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

Klaus Desmet
SMU

Joseph Gomes
Universidad de Navarra

Ignacio Ortuño-Ortín
UC3M

August 2017

Abstract

This paper analyzes the importance of local interaction between individuals of different linguistic groups for the provision of public goods at the national level. The micro-founded theory we develop predicts that a country's public goods (i) decrease in its overall *linguistic fractionalization*, and (ii) either increase or decrease in its *local learning multiplier*, a measure of how local interaction affects antagonism towards other groups in the society at large. After constructing a 5 km by 5 km dataset on language use for 223 countries, we empirically test our theory. While overall fractionalization worsens public goods outcomes, we find a positive causal effect of local learning. Conditional on a country's overall diversity, public goods outcomes are maximized when there are a few large-sized groups and the diversity of each location mirrors that of the country as a whole. Our large-scale study, spanning the entire globe, confirms experimental micro-evidence in favor of contact theory.

Keywords: linguistic diversity, geographic diversity, public goods, contact theory, conflict theory, antagonism, local learning

JEL Codes: H4, O18, O57, R1, Z10, Z13

1 Introduction

Although living in ethnolinguistically heterogeneous societies can be challenging, within countries diversity is often viewed more positively in highly diverse areas than in relatively homogenous locations.[1] This is consistent with *contact theory*, which argues that repeated interpersonal contact tends to reduce prejudice and conflict between individuals of different groups (Allport, 1954). Not everyone agrees with contact theory though. *Conflict theory* posits the exact opposite: interaction with other groups generates greater prejudice and antagonism (Blumer, 1958). With some notable exceptions, most of the micro-evidence falls on the

[1]For example, in sub-Saharan Africa more heterogenous countries exhibit relatively lower degrees of inter-ethnic trust, but within countries we see the opposite: inter-ethnic trust tends to be higher in highly heterogeneous localities (Robinson, 2013). As another example, while the pro-Brexit movement ran on a platform of regaining control over the inflow of migrants into the United Kingdom, the perception of there being too much diversity was especially strong in areas with few foreign-born residents, and much less so in cosmopolitan London (Lawton and Ackrill, 2016).

side of contact theory.[2] However, for lack of a conceptual framework and a comprehensive database, no large-scale analysis of the issue, spanning the entire globe, has been conducted.

This paper starts by conceptualizing how the intensity of local interaction and contact with other groups affects the antagonism an individual feels towards the rest of society. In the absence of local interaction, it is intuitive to assume that antagonism is increasing in the share of those other groups in the society at large. In the presence of local interaction, the intensity of this antagonism may be either mitigated or reinforced, depending on whether contact theory or conflict theory prevails. According to these theories, the change in antagonism depends on the degree of interaction, and this change extends to the entire out-group. Illustrating this with an example, the more local contact a Dutch-speaking Belgian has with French-speaking Belgians, the more this affects her prejudice towards all French speakers, not just those residing in her own locality but also those living in the rest of the country. That is, the change in prejudice experienced by a Dutch speaker depends on the local share of French speakers, but the effect of this change extends to the aggregate share of French speakers.

To formalize these insights, we assume that the intensity of an individual's local interaction with another group is increasing in the local share of that group; and how much this affects his antagonism towards that other group in the society at large depends on the aggregate share of that group. We interpret these assumptions in terms of learning: the larger the share of the other group locally, the more an individual learns about that group through local interaction; and the larger the share of the other group in the aggregate, the more this local learning affects the individual's overall antagonism. As a result, the local group share and the aggregate group share are complementary in their effects on an individual's antagonism. The intuition for this complementarity is as follows: the change in an individual's opinion about another group can only have an important impact on her overall antagonism if that other group is sizable in the society at large. Because the local group share gets multiplied by the aggregate group share when evaluating the effect of local interaction on antagonism, we refer to this effect as the *local learning multiplier*.

While the local learning multiplier tends to be greater in societies with lower degrees of segregation, both concepts differ in a fundamental way. Consider two societies, one with two equally-sized large groups, and another with ten equally-sized small groups. Suppose groups in both societies are perfectly mixed geographically, so that there is no segregation in either society. In spite of this perfect spatial mixing, the local learning multiplier is larger in the first society, with two big groups, than in the second society, with ten small groups. This result can be understood in the following way. When an individual interacts locally with another group, she learns something about that group. Learning a lot about few groups and then applying this larger knowledge to few groups has a bigger impact than learning a little about many groups and then applying this smaller knowledge to many groups. This is another way of stating that the complementarity between local group shares and aggregate group shares yields a greater local learning multiplier in the society with a few large groups.

---

[2]In a meta-analysis by Pettigrew and Tropp (2006), 94 percent of 713 independent samples of 515 studies support contact theory. In contrast, Putnam (2007) finds evidence in favor of conflict theory.

When taking a society's aggregate ethnolinguistic composition as given, more local mixing tends to increase the local learning multiplier. However, the above discussion clarifies that a society with a high degree of local mixing does not necessarily have a large local learning multiplier. Indeed, in highly fractionalized societies with many groups, the local learning multiplier will be small, even if the different groups are perfectly mixed at the local level. The local learning multiplier will only be large in societies where people are clustered into a few large-sized groups that are geographically mixed. This makes the local learning multiplier reminiscent of a polarization index, though it relies on very different micro-foundations. As we will explain in more detail in the theory section, in the framework of Esteban and Ray (1994) polarization has to do with an individual's identification with the own group, whereas in our framework local learning is related to an individual's identification with other groups.

Why do we care about inter-group interaction and the local learning multiplier? Antagonism between groups is a leading explanation for why more diverse countries have worse political economy outcomes. This is why many empirical papers have explored the effect of ethnolinguistic fractionalization on different economic, social and political outcomes.[3] Our theory suggests that these studies suffer from an omitted variable bias when not controlling for local interaction. To be more precise, our conceptual framework yields an antagonism index that is a linear combination of two terms: the society's overall ethnolinguistic fractionalization and the society's local learning multiplier. In the absence of local interaction, antagonism is measured by a society's overall ethnolinguistic fractionalization, henceforth referred to as its *global fractionalization*; in the presence of local interaction, that antagonism is either diminished or augmented by the local learning multiplier. Drawing the correct inference about the effect of diversity on different political economy outcomes requires us to control for both terms. In this paper, we focus on a variety of public goods outcomes related to health, education and infrastructure, and empirically explore the effects of a country's global fractionalization and its local learning multiplier.

In addition to micro-founding an antagonism index that captures how local interaction affects the degree of antagonism between groups at the national level, this paper makes three more contributions. First, we model the link between a society's antagonism and its public goods. We start by postulating that an individual's valuation of public goods depends on how much antagonism she feels towards others in society. We then show that if public goods are financed by private contributions, their provision depends on a society's average antagonism. Second, to bring the theory to the data, we develop a global database of language use at the local level, and empirically analyze the effect of our new antagonism measure on public goods for virtually all countries of the world. Third, using an instrumental variable approach, we make causal statements about whether increased local learning improves or worsens outcomes. Identifying the causal effect of local interaction on the provision of public goods is relevant for the design of policies that have an impact on the degree of local learning. We now discuss these last two contributions in some more detail.

---

[3]See, e.g., Alesina et al. (2003), Alesina and La Ferrara (2000) and Alesina, Baqir and Hoxby (2004).

Before empirically investigating the effect of the spatial distribution of ethnolinguistic 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 (Lewis, Simons and Fennig, 2014). 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. We cross-validate our algorithm using census data for sub-Saharan Africa from Gershman and Rivera (2016). Once we have information on the speakers of different languages at a 5 km by 5 km resolution, we compute our measures of global fractionalization and the local learning multiplier for each country. The underlying assumption is that local interaction occurs within cells of 5 km by 5 km.[4]

We then explore the effect of global fractionalization and the local learning multiplier on a wide variety of public goods outcomes in health, education and infrastructure at the country level. We start with an OLS approach and find that global fractionalization has a negative association with public good provision, whereas the local learning multiplier has a positive association with public good provision. This suggests that the local learning multiplier mitigates the antagonism felt towards other groups, hence providing support to the contact theory hypothesis. The magnitudes of the effects are considerable. For example, a one standard deviation increase in the local learning multiplier lowers child mortality by 7.5 per thousand live births. To put this figure into perspective, in its effect on child mortality, a one standard deviation increase in the local learning multiplier is equivalent to a 62 percent increase in GDP per capita. As another example, a one standard deviation increase in the local learning multiplier lowers illiteracy by 6.2 percentage points. This corresponds to a standardized β of -30 percent, i.e., a one standard deviation increase in the local learning multiplier decreases the illiteracy rate by 30 percent of its standard deviation.

Reverse causality is a potential concern. In societies with poor public goods individuals from the same linguistic group may prefer to geographically cluster to support each other. If so, public goods would affect the spatial sorting of individuals of different groups. This does not necessarily imply that our regressions suffer from reverse causality since the local learning multiplier is not a measure of geographic mixing. Still, to address any remaining concerns, we develop an instrument based on a country's geographic distribution of linguistic groups as predicted by the language use of its closest neighbor. More particularly, for each cell in a given country, we determine the languages spoken in the closest cell in a neighboring country, and use that information to predict language use in the original cell. With this information, we compute a predicted local learning multiplier that serves as an instrument. 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 the local learning multiplier lowers child mortality by 16.4 per thousand live births, compared to 7.5 in the OLS analysis. Using the methodology developed by Conley et al. (2012), we show that violations to the exclusion

---

[4]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).

restriction do not qualitatively affect our results. This allows us to conclude that there is a causal positive effect of the local learning multiplier on the quality of public goods. Overall, contact theory trumps conflict theory.

Our paper contributes to four strands of the literature. First, a few papers have analyzed how the spatial distribution of diversity matters for different outcomes at the country level. 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, 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 empirical paper by Alesina and Zhuravskaya (2011) who find 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 interaction 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 84,000 cells. Finally, in a recent empirical study of sixteen African economies, Robinson (2013) shows that local level diversity increases inter-ethnic trust at the country level.

In contrast to our work, these papers do not provide a theoretical framework that conceptualizes how local interaction affects the antagonism that derives from a country's overall diversity. Instead, they measure the impact of local interaction by either using a segregation index or a local fractionalization index. We have already highlighted the meaningful differences between standard measures of segregation and the local learning multiplier: segregation is low whenever groups are geographically mixed, whereas the local learning multiplier is high whenever there are a few large groups that are geographically mixed. As compared with a local fractionalization index, the local learning multiplier also differs in a crucial aspect. The effect of local learning on an individual's antagonism depends on her pre-existing prejudice towards the out-group. This explains why local group shares and aggregate group shares are complementary, thus entering multiplicatively in our antagonism measure. In contrast, under local fractionalization an individual's change in antagonism is unrelated to his pre-existing prejudice towards the out-group. There is no complementarity, and local group shares enter additively in the local fractionalization measure. Since contact theory and conflict theory are about the effect of local interaction on an individual's pre-existing prejudice, local learning captures this in a way that local fractionalization does not.

Second, some studies look at the relation between diversity and different political economy outcomes at the local level. In the context of local public goods, examples include Alesina, Baqir and Easterly (1999) at the level of cities in the U.S.; Dahlberg et al. (2012) at the level of Swedish municipalities; Munshi and Rosenzweig (2015) at the level of wards in India; and Algan et al. (2016) at the level of apartment blocks in France. Also noteworthy is the paper by Montalvo and Reynal-Querol (2016) who analyze the effect of local diversity on local economic growth at a spatial resolution of one degree by one degree. Rather than analyzing the effect of local diversity on local outcomes, our paper explores how local interaction affects outcomes at

the national level. In particular, controlling for a country's total diversity, we study how an increase in local learning affects the support for public goods at the country level.

Third, a number of experimental papers have explored the effect of contact on prejudice. For example, Burns, Corno and La Ferrara (2015) show how living with a randomly assigned roommate from a different ethnicity at the University of Cape Town reduces prejudice, and Carrell et al. (2015) find that white males in the U.S. Air Force Academy are more accepting of blacks if they are exposed to a greater number of black peers. Similarly, Boisjoly et al. (2006) conclude that white students who are randomly assigned African American roommates are more likely to endorse affirmative action and to be more empathetic towards other groups. In an overview of this literature, Bertrand and Duflo (2016) conclude that the overwhelming evidence is in favor of contact theory.[5] Rather than focusing on small-scale micro-evidence of the relation between interaction and prejudice, we conduct a large-scale analysis of the causal effect of local interaction on public goods provision at the national level across the globe. In doing so, we confirm the findings of this experimental literature.

Fourth, several papers have constructed datasets of language use at the local level. Like us, they combine linguistic maps with information on the number of speakers at the country level. Weidman et al. (2010) rely on the Atlas Narodov Mira to create the GREG (Geo-Referencing of Ethnic Groups) database, whereas Matuszeski and Schneider (2006) use maps from the Ethnologue with the same goal.[6] A recurring issue is how to allocate the shares of different language speakers in areas where more than one language is spoken. Our paper tackles this problem by using an iterative proportional fitting algorithm, commonly used in statistics. Unlike the methods used in other papers that allocate language speakers to cells, this algorithm respects both the cell populations and the country-language population totals. Another approach is to use census and survey data. Gershman and Rivera (2016) follow this method to estimate local diversity measures for around 400 first-level administrative regions in 36 countries of sub-Saharan Africa. To cross-validate our algorithm, we compare our local diversity measures to the ones in Gershman and Rivera (2016). Focusing on the regions for which Gershman and Rivera (2016) use census data, we find correlations in local diversity of 0.80 at the regional level and 0.96 at the country level. These high correlations for sub-Saharan Africa give us confidence in our algorithm. Compared to using census or survey data, our approach has the important advantage that we can generate a consistent dataset of local language use for the entire world.

