Journal of Development Economics 97 (2012) 322–338

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

# The political economy of linguistic cleavages[☆]

Klaus Desmet [a,b,*], Ignacio Ortuño-Ortín [a], Romain Wacziarg [b,c,d]

[a] Department of Economics, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain
[b] CEPR, United Kingdom
[c] UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles CA 90095, United States
[d] NBER, United States

ARTICLE INFO

*Article history:*
Received 6 March 2010
Received in revised form 10 February 2011
Accepted 23 February 2011

*JEL classification:*
H1
N4
O4
O5

*Keywords:*
Ethnolinguistic cleavages
Diversity
Linguistic trees
Civil conflict
Redistribution
Public goods
Growth

ABSTRACT

This paper uses a linguistic tree, describing the genealogical relationship between all 6912 world languages, to compute measures of diversity at different levels of linguistic aggregation. By doing so, we let the data inform us on which linguistic cleavages are most relevant for a range of political economy outcomes, rather than making ad hoc choices. We find that deep cleavages, originating thousands of years ago, lead to better predictors of civil conflict and redistribution. The opposite pattern emerges when it comes to the impact of linguistic diversity on growth and public goods provision, where finer distinctions between languages matter.


## 1. Introduction

How does ethnolinguistic diversity affect political and economic outcomes? In recent years, a vast literature has argued that such cultural heterogeneity impacts a wide range of outcomes, fostering civil war, undermining growth, hindering redistribution and the provision of public goods. However, evidence on this point remains subject to some disagreement. For instance, there is a vibrant debate on the role of ethnolinguistic divisions as determinants of civil wars.[1] Econometric results on growth, redistribution and public goods provision also vary widely across studies, raising issues of robustness.[2]

These inconclusive results may stem in part from the inability to convincingly define the ethnolinguistic groups used as primitives to construct measures of heterogeneity. When faced with the issue of how to define groups, researchers have either relied on readily available classifications, such as the ones based on the *Atlas Narodov Mira* or the *Encyclopedia Britannica*, or have carefully constructed their own classifications.[3] Both approaches are problematic: the former

☆ We thank Francesco Caselli, Jim Fearon, Thierry Mayer, John Roemer, Enrico Spolaore and three anonymous referees for helpful comments. Desmet gratefully acknowledges financial support from the Comunidad de Madrid (PROCIUDAD-CM), and the Spanish Ministry of Science (ECO2008-01300). Ortuño-Ortín gratefully acknowledges financial support from the Spanish Ministry of Science (ECO2010-19596). Wacziarg gratefully acknowledges financial support from Stanford University's Presidential Fund for Innovation in International Studies and from UCLA's Center for International Business Education and Research.

* Corresponding author at: Department of Economics, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain. Tel.: +34 91 624 9845, fax: +34 91 624 9875.
*E-mail addresses:* klaus.desmet@uc3m.es (K. Desmet), iortuno@eco.uc3m.es (I. Ortuño-Ortín), wacziarg@ucla.edu (R. Wacziarg).

0304-3878/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jdeveco.2011.02.003

[1] Fearon and Laitin (2003) show that ethnic fractionalization is not an important determinant of the onset of civil wars. Montalvo and Reynal-Querol (2005), in contrast, argue that ethnic polarization is a significant determinant of the incidence of civil conflict.

[2] Alesina et al. (2003) argue that while ethnic and linguistic fractionalization are usually negatively related to growth and the quality of government, the significance of these partial correlations is sensitive to the specification.

[3] For an excellent discussion of the difficulties raised by the issue of defining relevant or salient ethnolinguistic groups, see Alesina and La Ferrara (2005), section 5.2.1, page 792.

runs the risk of missing the relevant cleavages, whereas the latter is subject to the criticism that groups are defined based on how important they are expected to be for the problem at hand. In this paper, we propose a methodology that addresses both criticisms, and argue that the degree of coarseness of ethnolinguistic classifications has profound implications for inference on the role of diversity.

