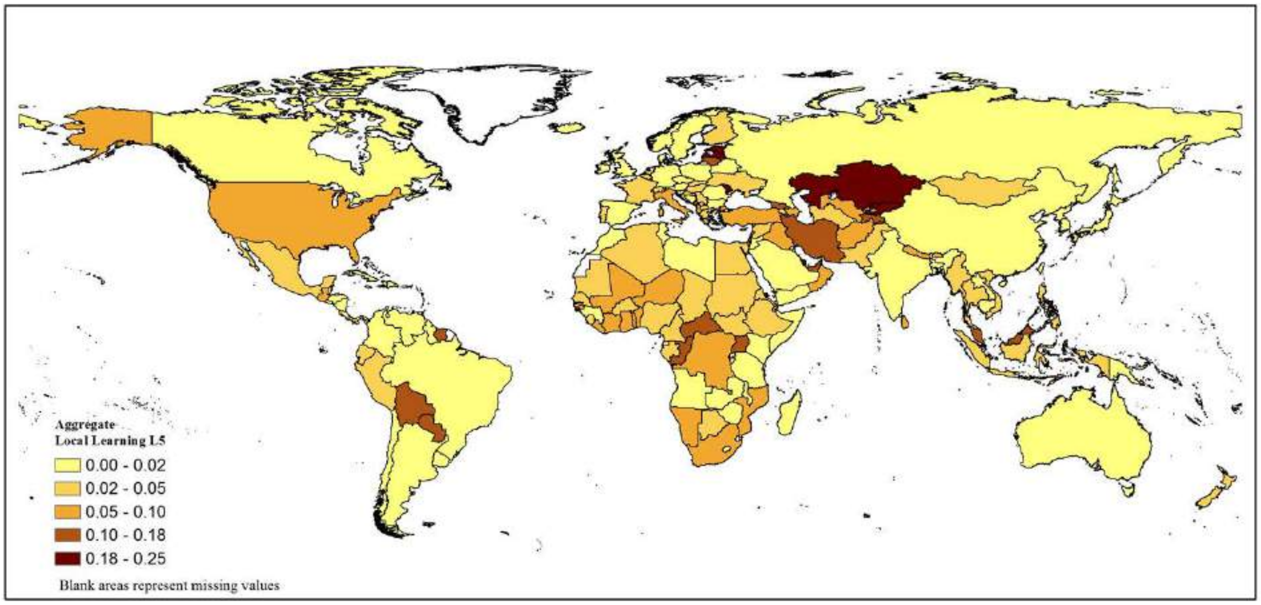
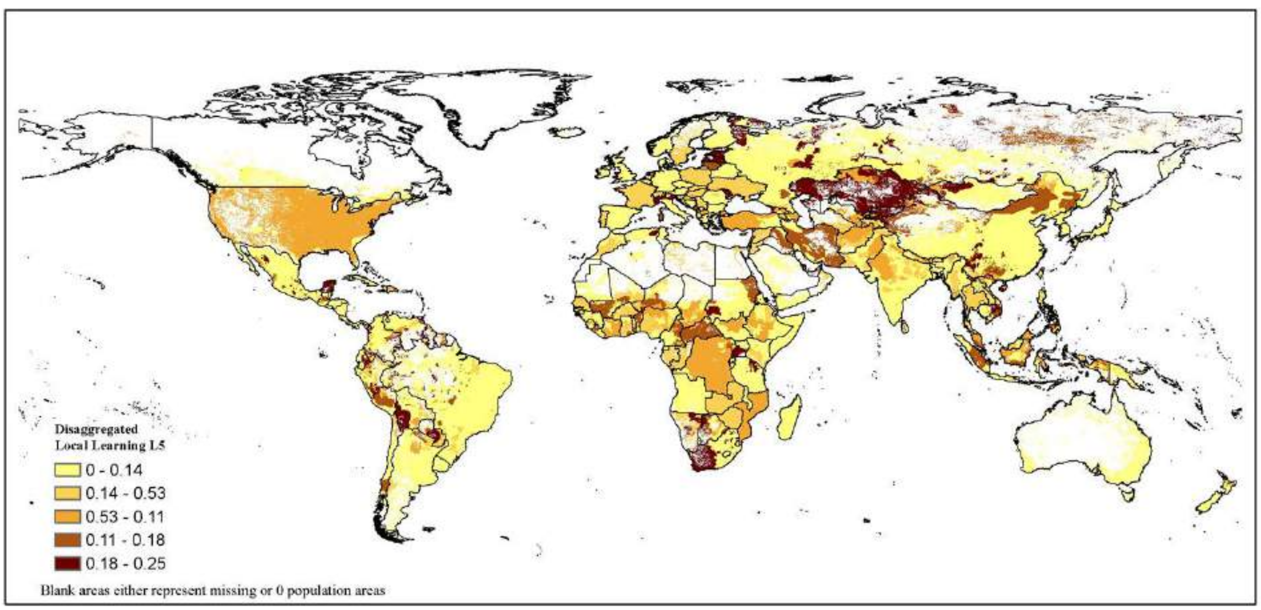