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.

---

[5]For a counter-example, an experiment by Enos (2014) who sent a small number of Spanish-speaking people to commuter trains in homogeneously white communities finds evidence in favor of conflict theory.

[6]One important difference is that the Ethnologue has information on 6,905 distinct languages, whereas the GREG data has 929 groups.

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 linguistic 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 others in 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.[7] In what follows we start by proposing a model of antagonism, and then relate it to the valuation of public goods.[8]

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 her own group. For now, where individuals live is assumed to be inconsequential; this is equivalent to assuming that local contact between individuals of different groups is irrelevant. 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 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 local 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

---

[7]Some of the robustness tests in the empirical part of the paper will control for overall income inequality and for income inequality between ethnic groups.

[8]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.

against that group.

Antagonism towards other groups: global fractionalization and local learning multiplier

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 the share $s_{\ell j}$ captures 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:

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

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 antagonism.

The average antagonism of all individuals of cell ℓ towards all 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

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

where the first term is the country's overall linguistic fractionalization and the second term is the probability that when an individual 1 is randomly matched with an individual 2 of his own cell and an individual 3

8

from the society at large, 2 and 3 belong to the same group and 1 belongs to a different group. The second term can be thought of as the average effect of local interaction on the antagonism experienced towards others in the society at large. That is, local learning about others can either increase or decrease the tension an individual feels towards others in the overall society. We call this second term LL the *local learning multiplier*.

We use the terminology of a multiplier when referring to LL because of the complementarity between the local group shares, $s_{\ell j}$, and the aggregate group shares, s_j. The intuition for this complementarity is easy to understand: how much an individual learns about other groups depends on the local shares; and how much that learning affects her antagonism towards those other groups in the society at large depends on the aggregate shares. That is, learning about another group locally can only have an important effect on antagonism if that other group is sizable in the aggregate. As we will now see, this complementarity is key when comparing the local learning multiplier to indices of segregation, polarization and fractionalization.

Relation with segregation

The local learning multiplier is different from standard indices of segregation used in the literature.[9] Segregation measures the extent to which cells differ from each other, whereas the local learning multiplier measures how local interaction affects antagonism in the society at large. To see the difference between both indices, take two countries, one with language shares $(1/3, 1/3, 1/3)$ and the other with language shares $(0.5, 1/6, 1/6, 1/6)$, so that ELF_{glob} is the same in both countries. There is perfect geographic mixing, so that the linguistic composition of each cell mirrors that of its country. As a result, segregation is zero in both countries. However, it is easy to show that the country with shares $(1/3, 1/3, 1/3)$ has a higher local learning multiplier than the country with shares $(0.5, 1/6, 1/6, 1/6)$.

The intuition is straightforward: learning about fewer, but bigger, groups is more useful than learning about more, but smaller, groups. Because of the complementarity between local group shares and aggregate group shares, local learning multipliers are bigger in societies with fewer, larger groups. As we will see in the data, this difference between segregation and the local learning multiplier is not just theoretically relevant. Controlling for global fractionalization, we find a correlation between the residuals of both indices of 0.54. As expected, the correlation is positive. However, it is far from perfect, implying that both concepts are empirically different.

Although segregation and local learning are different, when we take the aggregate linguistic composition of a country as given, segregation is minimized when local learning is maximized. This occurs when there is perfect spatial mixing.[10] 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 mirror that of the country. Again, this does not imply that there will be a lot of learning in countries that are

---

[9]For a survey, see Massey and Denton (1988), and for a recent application, see Alesina and Zhuravskaya, (2011).

[10]A formal proof is available upon request.

spatially mixed. As mentioned before, in highly fractionalized societies, with many small groups, the amount of learning will be limited.

Relation with polarization

The functional form of the local learning multiplier is related to that of a standard polarization index, though their micro-foundations are very different. The comparison between both indices is sharpest when focusing on a society that is spatially perfectly mixed. In that case $s_{\ell i} = s_i$ for all ℓ and all i, yielding a local learning multiplier equal to $\sum_i \sum_{j \neq i} s_i s_j^2$. Esteban and Ray (1994) micro-found an index of polarization by positing that the antagonism an individual experiences towards others increases in the size of his own group. The idea is that an individual feels greater antagonism if he identifies more strongly 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, $\sum_i \sum_{j \neq i} s_i^2 s_j$.[11]

As can be seen, in the case of perfect spatial mixing, the local learning multiplier is identical to an index of polarization, with one subtle difference in notation: in the polarization index the quadratic term shows up in the own-group share, whereas in the local learning multiplier the quadratic term shows up in the other-group share. Of course, since i and j are interchangeable, whether the quadratic term appears in the own group or in the other group is irrelevant for the final expression, so that both indices are identical in spatially mixed societies. Yet, the subtle difference in notation points to the relevant difference in micro-foundations between the two indices. In the polarization index, it is the identification with the own group that explains the quadratic term on the own-group share. Although we did not frame the local learning multiplier in terms of identification, we can use the same concept to interpret our index. Local interaction with another group leads to 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. Under that interpretation, identification with the own group underlies the polarization index, whereas identification with the other group underlies the local learning multiplier.

In addition to the different micro-foundations, one other difference between the two indices is important. Whereas polarization is commonly viewed as being detrimental to economic, political and social outcomes, local learning may have a benign effect by mitigating a society's antagonism if contact theory prevails. Hence, if we empirically find that local learning improves public goods outcomes, this would be consistent with an index that has micro-foundations based on contact theory, and it would not fit an index with micro-foundations related to polarization.

In spite of these parallels between local learning and polarization, their similarity should not be overstated. The equivalence in the functional forms of polarization and local learning only applies in the case of perfect spatial mixing. When groups are not perfectly mixed across a country's geography, the

---

[11]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).

equivalence breaks down.

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 inclusion of local fractionalization is to ignore the complementarity between local and global group shares by supposing 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

$$a_{\ell i j} = s_j + \beta s_{\ell j}, \tag{5}$$

so that the average antagonism experienced 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

$$
\begin{aligned}
A &= \left(1 - \sum_i s_i^2 \right) + \beta \sum_\ell s_\ell \left(1 - \sum_i s_{\ell i}^2 \right) \\
&= ELF_{glob} + \beta \sum_\ell s_\ell ELF_\ell \\
&= ELF_{glob} + \beta ELF_{loc}, \tag{6}
\end{aligned}
$$

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. These theories imply that local interaction can only have an effect on 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 is essentially zero.[12] However, expression (5) would imply that the antagonism experienced by individuals of group i and cell ℓ

---

[12]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 multiplier.

11

towards individuals of group j is β. That is, local fractionalization reduces antagonism even if there was no antagonism to start with. This happens because local fractionalization enters additively into the overall expression 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 by individuals of group i and cell ℓ towards the out-group would be zero.

2.2   Public Goods Provision

As shown in equation (1), 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

$$v_{\ell i} = \frac{\kappa_1}{\kappa_2 + a_{\ell i}}, \tag{7}$$

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.[13] 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

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

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 antagonism A. In particular,*

$$G = \frac{\kappa_1 N}{\kappa_1 + N(\kappa_2 + A)}. \tag{8}$$

Proof.   See Appendix A. ∎

For our empirical estimation it will be useful to linearly approximate (8). After dividing numerator and denominator by $\kappa_1 N$ and assuming that $1/N \approx 0$, we can write

$$G \approx \frac{\kappa_1}{\kappa_2 + A} \tag{9}$$

---

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

Assuming that κ_2 is sufficiently large and that A is sufficiently small, we can take a first-order Taylor approximation of (9) which yields

$$G \approx \frac{\kappa_1}{\kappa_2} - \frac{\kappa_1}{\kappa_2^2} A = \frac{\kappa_1}{\kappa_2} - \frac{\kappa_1}{\kappa_2^2} ELF_{glob} - \frac{\kappa_1}{\kappa_2^2} \beta LL. \tag{10}$$

This will serve as our estimating equation in the empirical part. From (10) we can conclude that the provision of public goods depends negatively on global fractionalization and either positively or negatively on the local learning multiplier (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.[14]

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, Ortuño-Ortín and Wacziarg (2012). The exact variables, their sources and summary statistics are given in Appendix B and Table B.1. 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 World Language Mapping System, the digitized version of the 17th edition of the Ethnologue. This gives us a polygon shapefile, where most of 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

---

[14]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.

is equivalent to having a polygon that consists of the entire country.[15] 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.[16] 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 $30''$. 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 according to the World Bank.

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.[17] We will later cross-validate the outcome of our algorithm with actual census data from Gershman and Rivera (2016).

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. 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 then 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,

$$\mathcal{T}^{(2n-1)}(\ell, i) = \frac{\mathcal{T}^{(2n-2)}(\ell, i)}{\sum_j \mathcal{T}^{(2n-2)}(\ell, j)} \mathcal{N}(\ell, 1), \tag{11}$$

---

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

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

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

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,

$$\mathcal{T}^{(2n)}(\ell, i) = \frac{\mathcal{T}^{(2n-1)}(\ell, i)}{\sum_k \mathcal{T}^{(2n-1)}(k, i)} \mathcal{L}(1, i) \tag{12}$$

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.

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 the local learning multiplier. 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 Walloon and Picard, two variations of French, be considered as two separate language groups or should they be aggregated into French? To address this issue, Desmet, Ortuño-Ortín and Wacziarg (2012) use the language tree of the Ethnologue to 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 Walloon and Picard are taken to be different groups. As Desmet, Ortuño-Ortín and Wacziarg (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.

In the case of public goods, Desmet, Ortuño-Ortín and Wacziarg (2012) find that intermediate levels of aggregation are most relevant. Hence, in our benchmark empirical analysis of public goods, we aggregate languages to level 5, though we also check the robustness of our results to using both higher and lower levels 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).

Using aggregation level 5, Figure 1 and Figure 2 show global fractionalization and the local learning multiplier by country.[18] As is immediately obvious, there are many differences between both indices. Compare, for example, Chad and the Central African Republic. Global fractionalization is much higher in Chad, but the local learning multiplier is much higher in the Central African Republic. Figure 3 also displays the local learning multiplier, but now for each grid cell separately. Consistent with the previous example, there are more areas of high local learning in the Central African Republic than in Chad.[19]

The left-hand panel of Figure 4 shows a scatterplot of both indices. Standing out is that countries with very high degrees of global fractionalization tend to have low local learning multipliers. The most fractionalized countries, such as Papua New Guinea (PNG), Mali (MLI) and Nigeria (NGA), have so many groups that local learning is limited, independently of how much the groups are locally mixed. Another relevant finding is the important heterogeneity in the degree of local learning between countries with inter-

---

[18]Table B.1 give summary statistics of the different indices and Table B.2 reports correlations between them.

[19]Appendix Figures C.1 through C.3 display the same maps, but for level 15 of aggregation. Compared to level 5, some of the countries in the central and the southern part of Africa now display much higher levels of diversity. For example, Zambia, where nearly everyone speaks a language of the Niger-Congo family, is much more diverse at level 15 than at level 2. Appendix Figures C.4 through C.6 also show maps for level 2 of aggregation.