The methodology we propose computes diversity measures at different levels of aggregation. We do so by exploiting the information of language trees. We refer to this as a phylogenetic approach, since tree diagrams describe the family structure of world languages. Depending on how finely or coarsely groups are defined, the measure of linguistic diversity will be different. For example, if one takes the different dialects of Italian to constitute different groups, then Italy appears to be very diverse. However, if one considers these different dialects to be only minor variations of Italian, then Italy looks homogeneous. Apart from allowing us to classify languages at different levels of aggregation, this approach has the advantage of giving a historical dimension to our analysis. Coarse linguistic divisions, obtained at high levels of aggregation, describe cleavages that go back thousands of years. In contrast, finer divisions, obtained at low levels of aggregation, are the result of more recent cleavages. Since we rely on data that cover the entire set of 6912 world languages, and examine effects of heterogeneity measures computed at all possible levels of aggregation, we are able to capture a wide range of linguistic classifications. Rather than choosing the "correct" classification ourselves, we let the data inform us as to which linguistic cleavages are most relevant for different outcomes of interest.[4]

Our empirical analysis reveals drastically different effects of linguistic diversity at different levels of aggregation. We also find that the relevant cleavages vary greatly across political economy outcomes. Starting from the data, specifications and estimation methods from major contributions to the literature on the political economy of ethnolinguistic diversity, we substitute our new measures of diversity for those commonly used. For civil conflict and the extent of redistribution, issues that inherently involve conflicts of interest, coarse divisions seem to matter most. While we find only weak evidence that diversity (whether measured by fractionalization or polarization) affects the onset of civil wars at any level of linguistic aggregation, the estimated effects do tend to be larger and more significant when considering a coarse classification. This finding is consistent with existing conflicts in African countries, such as Chad and Sudan, on the border between the Afro-Asiatic family and the Nilo-Saharan family. It may also help explain conflict in certain Latin American countries, such as Mexico and Bolivia, where the Indo-European family coexists with different Amerindian languages. For redistribution, the results are more robust, and suggest once again that measures based on a high level of aggregation matter most. In contrast, for economic growth, where coordination between individuals or groups is essential and market integration is important, we find that finer divisions lead to heterogeneity measures that matter more. The same pattern holds across a wide array of measures of public goods provision.

Thus, when the main issue involves conflicts of interest (as for the onset of civil wars and the extent of redistribution), deep differences originating thousands of years ago matter most: different groups' interests differ more when cleavages are more deeply rooted. In contrast, more superficial and recent divisions are negatively related to growth, an outcome related to the ease of coordination. For instance, to the extent that clusters of economic activity form around language lines, linguistic divisions may limit the integration of markets, and prevent economic growth. Even though Hindi and Gujarati are not so different, this linguistic cleavage may hinder the integration of the corresponding regions of India. What matters here is whether two individuals or groups can interact effectively. In fact, finer linguistic classifications deliver heterogeneity measures that matter more for outcomes such as economic growth, which is hindered by lack of coordination and integration. As for public goods, they fall somewhere in between both cases: although they have a redistributive aspect, their effective provision also requires coordination between groups or individuals. Empirically, we find that fine linguistic divisions, based on more superficial cleavages, are correlated with lower public goods provision across a wide array of indicators.

This paper is related to a vast literature in political economy. Various authors have studied how ethnolinguistic diversity affects redistribution, growth and civil conflict (Alesina and La Ferrara, 2005; Alesina et al., 1999, 2003; Easterly and Levine, 1997; Fearon and Laitin, 2003; La Porta et al., 1999, among many others). Measurement issues are central to recent research on these topics. One issue is that standard indices of diversity do not take into account the distance between groups (Desmet et al., 2009; Fearon, 2003; Spolaore and Wacziarg, 2009). Another possibility is that for certain issues, such as civil conflict, polarization may be more relevant than fractionalization (Esteban and Ray, 1994; Montalvo and Reynal-Querol, 2005), an issue we revisit below. A third problem is the difficulty of determining the right level of aggregation when computing heterogeneity measures, i.e., identifying the relevant ethnolinguistic cleavages. This issue has received little attention, and it is the main focus of the present study.[5]

This paper is organized as follows. Section 2 describes conceptual issues related to the measurement of heterogeneity based on language trees, and describes the data. Section 3 discusses the effects of diversity on civil conflict and redistribution. Section 4 covers the effects on public goods provision and economic growth. Section 5 explores a number of robustness issues, and Section 6 concludes.