Figure B.6. Local Learning at 5 km by 5 km Resolution – Level 5

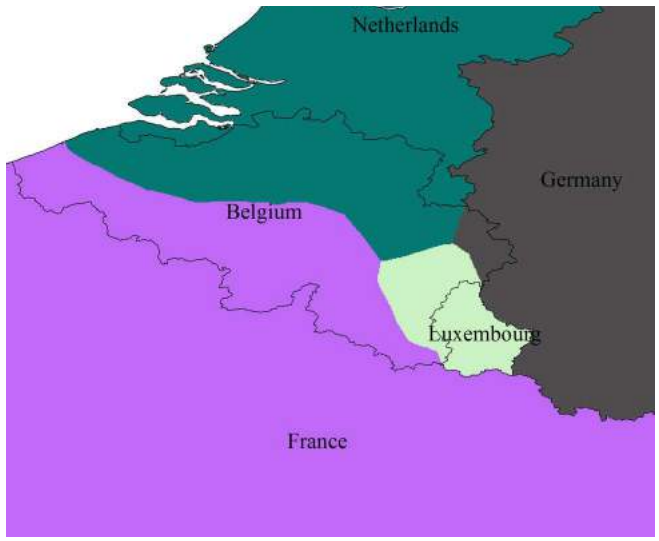


C. Instrument Construction

In this appendix we give an example of how neighboring countries are used to predict a country's geographic distribution of languages. We focus on the case of Belgium, and go through the following steps.

1. For each cell ℓ in Belgium, we determine the closest cell k in any of Belgium's neighboring countries of c. Figure C.1 illustrates this.

Figure C.1. Closest Neighbors of Each Cell in Belgium



2. Any language that is spoken in k and that is also spoken in Belgium is then assigned to ℓ. Figure C.2 represents one of the languages spoken in each cell k. In Figure C.3 we can see how the two previous figures are combined to assign a language to each cell ℓ in Belgium. Of course, more than one language may be spoken in a given cell k, including wide-spread languages. In that case, we use the same procedure for each of the languages spoken in k.

Figure C.2. Languages in Belgium's Neighbors

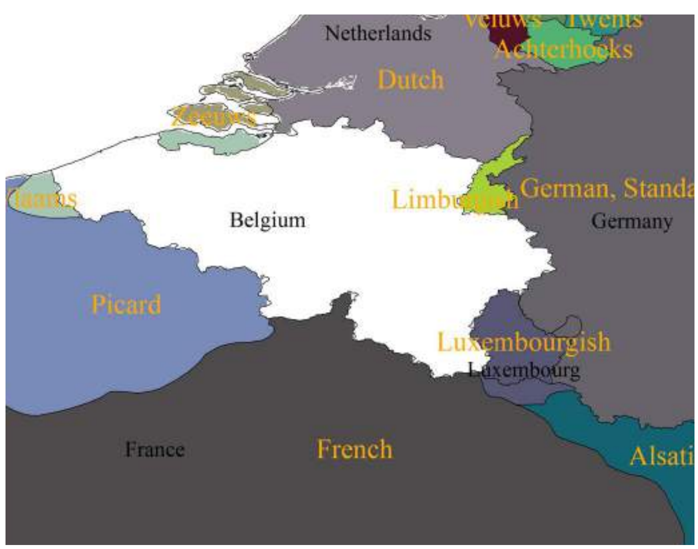
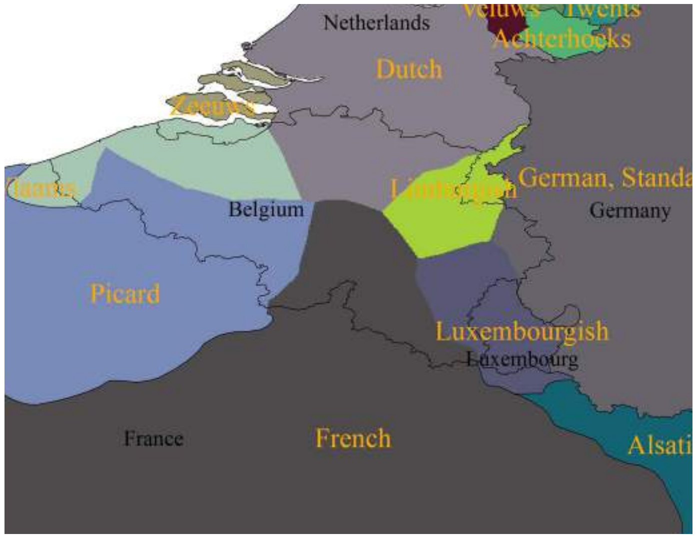


Figure C.3. Allocation of Neighbors' Languages to Belgium's Cells



3. For languages that are spoken in Belgium and that are not spoken in any of the closest cells in the neighboring countries, we assume that they are spoken in all of Belgium's cells.

4. The previous three steps yields a K x M binary matrix $\hat{\mathcal{B}}$ with predicted values of language use in Belgium. Note that the maps above are for languages at level 15; we can easily aggregate this information to the level we are interested in.

5. To allocate the number of language speakers to each cell in Belgium, we use the same algorithm as the one described in Section 3.2, but using $\hat{\mathcal{B}}$ instead of $\mathcal{B}$.

6. We then use this predicted allocation to construct our instrument for local learning in Belgium.