Figure 1. Global ELF by Country – Level 5
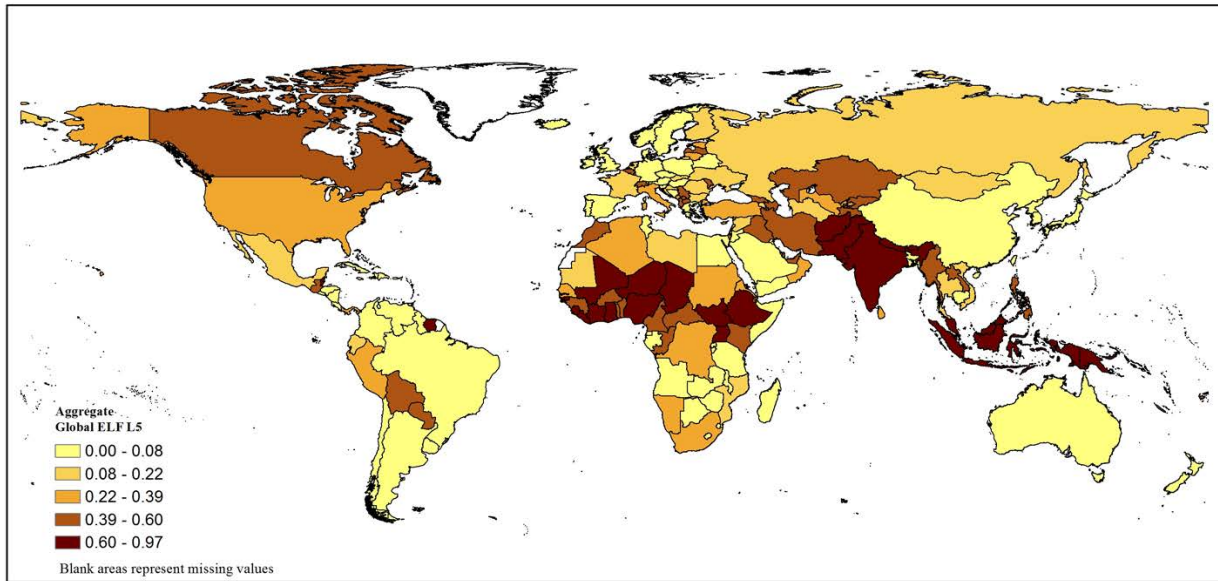
Figure 2. Local Learning by Country – Level 5
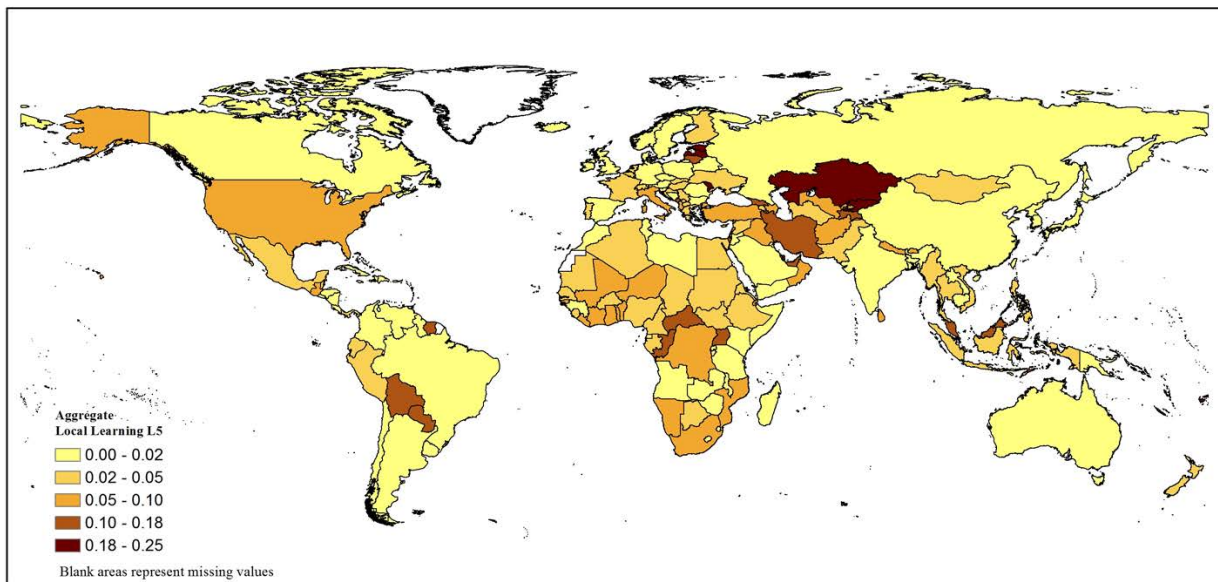
Figure 3. Local Learning at 5 km by 5 km Resolution – Level 5
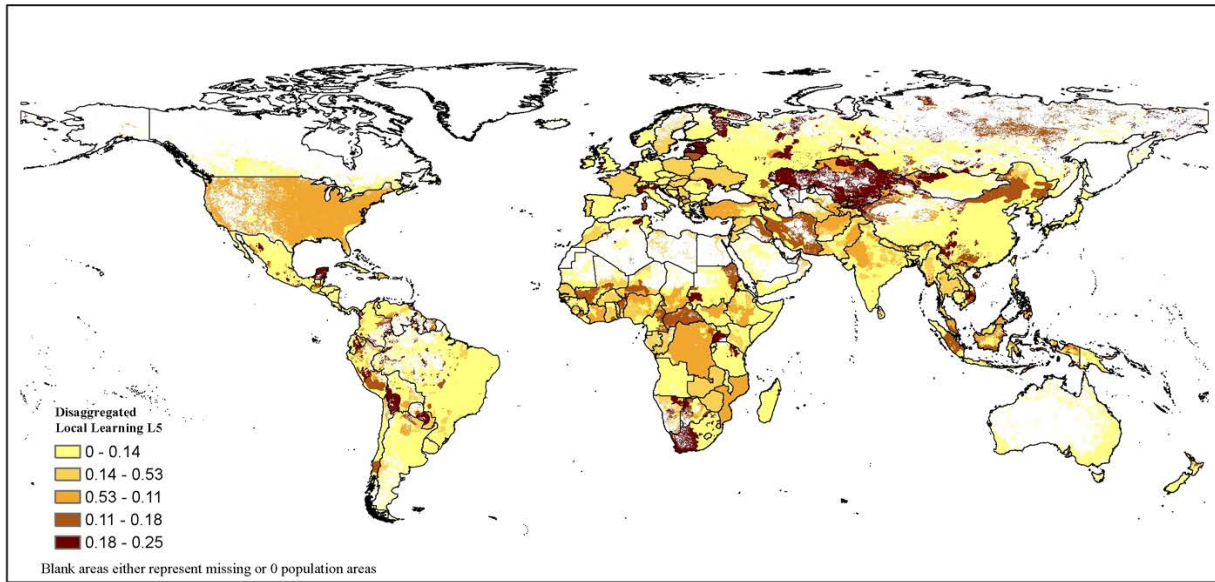
mediate levels of global fractionalization. Compare, for example, Guatemala (GTM) and Mauritius (MUS). In spite of overall diversity being essentially identical in Guatemala (0.53) and in Mauritius (0.52), local interaction is lower in Guatemala. The reason is that in Guatemala many of the speakers of indigenous languages live in the central and northwestern highlands, having limited contact with Spanish speakers. In contrast, according to Chiba (2006), Mauritians "switch languages according to the occasion in the way other people change clothes". As a result, the local learning multiplier is much higher in Mauritius (0.20) than in Guatemala (0.06).

Figure 4. Global Fractionalization, Local Learning Multiplier, and Segregation
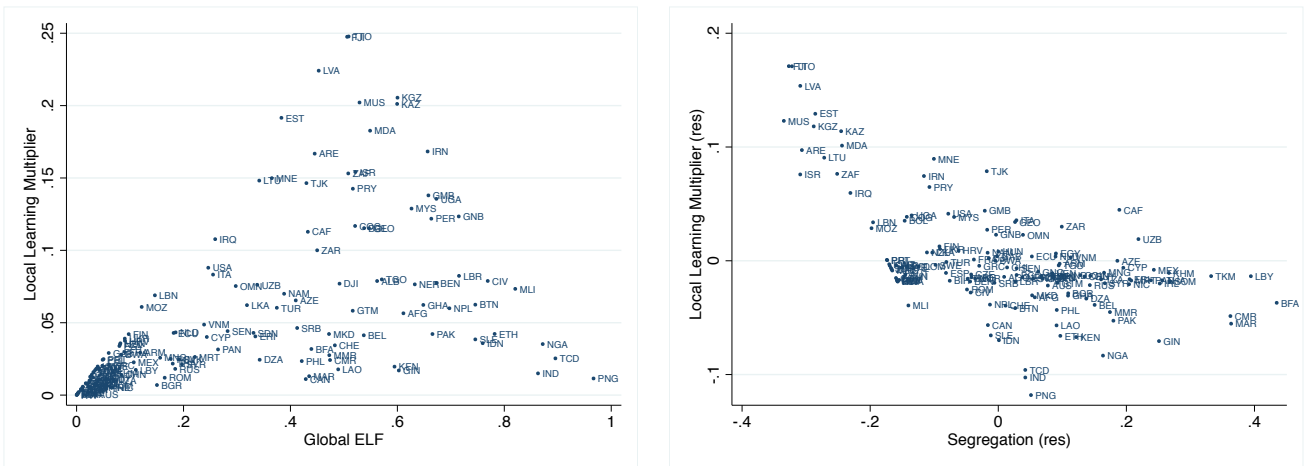
a. Global ELF and local learning multiplier   b. Segregation and local learning multiplier

In the theoretical section we already addressed the relation between the local learning multiplier

and segregation. The right-hand side panel of Figure 4 plots their empirical relation. For segregation, we compute the segregation index of Alesina and Zhuravskaya (2011) using our data at the 5 km by 5 km level. Comparing both indices is cleaner when controlling for global fractionalization, so the scatterplot represents residuals of regressions on global fractionalization. As expected, local learning tends to be greater in less segregated societies. However, the correlation is far from perfect, standing at 0.54. This suggests that there are important differences between both indices. Mali (MLI) is a case in point: there is a fair amount of local mixing, so that segregation is relatively low. However, because the country is made up of many different groups, the amount of local learning is limited.

3.4   Cross-Validation

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 at the level of first-level administrative regions. The local fractionalization measures in Gershman and Rivera (2016) that are based on census data are arguably of higher quality than those that are based on survey data. If we limit ourselves to first-level administrative regions for which local fractionalization comes from census data, the correlation with our index increases to 0.80. Since our empirical analysis uses population-weighted local indices at the country level, it is useful to compare country-level local fractionalization indices. When doing so, the correlation between our measure and the one based on Gershman and Rivera (2016) further increases to 0.96 for countries with census data. It is encouraging to find such high correlations, in spite of the differences in the underlying data on linguistic groups.[20] This cross-validation gives us confidence in the algorithm we use to allocate language speakers to cells.

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. Equation (10) from the theory provides the following estimating

---

[20]The Ethnologue is much more detailed in the number of linguistic groups than Gershman and Rivera (2016). For example, whereas Mali has 54 unique linguistic groups in the Ethnologue, it only has 9 unique groups in the Gershman and Rivera (2016) data.

equation:

$$G_c = \gamma_0 X_c + \gamma_1 ELF_{glob,c} + \gamma_2 LL_c + \varepsilon_c, \tag{13}$$

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 multiplier of country c, X_c are a set of additional controls, ε_c is an error term, $\gamma_0 = \kappa_1/\kappa_2$, $\gamma_1 = -\kappa_1/\kappa_2^2$, and $\gamma_2 = -(\kappa_1/\kappa_2^2)\beta$. Our main coefficient of interest will be γ_2: if its sign is consistent with an improvement in public goods, we will interpret this as implying that local learning reduces antagonism towards other groups, making individuals more willing to contribute to public goods. 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 the local learning multiplier. The coefficients switch signs, suggesting that a higher local learning multiplier is associated with less child mortality, but the coefficients are not statistically significant at the 10% level.

Motivated by the theory, in the last two columns we include simultaneously global fractionalization and the local learning multiplier. As before, countries with higher levels of global fractionalization continue to have greater rates of child mortality. In fact, the coefficients are larger in magnitude. Focusing on column (6), we find that a one standard deviation increase in global fractionalization raises child mortality by 11.3 per thousand live births. More interestingly, the local learning multiplier 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 substantial. Using once again column (6) as our preferred specification, a one standard deviation increase in the local learning multiplier lowers child mortality by 7.5 per thousand

Table 1. Child Mortality: Global ELF and Local Learning

	Child mortality					
	(1)	(2)	(3)	(4)	(5)	(6)
Global ELF	39.221***	23.991***			69.698***	42.573***
	(10.548)	(8.713)			(13.235)	(10.720)
Local Learning Multiplier			-35.929	-17.299	-214.470***	-124.861***
			(38.870)	(30.606)	(48.706)	(38.912)
Absolute Latitude	-0.713***	-0.139	-0.874***	-0.134	-0.646***	-0.116
	(0.180)	(0.203)	(0.169)	(0.203)	(0.170)	(0.197)
Latin America & Carib.	-12.372*	-5.362	-17.783**	-7.583	-10.189	-4.235
	(6.960)	(5.385)	(6.886)	(5.343)	(6.510)	(5.502)
Sub-Saharan Africa	77.872***	59.453***	78.650***	59.418***	74.768***	59.187***
	(9.403)	(8.352)	(9.696)	(8.553)	(8.642)	(7.882)
East and S.E. Asia	-1.009	-3.492	-1.671	-3.360	-7.234	-7.140
	(9.989)	(8.162)	(10.121)	(8.253)	(9.506)	(8.104)
Log GDP per Capita		-16.758***		-17.918***		-15.589***
		(1.959)		(2.001)		(1.852)
French Legal Origin		-17.976***		-13.925**		-13.244**
		(5.988)		(5.608)		(5.212)
German Legal Origin		-23.967***		-22.743***		-19.595***
		(4.723)		(4.762)		(4.188)
UK Legal Origin		-20.610***		-15.833***		-16.230***
		(5.906)		(5.708)		(5.402)
Constant	46.918***	191.137***	64.462***	204.318***	48.929***	178.437***
	(9.093)	(19.086)	(8.854)	(18.766)	(8.516)	(17.456)
Observations	178	171	178	171	178	171
R^2	0.685	0.815	0.655	0.804	0.719	0.825

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 Multiplier 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 B.

live births. To put this figure into perspective, in its effect on child mortality, a one standard deviation increase in the local learning multiplier has the same effect on child mortality as a 62% 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 accumulate knowledge about other groups 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 the local learning multiplier 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 the local learning multiplier are always statistically significant at the 5% level, with the

exception of road density. In terms of magnitude, a one standard deviation increase in the local learning multiplier increases the rate of measles immunization by 4.0 percentage points. The corresponding standardized β is 26%, meaning that a one standard deviation increase in the local learning multiplier increases the measles immunization rate by 26% of its standard deviation. In the case of illiteracy, the standardized β is -30%. These results are further evidence in favor of contact theory.