## 2. Aggregation and linguistic diversity

### 2.1. A tale of two countries

To illustrate our approach, we start with a comparative case study. Over the period 1965–2000, Chad and Zambia experienced some of the lowest growth rates on the globe, their income per capita shrinking by an average of 1 percentage point per year (Table 1). The 2005 Human Development Index ranked Chad 170 and Zambia 165 out of a total of 177 countries. It has long been argued that low growth may be related to high ethnolinguistic diversity. With 135 languages spoken in Chad, and between 40 and 70 in Zambia, these countries certainly are very diverse: taking the commonly used fractionalization index as a measure of diversity, the *Ethnologue* database on languages gives a value of 0.95 for Chad and 0.85 for Zambia, putting both countries in the top decile. As highlighted by Easterly and Levine (1997), data for a broad cross-section of countries point more formally to a general negative relationship between ethnic heterogeneity and economic

[4] Our approach is related to existing work arguing that people identify with different groups in different contexts (particularly the work of Crawford Young on situational identity — see Young, 1976). For instance, ethnolinguistic cleavages that matter for voting behavior in local elections may differ from those that matter for national elections. For a related point, see Posner's, 2005 book on ethnic politics in Zambia. More generally, cleavages that matter for some outcomes may not matter for others. There is no such thing as a "correct" classification of languages or ethnicities — this depends on the context.

[5] Fearon (2003) does discuss at length the issue of how to define the "right list" of ethnic groups serving as the basis for computing heterogeneity measures, and recognizes explicitly that not all cleavages may be relevant for a given outcome. However, he presents data on ethnic groups based on a single classification. Scarritt and Mozaffar (1999) present data on ethnic groups for Sub-Saharan countries using three different classifications, but do not examine the effects of using these different classifications on political and economic outcomes.

**Table 1**
Growth, conflict, redistribution and linguistic diversity in Chad and Zambia.

| | Chad | Zambia |
|---|---|---|
| Per capita growth 1960–1990 (Easterly-Levine),% | −1% | −1% |
| Per capita growth 1965–2000 (PWT 6.2),% | −1% | −1% |
| Years of civil war 1965–1999 | 35 | 0 |
| Redistribution as % of GDP, 1985–1995 | 0.9% | 3.8% |
| ELF (most disaggregated level) | 0.95 | 0.85 |
| ELF (at the aggregated level of language families) | 0.55 | 0.01 |
| Polarization (most disaggregated level) | 0.18 | 0.43 |
| Polarization (at the aggregated level of language families) | 0.89 | 0.02 |

performance. In our data, the 10% most diverse countries had an average per capita growth rate of a meager 0.54% over the period 1960–2004, whereas the 10% least diverse countries posted a much more sturdy figure of 2.59% (linguistic diversity here is measured using the most disaggregated classification of languages).

In spite of their high ethnolinguistic fractionalization, in terms of conflict and civil war Chad and Zambia have been at opposite sides of the spectrum. Chad has been at war almost continuously since independence, whereas Zambia has not witnessed any civil conflict worth speaking of. In Chad, during colonization, and after independence in 1960, the Christian South was privileged, and formed the political elite, to the detriment of the Islamic and partly Arab-speaking North. Dissatisfaction by the North led to a civil war, which started in 1965, and lasted for about a decade and a half, culminating in the rebels taking over the capital and ending Southern dominance. Since then the country has remained unstable, partly because of the inverted power relation, with the North now dominating the South, but also because of power struggles within these regions. In recent years, for example, there has been increasing ethnic tension between the Zaghawa and Tama, two non-Arab groups. Zambia, in contrast, has had a history of peaceful coexistence between the many groups and tribes. Although voting behavior in Zambia tends to run along language groups (Posner, 2003), it has not led to the violence seen in countries such as Chad. Income redistribution, which is an issue involving divergence of interests, is often interpreted as related to conflict. Data on redistribution confirm the contrast between both countries: figures on transfers and subsidies as a share of GDP reveal that on average between 1985 and 1995 Chad redistributed 0.9% of GDP, compared to 3.8% in Zambia.

This example illustrates the main point of this paper: although commonly used measures of diversity make Chad and Zambia look very similar, those measures mask one important difference between these countries in terms of diversity. Of the total population in Chad, one third speaks an Afro-Asiatic language, a little over half a Nilo-Saharan language, and the rest a language of the Niger-Congo family. In contrast, in Zambia, 99.5% of the population speaks a language from the Niger-Congo family. This raises an important point: whereas Chad and Zambia are amongst the most diverse countries on the globe, when considering language families rather than individual languages, we obtain a very different picture. While Chad continues to be one of the most diverse countries, ranking 7 out of 225, Zambia now looks very homogeneous, ranking 176 out of 225, similar to Portugal. In other words, when taking every language as being different, Zambia is very diverse, similar to Chad, whereas when aggregating into language families, Zambia no longer appears to be quite so heterogeneous.