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	-25.034***	-1.246**	31.697***	-0.244*	-25.065***	1.146
	(4.563)	(0.612)	(5.945)	(0.137)	(6.177)	(1.759)
Local Learning Multiplier	66.700***	7.037**	-102.057***	1.369***	81.892***	5.387
	(15.851)	(3.042)	(22.085)	(0.493)	(24.820)	(12.411)
Absolute Latitude	0.197**	0.101***	-0.301**	0.008***	0.136	0.173***
	(0.087)	(0.019)	(0.117)	(0.003)	(0.161)	(0.065)
Latin America & Carib.	4.500*	0.311	-6.256*	0.211**	-4.755	2.332
	(2.672)	(0.450)	(3.434)	(0.089)	(4.781)	(1.770)
Sub-Saharan Africa	-9.335***	0.119	8.128*	-0.125	-25.303***	2.958*
	(2.973)	(0.470)	(4.163)	(0.109)	(5.206)	(1.779)
East and S.E. Asia	2.434	0.983	-14.255***	0.085	-6.498	-1.298
	(3.777)	(0.903)	(4.778)	(0.102)	(6.484)	(2.124)
Log GDP per Capita	1.384*	0.257*	-4.348***	0.117***	10.189***	1.801***
	(0.831)	(0.155)	(1.137)	(0.022)	(1.279)	(0.452)
French Legal Origin	2.192	0.491	3.917	0.077	13.410***	-14.003*
	(2.949)	(1.448)	(3.495)	(0.080)	(4.079)	(7.885)
German Legal Origin	3.260	2.323		0.235***	13.573***	-11.608
	(2.765)	(1.483)		(0.064)	(3.585)	(7.914)
UK Legal Origin	7.458**	0.529	0.171	0.251***	10.584***	-10.932
	(3.302)	(1.454)	(4.142)	(0.082)	(4.038)	(8.094)
Constant	66.595***	-2.049	53.873***	0.647**	-17.277	-0.657
	(8.819)	(1.962)	(10.918)	(0.256)	(14.158)	(9.531)
Observations	171	173	138	136	171	172
R^2	0.572	0.597	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 Multiplier 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 B.

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 that 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 the local learning multiplier continues to have a benign effect, with both coefficients being statistically significant at the 1% level.

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	41.888***	23.267	42.059***	24.922	48.944***	34.830**
	(12.288)	(15.098)	(14.144)	(17.932)	(13.711)	(14.556)
Local Learning Multiplier	-144.832***	-122.398**	-151.335***	-135.434**		-183.511**
	(49.096)	(54.170)	(54.920)	(64.481)		(85.957)
Local ELF					-47.828**	31.572
					(18.856)	(42.139)
Decentralization[1]	-0.138	0.956	-3.336	-1.895		
	(5.278)	(4.882)	(5.588)	(6.856)		
Decentralization[1] x Global ELF			0.415	4.769		
			(28.247)	(30.509)		
Decentralization[1] x Local Learning			80.245	81.601		
			(140.199)	(138.609)		
Absolute Latitude	-0.115	-0.176	-0.086	-0.166	-0.146	-0.101
	(0.200)	(0.197)	(0.212)	(0.205)	(0.199)	(0.200)
Latin America & Carib.	-4.137	-5.530	-3.183	-4.689	-4.339	-4.381
	(6.389)	(6.954)	(6.553)	(7.191)	(5.461)	(5.487)
Sub-Saharan Africa	61.305***	57.497***	61.806***	57.085***	59.323***	59.148***
	(9.102)	(9.707)	(9.156)	(9.877)	(8.099)	(7.848)
East and S.E. Asia	-7.278	-6.730	-7.720	-7.803	-5.473	-7.545
	(8.909)	(8.716)	(9.040)	(9.061)	(8.072)	(8.179)
Log GDP per Capita	-15.899***	-16.829***	-16.021***	-17.002***	-15.930***	-15.587***
	(1.787)	(1.829)	(1.817)	(1.857)	(1.853)	(1.868)
French Legal Origin	-12.351**	-13.567**	-11.645*	-13.543**	-16.067***	-12.281**
	(5.993)	(6.172)	(6.103)	(6.168)	(5.457)	(5.476)
German Legal Origin	-19.460***	-20.811***	-18.677***	-20.214***	-22.047***	-18.809***
	(4.701)	(4.633)	(4.749)	(4.682)	(4.217)	(4.442)
UK Legal Origin	-15.868**	-14.043**	-15.512**	-14.156**	-19.178***	-15.118***
	(6.173)	(6.034)	(6.306)	(6.063)	(5.581)	(5.531)
Constant	181.963***	195.991***	181.571***	197.336***	183.260***	177.671***
	(20.276)	(20.039)	(20.401)	(20.148)	(17.735)	(17.546)
Observations	149	122	149	122	171	171
R^2	0.849	0.848	0.850	0.849	0.821	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 Multiplier 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 B.

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., whether 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 decentraliza-

tion and global fractionalization and the local learning multiplier. 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, the local learning multiplier 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 the local learning multiplier 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, the local learning multiplier trumps local fractionalization.

Different levels of linguistic and geographic aggregation. Desmet, Ortuño-Ortín and Wacziarg (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 Walloon and Picard, are considered to be different languages, whereas at level 2 all Romance languages, such as French and Spanish, 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 the local learning multiplier, from 1% to 5%.

Our regressions so far are based on local learning that takes 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) in Table 4 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 the local learning multiplier lowered child mortality by 7.5 per thousand live births, it now lowers it by 7.8 per thousand live births.

Income inequality, linguistic inequality and segregation. 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 the local learning multiplier 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

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

	(1) Level 15	(2) Level 2	(3) 10km x 10km	(4) Gini	(5) Linguistic Inequality	(6) Linguistic Segregation
Global ELF	30.154***	37.885**	43.074***	48.115***	40.184***	35.494***
	(9.238)	(15.410)	(10.595)	(11.287)	(12.099)	(13.169)
Local Learning Multiplier	-83.219**	-90.960**	-128.141***	-165.075***	-113.338***	-109.944**
	(38.757)	(45.239)	(38.250)	(53.191)	(40.827)	(46.750)
Gini				-0.036		
				(0.442)		
Linguistic Inequality					10.531	
					(8.378)	
Linguistic Segregation						9.291
						(13.291)
Absolute Latitude	-0.015	-0.137	-0.118	0.080	0.028	-0.060
	(0.214)	(0.204)	(0.196)	(0.254)	(0.207)	(0.206)
Latin America & Carib.	0.292	-7.280	-4.198	-0.257	-2.043	-3.087
	(6.117)	(5.461)	(5.442)	(9.362)	(5.897)	(5.820)
Sub-Saharan Africa	56.759***	59.321***	59.105***	66.139***	62.983***	62.331***
	(8.247)	(8.412)	(7.847)	(9.868)	(7.958)	(8.078)
East and S.E. Asia	-5.380	-4.927	-7.057	-3.577	-8.278	-9.591
	(8.750)	(7.876)	(8.171)	(9.102)	(8.284)	(8.953)
Log GDP per Capita	-15.674***	-16.810***	-15.513***	-17.004***	-14.945***	-15.715***
	(1.820)	(1.913)	(1.850)	(2.316)	(1.897)	(1.910)
French Legal Origin	-12.739**	-14.424**	-13.091**	-14.908**	-8.356	-11.672**
	(5.461)	(5.668)	(5.201)	(6.084)	(6.071)	(5.146)
German Legal Origin	-18.853***	-22.418***	-19.670***	-19.544***	-14.420***	-17.840***
	(4.744)	(4.474)	(4.169)	(4.730)	(5.267)	(4.304)
UK Legal origin	-15.993***	-16.821***	-16.166***	-15.372**	-10.705	-14.619***
	(5.956)	(5.878)	(5.397)	(6.913)	(6.534)	(5.562)
Constant	172.810***	192.504***	177.830***	182.767***	158.994***	174.944***
	(18.404)	(18.255)	(17.404)	(21.897)	(19.528)	(17.820)
Observations	171	171	171	124	165	161
R^2	0.819	0.812	0.826	0.842	0.827	0.833

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 Multiplier variables are measured at level 5 of aggregation, except in columns 1 and 2 where they are measured at, respectively, levels 15 and 2 of aggregation. In column 3, we use the Local Learning Multiplier 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, in column 5 we control for income inequality between linguistic groups, and in column 6 we control for linguistic segregation. The variable definitions and data sources for each of the variables are provided in Appendix B.

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 the local learning multiplier are very similar to our baseline regression.

As discussed in our theoretical section, though our local learning multiplier is related to the degree of local mixing, it differs from segregation in an important way. In particular, the local learning multiplier tends to be high when there is spatial mixing *and* there are relatively few large-sized groups, whereas the degree of segregation only depends on the degree of local mixing. To further explore the relative importance of both indices, we use our data to calculate the linguistic segregation measure of Alesina and Zhuravskaya (2011), and include it jointly with the local learning multiplier in our regression. Table 4, column (6), reports the results. Our coefficients of interest do not change, while the segregation variable itself is not statistically significant. From this we conclude that the local learning multiplier captures a relevant determinant of public goods provision in a way that segregation does not.

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: the local learning multiplier 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 the local learning multiplier 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 a bigger local learning multiplier 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 the local learning multiplier 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	28.186***	46.351***	48.739***	40.303***	40.505***	58.777***	39.275***
	(9.254)	(11.893)	(11.691)	(10.807)	(11.545)	(13.298)	(10.413)
Local Learning Multiplier	-70.061**	-126.716***	-159.589***	-107.179**	-112.393***	-195.534***	-138.121***
	(27.660)	(40.668)	(45.200)	(44.350)	(42.105)	(48.935)	(37.835)
Absolute Latitude	-0.035	-0.157	-0.166	-0.077	-0.068	-0.800***	-0.032
	(0.173)	(0.201)	(0.221)	(0.200)	(0.236)	(0.292)	(0.264)
Log GDP per Capita	-14.398***	-15.392***	-15.121***	-15.597***	-15.295***		-14.101***
	(1.616)	(1.921)	(1.926)	(2.006)	(2.062)		(2.252)
Soil Fertility						1.790	-16.133**
						(7.463)	(6.944)
Roughness						-43.000	10.334
						(28.485)	(30.316)
Elevation						15.793*	-2.764
						(7.997)	(9.338)
Island						2.832	2.858
						(10.051)	(8.643)
Landlocked						18.086***	13.708**
						(6.539)	(6.221)
Log Population							1.168
							(1.354)
Constant	160.425***	178.738***	176.836***	163.803***	181.942***	64.274***	125.271***
	(16.759)	(18.172)	(18.190)	(15.170)	(25.033)	(24.155)	(30.805)
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	0.687	0.832	0.826	0.826	0.826	0.809	0.868

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 Multiplier 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 B.

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	-26.074***	-1.038	31.271***	-0.276**	-29.751***	2.044
	(4.992)	(0.642)	(6.626)	(0.127)	(6.726)	(2.015)
Local Learning Multiplier	61.983***	8.237**	-113.923***	1.451***	98.518***	-1.128
	(20.909)	(3.725)	(25.651)	(0.520)	(31.794)	(15.601)
Log GDP per Capita	0.897	0.245	-6.451***	0.153***	9.254***	1.982***
	(1.025)	(0.176)	(1.527)	(0.022)	(1.481)	(0.624)
Absolute Latitude	0.080	0.082***	-0.090	0.000	-0.038	0.138
	(0.115)	(0.025)	(0.166)	(0.003)	(0.170)	(0.107)
Soil Fertility	3.587	0.940	-9.472**	0.315***	2.197	2.118
	(2.977)	(0.595)	(4.610)	(0.078)	(4.196)	(2.431)
Roughness	-12.824	-0.047	1.743	-0.588**	26.825	-24.410***
	(12.409)	(2.042)	(16.160)	(0.288)	(18.806)	(7.714)
Elevation	1.584	-0.608	-4.930	0.203**	-6.554	5.481***
	(3.446)	(0.609)	(4.739)	(0.088)	(5.632)	(1.799)
Island	-7.659*	1.038	-2.186	0.042	-7.494	3.468
	(4.023)	(0.787)	(6.011)	(0.083)	(5.559)	(2.245)
Landlocked	-1.311	1.072**	-1.973	-0.082	2.885	-2.661*
	(2.548)	(0.474)	(3.737)	(0.069)	(3.423)	(1.410)
Log Population	-0.889	-0.064	0.118	-0.024	-0.410	-1.244***
	(0.640)	(0.124)	(0.871)	(0.016)	(1.041)	(0.414)
Constant	98.263***	0.941	60.389***	1.395***	11.148	21.100*
	(15.741)	(3.732)	(19.651)	(0.377)	(23.928)	(12.029)
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	0.632	0.677	0.709	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 Multiplier 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 B.

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 potential 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.

Does this link between public goods provision and spatial mixing matter for the local learning multiplier? We address this question in three ways. First, recall that the local learning multiplier is not a measure of the degree of geographic mixing of different ethnolinguistic groups. Since there is a fair amount

of variation in the local learning multiplier for countries with similar levels of spatial mixing, this somewhat weakens the potential endogeneity concern. Second, we consider the empirical evidence on whether spatial sorting by ethnolinguistic groups 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 diversity at the local level is very stable over time. In particular, the correlation in local diversity over two to three decades exceeds 0.95. More importantly for our purpose, their results indicate that the small changes in subnational diversity are not related to regional economic performance. This also alleviates the endogeneity concern. Third, in what follows we design an instrumental variable strategy, with the aim of causally establishing the relation between the local learning multiplier and public goods.

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. We therefore want to predict $\mathcal{B}$, our K x M binary matrix of language use by cell, in a way that is independent of any spatial sorting. Once we have a predicted matrix of $\mathcal{B}$, we can apply our iterative proportional fitting algorithm to get an instrument for the local learning multiplier.

To create a predicted measure of local language use, we follow an approach similar to the one in Alesina and Zhuravskaya (2011) 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 a predicted measure of local learning.[21] For this to be a valid instrument, the exclusion restriction requires that the spatial distribution of languages in the closest cells of the neighboring countries does not directly affect the public goods outcomes in the own country. We will discuss the plausibility of the exclusion restriction at length at the end of this section.

IV results. Table 7 reports the IV results for our baseline regression. For five of the seven public goods, the local learning multiplier 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 instrument is 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 the local learning multiplier lowered child mortality by 7.5 per thousand in the OLS regression, it now lowers child mortality by 16.4 per thousand. The corresponding

---

[21]Appendix D provides a simple example of how neighboring countries are used to predict a country's geographic distribution of languages.

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	64.453***	-29.824***	-0.400	40.948***	-0.360**	-30.904***	4.326**
	(13.435)	(5.317)	(0.804)	(7.208)	(0.167)	(7.433)	(2.003)
Local Learning Multiplier	-271.882***	98.887***	1.368	-164.212***	2.107**	120.844***	-15.960
	(73.983)	(27.932)	(4.253)	(42.787)	(0.964)	(41.819)	(12.495)
Absolute Latitude	-0.089	0.192**	0.102***	-0.249**	0.008***	0.130	0.177***
	(0.204)	(0.086)	(0.019)	(0.121)	(0.003)	(0.159)	(0.062)
Latin America & Carib.	-2.908	4.209	0.345	-5.263	0.212**	-5.121	2.511
	(6.154)	(2.734)	(0.430)	(3.569)	(0.087)	(4.830)	(1.687)
Sub-Saharan Africa	58.875***	-9.267***	0.096	8.990**	-0.128	-25.216***	2.908*
	(7.448)	(2.876)	(0.458)	(4.068)	(0.103)	(5.038)	(1.691)
East and S.E. Asia	-11.434	3.374	0.809	-15.138***	0.107	-5.347	-1.924
	(7.894)	(3.634)	(0.860)	(4.893)	(0.101)	(6.378)	(2.093)
Log GDP per Capita	-14.214***	1.083	0.307**	-3.548***	0.111***	9.810***	1.998***
	(1.904)	(0.835)	(0.154)	(1.135)	(0.022)	(1.253)	(0.447)
French Legal Origin	-7.672	0.972	0.693	3.942	0.050	11.930***	-13.203*
	(5.985)	(2.962)	(1.415)	(2.681)	(0.087)	(4.184)	(7.599)
German Legal Origin	-14.447***	2.133	2.513*	-0.076	0.211***	12.200***	-10.865
	(5.038)	(2.798)	(1.438)	(4.260)	(0.073)	(3.764)	(7.737)
UK Legal Origin	-11.073*	6.329**	0.712	0.000	0.232***	9.222**	-10.185
	(5.915)	(3.204)	(1.426)	(.)	(0.084)	(4.116)	(7.787)
Constant	163.483***	69.868***	-2.579	46.933***	0.712***	-13.221	-2.798
	(18.745)	(8.960)	(1.960)	(11.373)	(0.258)	(14.085)	(9.148)
First-Stage F-Statistic	60.885	60.885	61.604	41.292	36.568	63.386	60-943
Observations	171	171	173	138	136	171	172
R^2	0.810	0.562	0.589	0.646	0.670	0.780	0.445

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 Multiplier 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 B.

standardized β is -29%.

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 the local learning multiplier become stronger. A one-standard deviation increase in the local learning multiplier lowers child mortality by an estimated 19.2 per thousand, with a corresponding standardized β of -33%. From this we can conclude that there is a causal effect between a higher local learning multiplier and improved public goods outcomes. The evidence hence overwhelmingly supports contact theory.

Plausibility of the exclusion restriction. We now discuss the plausibility of our identification strategy which relies on using predicted local learning as an instrument for observed local learning. For the exclusion restriction to hold, the predicted local learning multiplier must not have a direct effect on public goods

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	66.432***	-32.883***	0.222	47.484***	-0.344*	-39.049***	8.439***
	(13.729)	(6.292)	(1.037)	(9.072)	(0.180)	(9.515)	(3.016)
Local Learning Multiplier	-351.169***	115.402***	-1.651	-235.940***	1.932	171.615**	-51.297**
	(89.006)	(43.620)	(7.496)	(67.851)	(1.316)	(68.807)	(22.084)
Log GDP per Capita	-12.996***	0.620	0.297*	-5.451***	0.151***	8.911***	2.243***
	(2.324)	(1.017)	(0.166)	(1.567)	(0.022)	(1.449)	(0.627)
Absolute Latitude	0.057	0.058	0.086***	0.007	0.000	-0.069	0.159
	(0.264)	(0.111)	(0.024)	(0.183)	(0.003)	(0.167)	(0.099)
Soil Fertility	-15.957**	3.543	0.948*	-9.151**	0.316***	2.197	2.159
	(6.804)	(2.837)	(0.565)	(4.513)	(0.072)	(4.108)	(2.253)
Roughness	-12.517	-7.095	-1.107	-13.487	-0.535*	34.978*	-29.791***
	(31.618)	(13.156)	(2.130)	(17.791)	(0.308)	(19.036)	(8.222)
Elevation	3.417	0.035	-0.321	-0.934	0.190**	-8.727	6.936***
	(9.307)	(3.494)	(0.655)	(4.933)	(0.088)	(5.333)	(1.936)
Island	2.877	-7.664**	1.039	-0.401	0.042	-7.517	3.472*
	(7.916)	(3.736)	(0.745)	(6.184)	(0.077)	(5.149)	(2.037)
Landlocked	11.743*	-0.819	0.981**	-2.704	-0.082	3.568	-3.123**
	(6.374)	(2.509)	(0.469)	(3.587)	(0.064)	(3.464)	(1.546)
Log Population	-0.644	-0.435	-0.148	-0.820	-0.020	0.231	-1.671***
	(1.592)	(0.707)	(0.136)	(0.916)	(0.016)	(1.110)	(0.518)
Constant	138.714***	94.892***	1.565	66.623***	1.375***	5.912	24.265**
	(31.504)	(15.570)	(3.618)	(22.155)	(0.345)	(23.192)	(11.738)
Legal Origins	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-Stage F-Statistic	29.078	29.078	29.078	20.204	21.688	28.456	29.078
Observations	147	147	147	121	126	145	147
R^2	0.847	0.614	0.661	0.655	0.737	0.813	0.470

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 Multiplier 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 B.

over and above its indirect effect through the observed local learning multiplier. This exclusion restriction might be violated if public goods in the neighboring countries have a direct effect on public goods in the home country. To see why such a situation might be problematic for our identification strategy, recall that the spatial distribution of languages in the neighboring countries affects both the predicted local learning multiplier in the home country and public goods in the neighboring countries. The presence of cross-border spillovers in public goods might then lead to a direct relation between the predicted local learning multiplier and public goods in the home country. For example, if higher immunization rates in the neighboring countries lowers the spread of infectious diseases in the broader region, then any drop in child mortality in the neighboring countries would extend to the home country. If there is a strong correlation between the predicted local learning multiplier in the home country and the observed local learning multiplier in the neighboring countries, such spillovers could imply a violation of the exclusion restriction. To address this

potential concern, we adopt a three-pronged strategy.

First, we clarify that we do not directly use the local learning multiplier of the neighboring countries to predict the local learning index of the home country. Instead, for each cell in the home country we only use information on which languages are spoken in the closest cell of its neighboring countries to predict which languages are spoken in the home country cell. This procedure yields a predicted binary matrix of language use for all cells in the home country, which is then used to construct a predicted local learning multiplier. Hence, the predicted local learning multiplier of the home country may be very different from the local learning multipliers of its neighbors. In fact, the correlation between both measures is essentially zero (to be precise, between 0.03 and 0.06, depending on how we average the neighbors' local learning multipliers). With such a low correlation, it would be highly improbable that cross-border spillovers in public goods would invalidate the exclusion restriction.

Second, to be exhaustive, we can of course still investigate whether the neighbors' local learning multiplier, global fractionalization and public goods affect outcomes. When controlling for these additional variables in our otherwise identical IV specification of Table 8, the results are unchanged.[22] In particular, in Table 9 the local learning multiplier continues to have a significant and robust effect on the different public goods. The magnitudes of the effects are very similar to those found in the specification without the additional controls. This confirms our prior conclusion that the possible presence of cross-border spillovers in public goods does not undermine our identification strategy. As an alternative specification, we could separately control for, on the one hand, public goods, and on the other hand, local learning and global fractionalization. Doing so does not substantially alter the magnitudes of the original effects of the local learning multiplier.[23]

Third, while in the previous paragraphs we have argued that our identification strategy is likely to hold, next we check for the sensitivity of our results to small violations of the exclusion restriction. To that end, we follow the method outlined in Conley et al. (2012). Consider the equation we are interested in estimating, but now allow the instrument to have a direct impact on the outcome. That is, $G_c = \gamma_0 X_c + \gamma_1 ELF_{glob,c} + \gamma_2 LL_c + \gamma_3 LL_{nc} + \varepsilon_c$, where LL_{nc} is the instrument. The exclusion restriction assumes that $\gamma_3 = 0$. Conley et al. (2012) provide a procedure that allows for inference even if γ_3 is not exactly equal to zero. The idea is intuitive. If we were to know the true value of γ_3, it would be straightforward to estimate the above equation by 2SLS using LL_{nc} as an instrument for LL_c.[24] Of course, we do not know the true value of γ_3. Conley et al. (2012) tackle this problem by making assumptions on the support of γ_3, and then run 2SLS on different possible values of γ_3. Following their *Local to Zero* approach, we assume that $\gamma_3 \sim N(0, \delta^2)$. Using their methodology, we can obtain a 95% confidence interval for our coefficient of

---

[22]Neighboring country variables are computed as population-weighted averages of the neighboring countries' variables. Results do not change when either taking simple averages or using the populations in the home country cells that are exposed to each neighboring country as weights.

[23]Results are available upon request.

[24]This requires first subtracting $\gamma_3 LL_{nc}$ from both sides of the above equation.

Table 9. Neighbors' Global ELF, Local Learning and Public Goods, 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	61.371***	-30.656***	-0.084	40.548***	-0.346*	-38.103***	6.089*
	(13.634)	(6.888)	(1.105)	(9.041)	(0.183)	(10.341)	(3.317)
Local Learning Multiplier	-314.309***	105.024**	-0.886	-211.004***	1.798	168.179**	-41.570**
	(81.949)	(46.363)	(7.414)	(60.452)	(1.280)	(68.696)	(20.454)
Neighbors Global ELF	-26.123	5.872	0.320	0.018	0.141	2.503	3.592
	(17.175)	(6.482)	(1.025)	(10.154)	(0.168)	(9.931)	(4.998)
Neighbors Local Learning	-53.084	54.277*	-1.789	-71.297	0.034	33.850	-24.927
	(99.801)	(32.644)	(6.297)	(57.609)	(1.085)	(63.486)	(26.124)
Neighbors Public Goods	0.325***	0.080	0.352***	0.273**	0.090	-0.005	0.136
	(0.125)	(0.120)	(0.124)	(0.121)	(0.097)	(0.101)	(0.152)
Controls Table 8	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-Stage F-Statistic	31.922	27.839	27.586	20.045	22.119	34.871	27.949
Observations	146	146	146	121	125	144	146
R^2	0.864	0.634	0.688	0.700	0.745	0.813	0.444

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 Multiplier variables are measured at level 5 of aggregation and are based on the authors' calculations. The Local Learning Multiplier 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 B.

interest, γ_2.

How much this violation of the exclusion restriction affects our results for local learning depends of course on the magnitude of δ. Table 10 presents the results when we allow the direct effect of the instrument to be up to 10% to 50% of the marginal effect of the local learning multiplier on public goods from the IV estimation. To interpret our findings, consider, for example, the row which says CI ($2\delta = 15\%$). It reports the 95% confidence interval for the local learning coefficient when we allow the direct effect of the instrument to be up to 15% of the marginal effect of local learning from the original instrumental variable specification. When comparing the results with our original findings, reported in the upper panel of Table 10, we notice that the confidence intervals do not change the significance of our original results. Even if we allowed for a rather strong violation of the exclusion restriction so that the direct effect of the instrument were up to 25% of the marginal effect of the local learning multiplier, our findings continue to be statistically significant.