In the example of Chad and Zambia, both countries are very diverse at low levels of aggregation, but only Chad continues to be very diverse at high levels of aggregation. It may be useful to consider an example that goes in the other direction. Afghanistan has about 50 languages, whereas Sri Lanka only 7, making Afghanistan relatively much more diverse at low levels of aggregation. However, at high levels of aggregation both countries are similar: 80% of their populations speak an Indo-European language, with a 20% minority of mostly Dravidian in the case of Sri Lanka and Altaic in Afghanistan. In this case, we would expect both countries to exhibit similar levels of conflict, but Sri Lanka should outperform Afghanistan in terms of economic growth. Consistent with this prediction, between 1948 and 1999 Afghanistan experienced civil conflict for 22 years and Sri Lanka for 18 years. In contrast, annual growth in real GDP per capita between 1970 and 2000 was −4% in Afghanistan, and +4% in Sri Lanka.

The experience of these different country pairs suggests that the type of diversity that matters for economic growth is different from the type of diversity that matters for civil conflict and redistribution. The essential difference between the two types of diversity is the degree of aggregation. The relevant degree of aggregation, and thus the relevant definition of a group, depends on the problem at hand. This case study suggests that, for economic growth, fine differences between languages may matter, whereas for civil conflict and redistribution, only coarse differences may play a role — as is confirmed below in large samples.[6]

### 2.2. Language trees and linguistic diversity

#### 2.2.1. The construction of language trees

This paper seeks to measure linguistic diversity at different levels of aggregation. To do so, we use language trees. We refer to this as a phylogenetic approach (as the linguistics literature does), referring to the fact that tree diagrams capture the genealogy of languages, classified in terms of their family structure.[7] Using language trees gives a historical dimension to our analysis. Coarse linguistic divisions, such as that between Indo-European and non Indo-European languages, describe cleavages that originate several thousand years ago. In contrast, finer divisions, such as that between Dutch and German, tend to be the result of more recent splits. For instance, Gray and Atkinson (2003) estimate separation times between language groups within the Indo-European family. While the separation between Indo-European languages and all others is estimated to have occurred prior to 8700 years ago, the separation time between different dialects of Modern Greek is estimated to have occurred only 800 years ago. There are differences of opinion between linguists on the precise dates, but the general point of an association between tree structure and separation times remains. We do not require that there be a strict association between the coarseness of the linguistic classification and the time since the linguistic split between groups occurred — we only point out that coarse classifications capture cleavages that tend to go back deeper in the past.

[6] The difference in the experience of Chad and Zambia (or Afghanistan and Sri Lanka) is not related to the use of measures of linguistic fractionalization rather than polarization, but to the issue of aggregation. As Table 1 reveals, using a standard measure of polarization instead of fractionalization leads to the same conclusion: the difference in polarization between Zambia and Chad is much more pronounced for highly aggregated linguistic classifications than for disaggregated ones. Correspondingly, conflict and war has been continuous in Chad, but absent in Zambia. We discuss the important issue of how the distinction between polarization and fractionalization (which has to do with the functional form used to calculate measures of diversity) relates to the level aggregation (which has to do with the definition of relevant groups) in Section 2.3.

[7] This point was recognized going at least as far back as Charles Darwin, who wrote: "If we possessed a perfect pedigree of the mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world; and if all extinct languages, and all intermediate and slowly changing dialects, were to be included, such an arrangement would be the only possible one. Yet it might be that some ancient language had altered very little and had given rise to few new languages, whilst others had altered much owing to the spreading, isolation, and state of civilization of the several co-descended races, and had thus given rise to many new dialects and languages. The various degrees of difference between the languages of the same stock, would have to be expressed by groups subordinate to groups; but the proper or even the only possible arrangement would still be genealogical; and this would be strictly natural, as it would connect together all languages, extinct and recent, by the closest affinities, and would give the filiation and origin of each tongue." (Darwin, 1902, p. 380).