5   Concluding Remarks

In this paper we developed a model of antagonism which suggests that both a society's overall linguistic diversity and its local learning about other groups should affect the provision of public goods. Theoretically, we showed that local learning tends to be high when there are a few large-sized groups that are spatially mixed. This makes local learning different from both segregation and local diversity. Empirically, we found that overall linguistic diversity worsens the provision of public goods, whereas the opposite is true for the local learning multiplier. To assuage concerns about reverse causality, we developed an instrumental variable

Table 10. Plausibly Exogenous IV Estimation, Comprehensive Specification

	(1) Child Mortality	(2) Measles Immunization	(3) Hospital Beds	(4) Illiteracy Rate	(5) Schooling	(6) Improved Sanitation	(7) Road Density
Panel A: IV							
Local Learning	-351.169***	115.402***	-1.651	-235.940***	1.932	171.615**	-51.297**
	(89.006)	(43.620)	(7.496)	(67.851)	(1.316)	(68.807)	(22.084)
Panel B: 95% confidence intervals for Local Learning under $\gamma_3 \sim N(0, \delta^2)$							
CI ($2\delta = 10\%$)	(-539.5 -162.9)	(26.8 204.0)	(-16.3 13.0)	(-377.5 -94.4)	(-0.7 4.5)	(32.4 310.9)	(-95.8 -6.8)
CI ($2\delta = 15\%$)	(-555.5 -146.9)	(23.0 207.8)	(-16.4 13.0)	(-387.5 -84.4)	(-0.7 4.6)	(27.1 316.2)	(-97.3 -5.3)
CI ($2\delta = 25\%$)	(-599.8 -102.6)	(12.0 218.8)	(-16.4 13.1)	(-415.9 -56.0)	(-0.8 4.7)	(11.3 331.9)	(-101.7 -0.9)
CI ($2\delta = 50\%$)	(-746.2 43.8)	(-29.1 259.9)	(-16.4 13.1)	(-512.5 40.6)	(-1.3 5.2)	(-48.0 391.2)	(-118.8 16.2)
Observations	147	147	147	121	126	145	147

Panel A: Robust standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This panel presents the coefficients and standard errors for the Local Learning Multiplier variable from the second stage of 2SLS IV regressions estimated in Table 8. The column headings give the dependent variables for each of the columns. The Local Learning Multiplier variable has been instrumented using a predicted Local Learning Multiplier variable based on languages spoken in neighboring countries. Please refer to the notes from Table 8 for further details. Panel B: The 95% confidence intervals are computed allowing for violations in the exclusion restriction following the local to zero approach of Conley et al. (2012) and using the Stata ado program "plusexog" created by Clarke (2014). The row which says CI ($2\delta = x\%$) reports the 95% confidence interval for the local learning coefficient when we allow the direct effect of the instrument to be up to $x\%$ of the marginal effect of local learning from the original instrumental variable specification.

strategy. Doing so allowed us to causally establish the positive effect of the local learning multiplier on public goods outcomes. This finding is consistent with contact theory which hypothesizes that prejudice against other groups is mitigated by frequent interaction with those groups.

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. By using data that span virtually all countries of the world, our results provide important macro support to experimental evidence in favor of contact theory. This database should also be useful for researchers interested in analyzing the effect of local diversity on a variety of political economy outcomes, such as development and conflict.

Our paper suggests two important policy lessons. A first lesson concerns public policies that affect the geographic distribution of ethnolinguistic groups within countries. Compare, for example, Guatemala and Mauritius, two countries with identical levels of overall ethnolinguistic diversity, but with very different local learning multipliers. While geographic separation limits contact between Spanish and indigenous language speakers in Guatemala, it is common for individuals in Mauritius to interact on a daily basis with speakers of English, Kreol Morisien, French and Bhojpuri. Our result suggests that policies that improve the geographic mixing of ethnolinguistic groups in Guatemala may help narrow the gap with Mauritius in public goods outcomes. Although policy makers often shy away from attempts at social engineering, many European governments use social housing to geographically spread ethnic minorities, making neighborhoods and cities more equal in their diversity. Singapore goes even much further by implementing a quota system ensuring

that each housing block resembles the nation's ethnic make-up.[25]

A second lesson concerns migration policy. Taking a country's global fractionalization as given, outcomes tend to be better when there are few groups. A good example of a diverse country that achieves high local learning is Luxembourg: with around 45% of its population foreign-born, more than a third are of Portuguese origin and more than half come from just two countries. Conditional on a country's overall fractionalization, the predominance of a few large groups that are spatially mixed makes for a high local learning multiplier, and hence improved public goods outcomes. Policy makers may want to take this into account when designing migration policies.

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., Baqir, R., and Hoxby, C., 2004. "Political Jurisdictions in Heterogeneous Communities," *Journal of Political Economy*, 112, 348-396.

[3] Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R., 2003. "Fractionalization," *Journal of Economic Growth*, 8, 155-194.

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

[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] Algan, Y., Hémet, C., and Laitin, D., 2016. "The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation," *Journal of Political Economy* 124, 696-733.

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

[9] Ashraf, Q. and Galor, O., 2013. "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103, 1-46.

[10] Barro, R. and Lee, J., 2010. "A New Data Set of Educational Attainment in the World, 1950-2010," NBER Working Paper #15902.

---

[25]Similarly, the causal interpretation of contact theory was an important argument in the landmark U.S. Supreme Court case Brown v. Board of Education in 1954 that led to the racial desegregation of public schools (Putnam, 2007). Naturally, these policies are not exempt of controversy, as they trade off individual freedom of choice with social outcomes that are deemed desirable.

[11] Bertrand, M. and Duflo, E., 2016. "Field Experiments on Discrimination," NBER Working Paper #22014.

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

[13] Blumer, H., 1958. "Race Prejudice as a Sense of Group Position," *Pacific Sociological Review,* 1, 3-7.

[14] Boisjoly, J., Duncan, G.J., Kremer, M., Levy, D. M., and Eccles, J., 2006. "Empathy or Antipathy? The Impact of Diversity," *American Economic Review*, 96, 1890-1905.

[15] Burns, J., Corno, L. and La Ferrara, E., 2014. "Does Interaction Affect Racial Prejudice And Cooperation? Evidence From Randomly Assigned Peers In South Africa," work in progress.

[16] Carrell, S. E. , Hoekstra, M. and West, J. E., 2015. "The Impact of Intergroup Contact on Racial Attitudes and Revealed Preferences, NBER Working Paper #20940.

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

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

[19] Clarke, D. (2014). PLAUSEXOG: Stata module to implement Conley et al's plausibly exogenous bounds. Statistical Software Components

[20] Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1), 260-272.

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

[22] Dahlberg, M., Edmark, K., and Lundqvist, H., 2012. "Ethnic Diversity and Preferences for Redistribution," *Journal of Political Economy*, 120, 41-76.

[23] 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.

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

[25] Enos, R. D., 2014. "Causal Effect of Intergroup Contact on Exclusionary Attitudes," *PNAS*, 111, 3699-3704.

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

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

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

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

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

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

[32] La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R., 1999. "The Quality of Government," *Journal of Law, Economics, and Organization*, 15, 222-79.

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

[34] Lawton, C. and Ackrill, R. 2016. "Hard Evidence: How Areas with Low Immigration Voted Mainly for Brexit," *The Conversation*, `http://theconversation.com/hard-evidence-how-areas-with-low-immigration-voted-mainly-for-brexit-62138` (accessed 15 January 2017).

[35] Lewis, P. M., Simons, G. F., and Fennig, C. D., 2014. *Ethnologue: Languages of the World*, 17th edition, Dallas: SIL International.

[36] Massey, D. S. and Denton, N. A., 1988. "The Dimensions of Residential Segregation," *Social Forces*, 67, 281-315.

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

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

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

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

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

[42] 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.

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

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

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

[46] Weidmann, N. B., Rod, J. K., and Cederman, L., 2010, "Representing Ethnic Groups in Space: A New Dataset," *Journal of Peace Research*, 47(4), 491-499.

A. Proof of Proposition 1

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 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

$$\sum_{\ell}\sum_{i} N\ s_{\ell}\ s_{\ell i}\ (1 - \frac{G}{v_{\ell i}}) = G \tag{14}$$

which can be written as

$$\sum_{\ell}\sum_{i} N\ s_{\ell}\ s_{\ell i} - \sum_{\ell}\sum_{i} N\ s_{\ell}\ s_{\ell i} \left(\frac{\kappa_2 + a_{\ell i}}{\kappa_1}\right) G = G \tag{15}$$

or

$$N - NG\frac{\kappa_2}{\kappa_1} - NG\sum_{\ell}\sum_{i} s_{\ell}\ s_{\ell i}\frac{a_{\ell i}}{\kappa_1} = G,$$

so

$$N - \frac{NG}{\kappa_1}(\kappa_2 - \sum_{\ell} s_{\ell}a_{\ell}) = G.$$

Recall that average antagonism is

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

so the solution to (15) can be written as

$$G = \frac{\kappa_1 N}{\kappa_1 + N(\kappa_2 + A)} \tag{16}$$

Expression (16) shows that the equilibrium total contribution G is a decreasing function of average antagonism A. ∎

B. 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).*

Linguistic segregation. Linguistic segregation between subnational administrative regions. *Source: Alesina and Zhuravskaya (2011).*

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.*

Ethnolinguistic maps. For information on linguistic groups we use the digitized version of the 17th 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. *Source: World Language Mapping System Version 17, 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/`

Table B.1. Summary Statistics

	Obs	Mean	SD	Min	Max
1. Indices					
Local learning (level 2)	223	0.05	0.07	0.00	0.25
Local learning (level 5)	223	0.05	0.06	0.00	0.25
Local learning (level 15)	223	0.06	0.06	0.00	0.25
Local learning (level 5, 10 km)	223	0.05	0.07	0.00	0.25
Global ELF (level 2)	223	0.19	0.21	0.00	0.90
Global ELF (level 5)	223	0.25	0.26	0.00	0.97
Global ELF (level 15)	223	0.38	0.31	0.00	0.99
Local ELF (level 2)	223	0.12	0.15	0.00	0.71
Local ELF (level 5)	223	0.14	0.16	0.00	0.73
Local ELF (level 15)	223	0.20	0.17	0.00	0.73
Local ELF (level 5, 10 km)	223	0.14	0.16	0.00	0.73
Local learning (level 5, instrument)	223	0.06	0.06	0.00	0.25
Local ELF (level 5, instrument)	223	0.18	0.18	0.00	0.83
2. Dependent variables					
Child mortality	184	56.49	55.49	4.20	230.80
Measles immunization	183	81.72	15.48	27.00	99.00
Hospital beds	188	3.64	3.25	0.17	20.62
Illiteracy	147	20.08	20.39	0.00	80.97
School attainment	141	1.98	0.44	0.67	2.63
Improved sanitation	190	69.35	30.70	7.03	100.00
Roads	181	7.43	8.52	0.32	54.12

Table B.2. Correlations between Indices

	LL (2)	LL (5)	LL (15)	LL (5,10k)	G-ELF (2)	G-ELF (5)	G-ELF (15)	L-ELF (2)	L-ELF (5)	L-ELF (15)	L-ELF (5,10k)	LL Inst (5)
Local learning (level 5)	0.95	1.00										
Local learning (level 15)	0.69	0.73	1.00									
Local learning (level 5, 10 km)	0.94	0.98	0.70	1.00								
Global ELF (level 2)	0.74	0.66	0.39	0.66	1.00							
Global ELF (level 5)	0.59	0.58	0.28	0.58	0.84	1.00						
Global ELF (level 15)	0.38	0.39	0.26	0.40	0.59	0.76	1.00					
Local ELF (level 2)	0.97	0.89	0.62	0.88	0.83	0.65	0.43	1.00				
Local ELF (level 5)	0.91	0.92	0.62	0.91	0.78	0.78	0.55	0.92	1.00			
Local ELF (level 15)	0.67	0.69	0.75	0.68	0.60	0.63	0.76	0.68	0.76	1.00		
Local ELF (level 5, 10 km)	0.89	0.89	0.58	0.92	0.79	0.78	0.56	0.91	0.98	0.73	1.00	
Local learning (level 5, instr)	0.79	0.83	0.59	0.82	0.67	0.66	0.44	0.75	0.79	0.60	0.77	1.00
Local ELF (level 5, instr)	0.68	0.68	0.41	0.67	0.80	0.90	0.64	0.74	0.83	0.65	0.81	0.85

LL refers to local learning, G-ELF refers to global ELF, L-ELF refers to local ELF. First number in brackets corresponds to the linguistic level of aggregation, whereas the second number corresponds to the geographic size of cells, with a default cell size of 5 km by 5 km. For example, L-ELF (5,10k) refers to the local fractionalization index at linguistic aggregation level 5, using a cell size of 10 km by 10 km.

Table B.3. Global Fractionalization and Local Learning by Country

Country	Global ELF	LL	Country	Global ELF	LL
Afghanistan	0.681	0.060	Denmark	0.022	0.006
Albania	0.562	0.079	Djibouti	0.492	0.077
Algeria	0.342	0.024	Dominica	0.313	0.155
American Samoa	0.066	0.033	Dominican Republic	0.035	0.017
Andorra	0.000	0.000	East Timor	0.635	0.152
Angola	0.011	0.005	Ecuador	0.182	0.043
Anguilla	0.141	0.070	Egypt	0.080	0.034
Antigua and Barbuda	0.500	0.250	El Salvador	0.004	0.002
Argentina	0.011	0.005	Equatorial Guinea	0.031	0.012
Armenia	0.119	0.029	Eritrea	0.334	0.041
Aruba	0.328	0.143	Estonia	0.383	0.192
Australia	0.033	0.000	Ethiopia	0.782	0.042
Austria	0.017	0.008	Fiji	0.506	0.248
Azerbaijan	0.410	0.065	Finland	0.097	0.042
Bahamas	0.295	0.147	France	0.090	0.030
Bahrain	0.000	0.000	French Polynesia	0.430	0.212
Bangladesh	0.018	0.008	Gabon	0.060	0.029
Barbados	0.092	0.046	Gambia	0.658	0.138
Belarus	0.000	0.000	Georgia	0.547	0.115
Belgium	0.537	0.041	Germany	0.011	0.006
Belize	0.755	0.046	Ghana	0.649	0.062
Benin	0.552	0.057	Greece	0.061	0.021
Bermuda	0.000	0.000	Grenada	0.064	0.030
Bhutan	0.746	0.062	Guadeloupe	0.033	0.016
Bolivia	0.538	0.115	Guam	0.499	0.250
Bosnia and Herzegovina	0.002	0.001	Guatemala	0.517	0.058
Botswana	0.084	0.028	Guinea	0.603	0.017
Brazil	0.034	0.007	Guinea-Bissau	0.773	0.092
British Virgin Islands	0.167	0.084	Guyana	0.055	0.011
Brunei	0.578	0.135	Haiti	0.000	0.000
Bulgaria	0.150	0.007	Honduras	0.044	0.011
Burkina Faso	0.439	0.032	Hungary	0.082	0.036
Burundi	0.000	0.000	Iceland	0.000	0.000
Cambodia	0.067	0.015	India	0.863	0.015
Cameroon	0.474	0.024	Indonesia	0.760	0.036
Canada	0.429	0.011	Iran	0.513	0.126
Cape Verde Islands	0.000	0.000	Iraq	0.436	0.102
Cayman Islands	0.000	0.000	Ireland	0.063	0.006
Central African Republic	0.433	0.113	Israel	0.521	0.154
Chad	0.848	0.044	Italy	0.255	0.083
Chile	0.042	0.017	Jamaica	0.000	0.000
China	0.084	0.014	Japan	0.036	0.008
Colombia	0.021	0.008	Jordan	0.031	0.016
Comoros	0.008	0.004	Kazakhstan	0.600	0.201
Congo	0.521	0.117	Kenya	0.595	0.020
Cook Islands	0.000	0.000	Kiribati	0.017	0.008
Costa Rica	0.040	0.009	Kuwait	0.040	0.020
Cote d'Ivoire	0.715	0.059	Kyrgyzstan	0.600	0.205
Croatia	0.090	0.037	Laos	0.489	0.018
Cuba	0.000	0.000	Latvia	0.453	0.224
Cyprus	0.244	0.040	Lebanon	0.146	0.069
Czech Republic	0.024	0.009	Lesotho	0.000	0.000
DRC	0.370	0.072	Liberia	0.715	0.082

LL refers to local learning.

34

Country	Global ELF	LL	Country	Global ELF	LL
Libya	0.111	0.017	Russian Federation	0.184	0.018
Liechtenstein	0.000	0.000	Rwanda	0.001	0.000
Lithuania	0.342	0.148	St. Kitts and Nevis	0.010	0.005
Luxembourg	0.353	0.177	St. Lucia	0.020	0.010
Macedonia	0.472	0.042	St. Pierre and Miquelon	0.069	0.035
Madagascar	0.025	0.013	St. Vincent & Grenadines	0.007	0.004
Malawi	0.002	0.001	Samoa	0.002	0.001
Malaysia	0.626	0.129	San Marino	0.000	0.000
Maldives	0.000	0.000	Saudi Arabia	0.000	0.000
Mali	0.821	0.073	Senegal	0.282	0.044
Malta	0.076	0.038	Serbia	0.413	0.046
Marshall Islands	0.000	0.000	Seychelles	0.099	0.048
Martinique	0.041	0.021	Sierra Leone	0.746	0.038
Mauritania	0.221	0.026	Slovakia	0.192	0.025
Mauritius	0.529	0.202	Slovenia	0.012	0.006
Mayotte	0.437	0.208	Solomon Islands	0.238	0.066
Mexico	0.107	0.023	Somalia	0.057	0.007
Micronesia	0.075	0.037	South Africa	0.393	0.080
Moldova	0.549	0.183	South Korea	0.000	0.000
Monaco	0.000	0.000	South Sudan	0.717	0.039
Mongolia	0.156	0.026	Spain	0.027	0.011
Montenegro	0.365	0.150	Sri Lanka	0.319	0.062
Montserrat	0.026	0.013	Sudan	0.331	0.043
Morocco	0.435	0.013	Suriname	0.788	0.152
Mozambique	0.122	0.061	Swaziland	0.037	0.018
Myanmar	0.473	0.028	Sweden	0.048	0.020
Namibia	0.388	0.070	Switzerland	0.483	0.034
Nauru	0.189	0.095	Syria	0.197	0.021
Nepal	0.697	0.060	Taiwan	0.035	0.006
Netherlands	0.185	0.043	Tajikistan	0.430	0.146
Netherlands Antilles	0.455	0.062	Tanzania	0.070	0.010
New Caledonia	0.507	0.194	Thailand	0.180	0.022
New Zealand	0.081	0.035	Togo	0.571	0.080
Nicaragua	0.068	0.005	Tonga	0.000	0.000
Niger	0.633	0.076	Trinidad and Tobago	0.509	0.248
Nigeria	0.872	0.035	Tunisia	0.008	0.003
Niue	0.071	0.036	Turkey	0.375	0.060
Norfolk Island	0.311	0.155	Turkmenistan	0.176	0.025
North Korea	0.000	0.000	Turks and Caicos Islands	0.146	0.073
N. Mariana Islands	0.538	0.214	Uganda	0.673	0.136
Norway	0.011	0.004	Ukraine	0.090	0.039
Oman	0.298	0.075	United Arab Emirates	0.445	0.167
Pakistan	0.666	0.042	United Kingdom	0.025	0.006
Palau	0.063	0.032	United States	0.246	0.088
Palestinian WB & Gaza	0.000	0.000	U.S. Virgin Islands	0.454	0.227
Panama	0.265	0.032	Uruguay	0.004	0.002
Papua New Guinea	0.967	0.011	Uzbekistan	0.338	0.076
Paraguay	0.517	0.143	Vanuatu	0.179	0.082
Peru	0.315	0.048	Venezuela	0.024	0.010
Philippines	0.421	0.023	Viet Nam	0.181	0.039
Poland	0.042	0.018	Wallis and Futuna	0.018	0.009
Portugal	0.050	0.025	Yemen	0.014	0.000
Puerto Rico	0.049	0.024	Zambia	0.028	0.014
Qatar	0.485	0.242	Zimbabwe	0.032	0.016
Romania	0.165	0.012			

LL refers to local learning.

C. Appendix Figures

Figure C.1. Global ELF by Country – Level 15
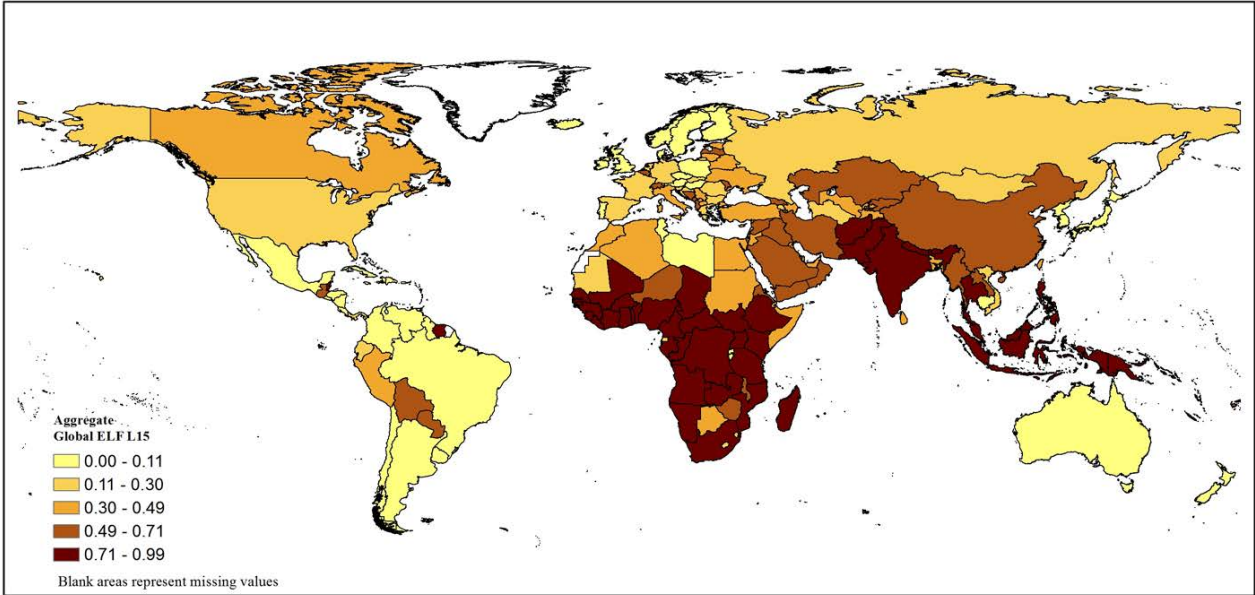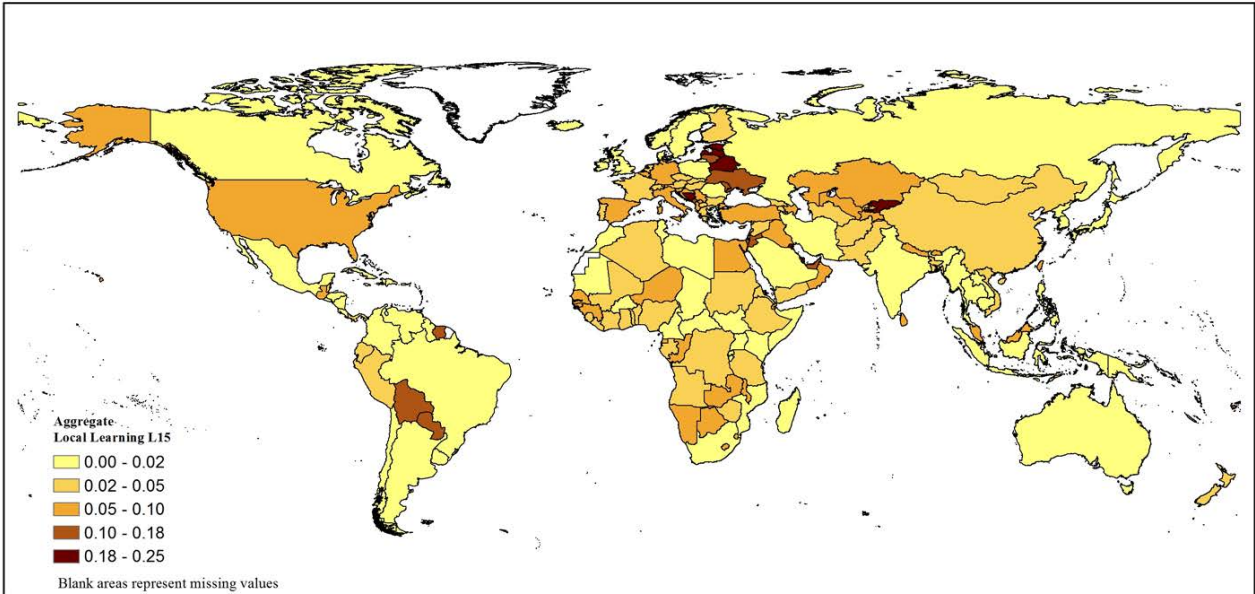