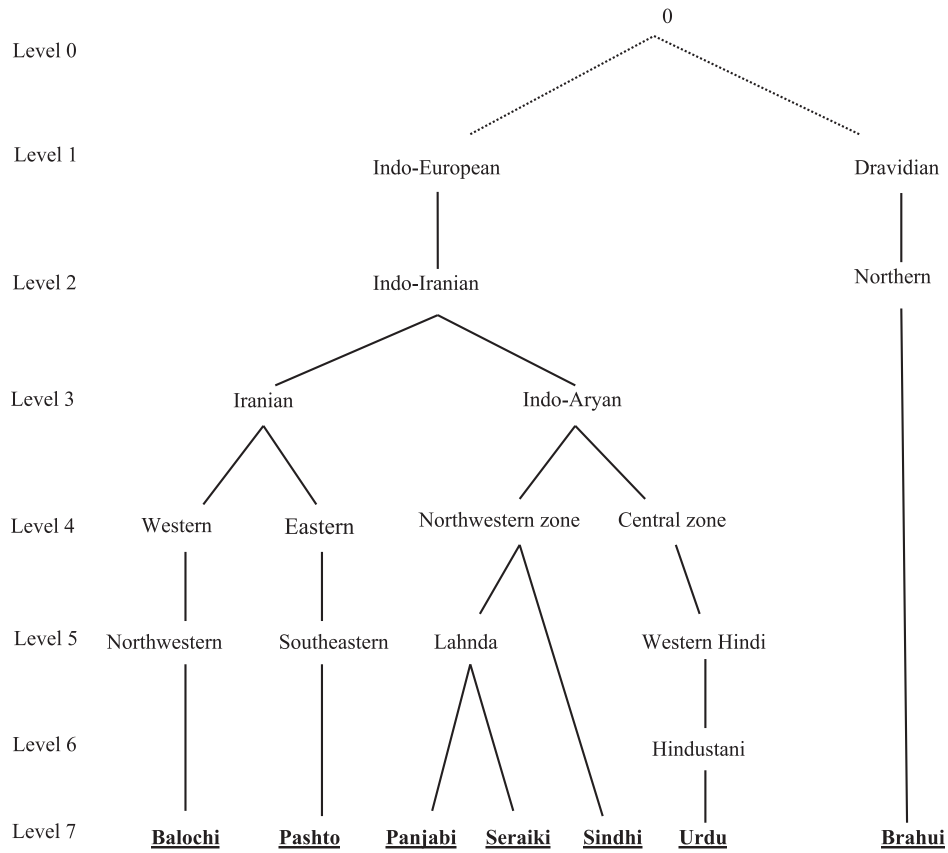


**Fig. 1.** Phylogenetic tree of major languages in Pakistan.

Linguistic differentiation occurs because specific human populations become relatively isolated from each other and, as a result, develop specific languages over time. In general three major factors can affect the degree to which languages differ. The first factor is the time since the populations speaking these languages have split from each other. As noted, populations speaking French and Spanish have split from each other much more recently than populations speaking, say, Swahili and Tibetan. The second factor, known by linguists as *Sprachbund* (or language union), results from interactions between populations that are already linguistically distinct (Emeneau and Anwar, 1980). For example, historically the spread of Latin words likely had a homogenizing influence on European languages, keeping Romance and Germanic languages more similar than would have been the case without commercial and political interactions. The third factor is the size of the population. Linguistic drift tends to be faster in smaller populations. For instance, Lithgow (1973) studies the Muyuw language, spoken on Woodlark Island (New Guinea): 13% of the Muyuw vocabulary was replaced in a period of 50 years during the middle of the 20th Century (see also Dixon, 1997, for a discussion). This language is spoken by only 6000 individuals, according to *Ethnologue*. Empirically, this determinant of linguistic differentiation does not greatly affect our measures of diversity, as it only affects very small linguistic groups. Linguistic trees such as the one from *Ethnologue*, which we use in our empirical analysis, are constructed by linguists to mainly capture the first factor.[8]

Given that linguistic divisions arise because of splits between populations, one would expect coarser linguistic divisions to reflect deeper population cleavages. Consistent with this view, Cavalli-Sforza et al. (1988) suggest that there is a link between the major language families and the main human genetic clusters. Although there are some notable exceptions, subsequent studies have tended to confirm this finding. For example, Belle and Barbujani (2007), using more recent data on genetic polymorphism and languages from Ethnologue, conclude that “globally speaking, the genetic differences between the main language phyla probably reflect relatively ancient demographic subdivisions”.