Figure C.2. Local Learning by Country – Level 15

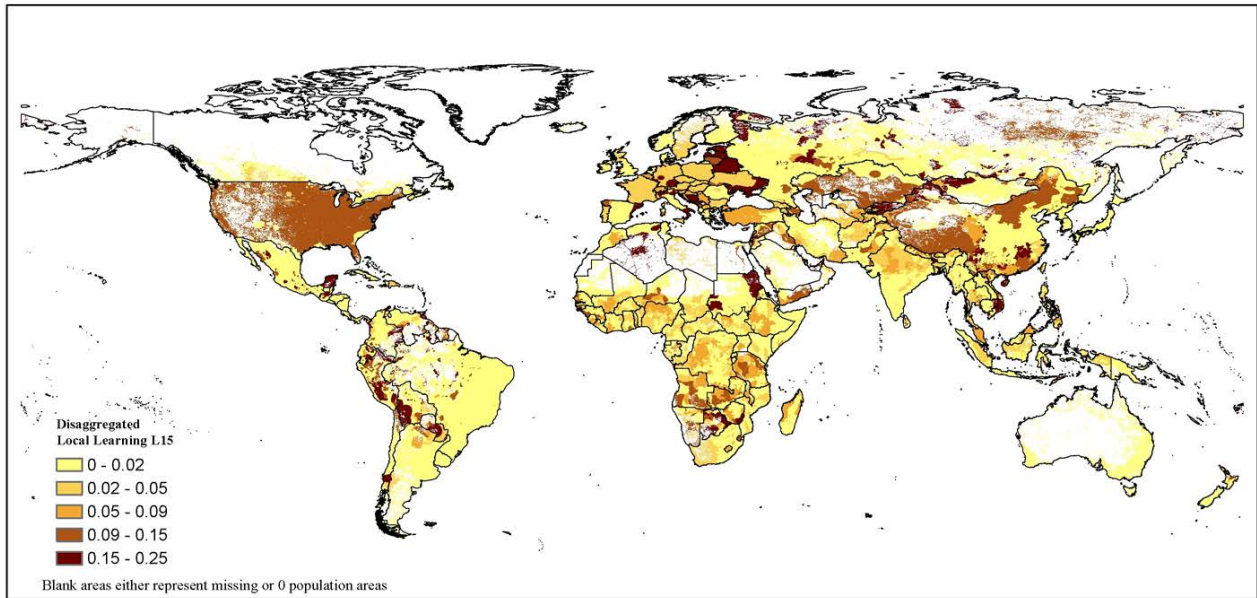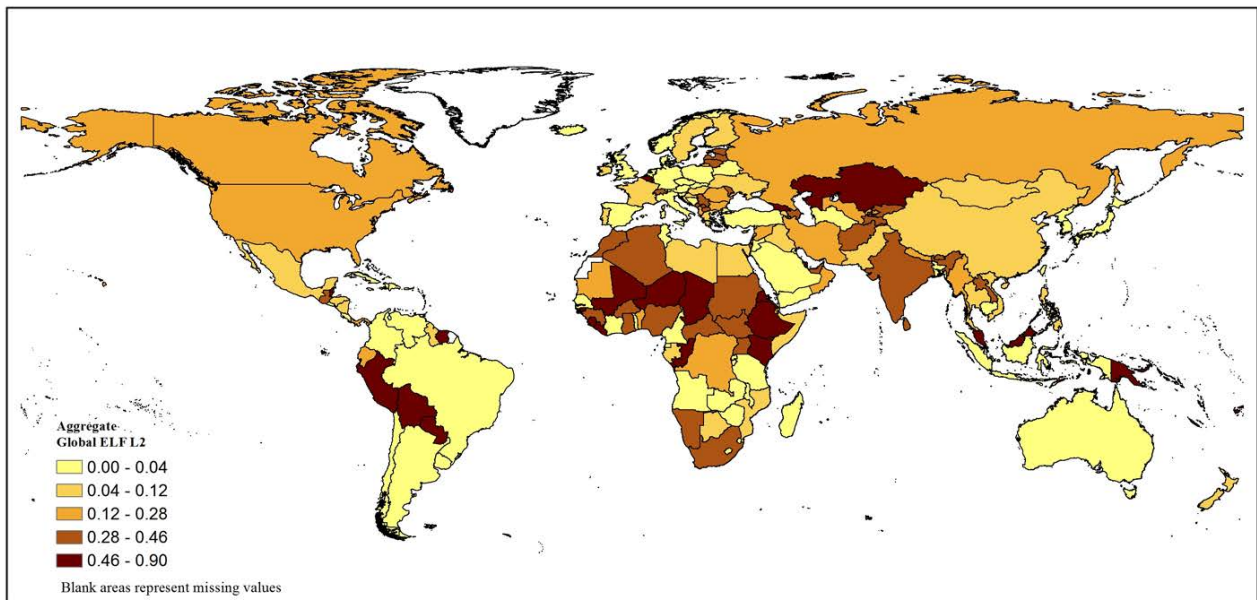Figure C.3. Local Learning at 5 km by 5 km Resolution – Level 15

Disaggregated Local Learning L15
- 0 - 0.02
- 0.02 - 0.05
- 0.05 - 0.09
- 0.09 - 0.15
- 0.15 - 0.25

Blank areas either represent missing or 0 population areas

Figure C.4. Global ELF by Country – Level 2

Aggregate Global ELF L2
- 0.00 - 0.04
- 0.04 - 0.12
- 0.12 - 0.28
- 0.28 - 0.46
- 0.46 - 0.90

Blank areas represent missing values

Figure C.5. Local Learning by Country – Level 2
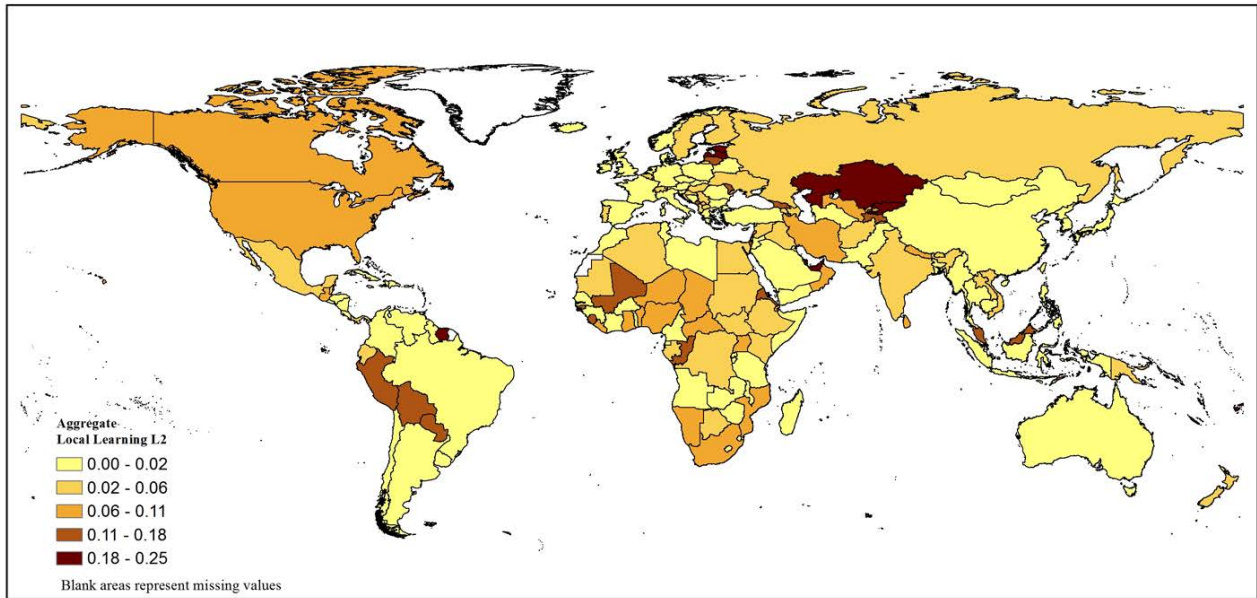
Figure C.6. Local Learning at 5 km by 5 km Resolution – Level 2
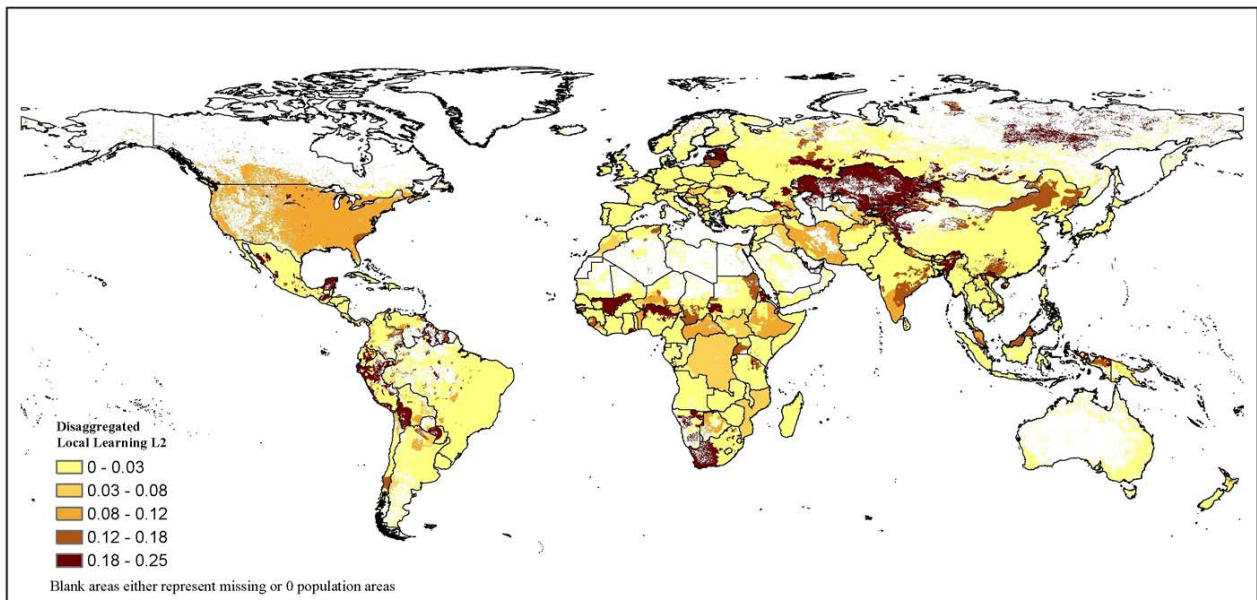
D. Instrument Construction

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 D.1 illustrates this.

Figure D.1. Closest Neighbors of Each Cell in Belgium
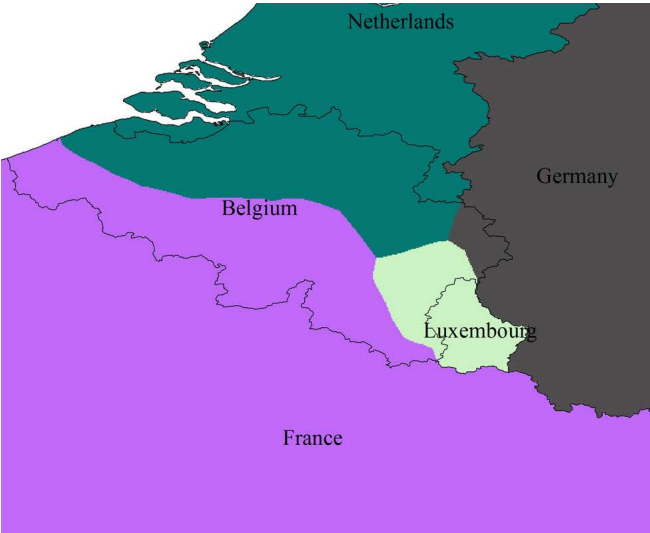
2. Any language that is spoken in k and that is also spoken in Belgium is then assigned to ℓ. Figure D.2 represents one of the languages spoken in each cell k. In Figure D.3 we can see how the two previous figures are combined to assign a language to each cell ℓ in Belgium. Of course, more than one language 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.

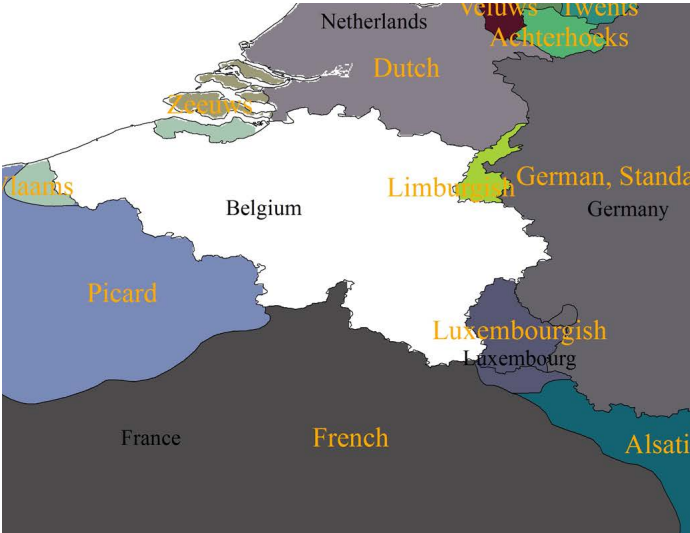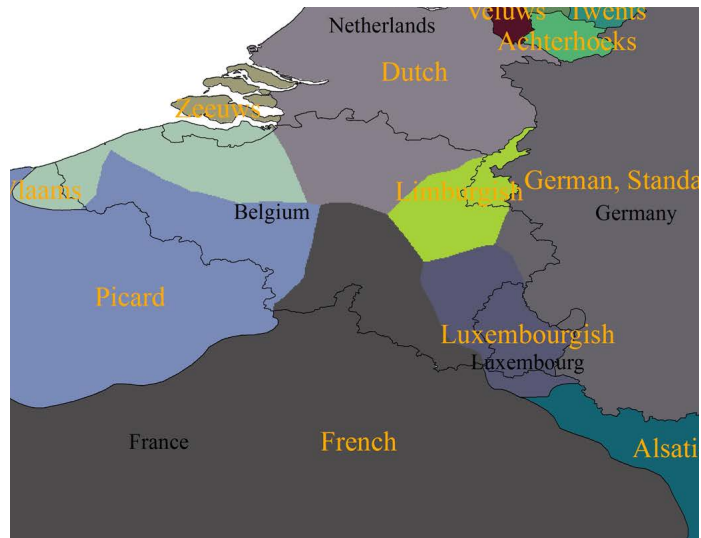Figure D.2. Languages in Belgium's Neighbors

Figure D.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.