We emphasize that the issue of aggregation is separate from (although related to) the issue of how to capture the distance separating languages when computing measures of diversity (for a paper that accomplishes the latter goal, see Desmet et al., 2009; for indices of fractionalization that take into account distances, see Greenberg, 1956 and Bossert et al., in press). We are after identifying the level of aggregation that corresponds to the most relevant cleavages for the various dependent variables we examine. A focus on the level of aggregation that captures the relevant cleavages retains a strong focus on ethnolinguistic groups as the basis for individuals' identification with, or alienation from, a given ethnolinguistic identity or group (we borrow the identification/alienation terminology from Esteban and Ray, 1994). In contrast, distance-weighted measures of diversity (such as the measure proposed by Greenberg, 1956), capture the expected distance between individuals, and relegates the group structure to the background. Our approach is therefore distinct from approaches that make use of distances between groups: we are interested in identifying the group structures (or classifications) that matter most for political economy outcomes. At the same time, by construction, more aggregated classifications retain groups that tend to be more distant from each other (in terms, say, of separation times, or in terms of how different the languages are), compared to more disaggregated classifications.

To illustrate the discussion above, Fig. 1 displays the tree for the major languages in Pakistan. On the left side of the figure, we list the

[8] There are controversies among linguists on the right classification of languages. For example, Greenberg (1987) considers that all Native American languages can be classified into three groups (Eskimo-Aleut, Na-Dene and Amerindian) whereas the *Ethnologue* contemplates dozens of unrelated families. However, the classification provided in the *Ethnologue* is the most widely used and, to the best of our knowledge, the only one available in electronic format covering all of the languages of the world.

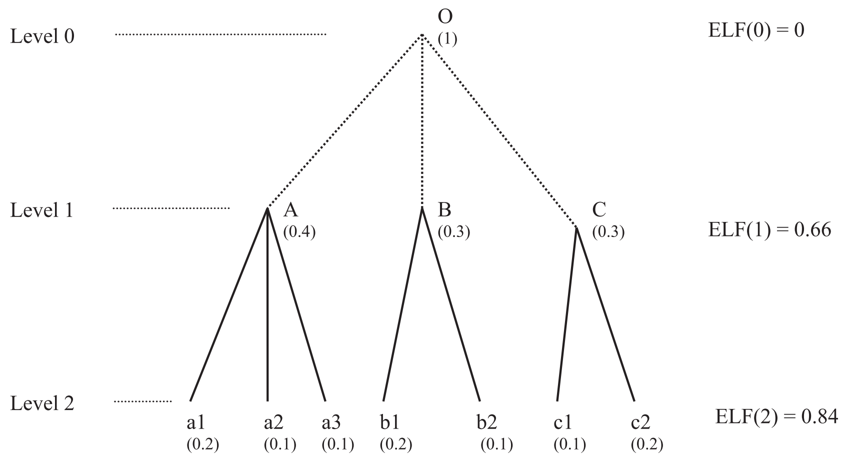


**Fig. 2.** Hypothetical language tree.

level of aggregation. At level 7, the most disaggregated level, there are seven main languages: Panjabi, Pashto, Sindhi, Seraiki, Urdu, Balochi, and Brahui. Going up the tree, the number of groups declines, as the level of aggregation rises. For instance, at level 4, there are only five linguistic groups — at that level, Panjabi, Seraiki and Sindhi are classified as one and the same. At level 3, only three linguistic groups are left (Iranian, Indo-Aryan and Northern Dravidian). Finally, at aggregation level one, there are two groups: Dravidian (Brahui) and Indo-European (all others). These classifications allow us to compute measures of diversity at each level of aggregation.

### *2.2.2. Measuring diversity at different levels of aggregation*

How precisely are the measures of diversity computed? An example of how a language tree looks like is shown in Fig. 2. The *root* of the tree is represented by the upper-case letter *O*, whereas the *leafs* of the tree are represented by lower-case letters *a* through *c*. In Fig. 2, all leafs have a common root, so that the tree is rooted (this terminology is borrowed from the field of linguistics). As can be seen, the tree has three different levels. Each of the seven leafs at level 2 represents a living language. The three nodes at level 1 represent the (extinguished) mother languages of the existing languages. These correspond to the proto-languages of the different families, such as Indo-European or Sino-Tibetan. The node at level 0 represents the hypothetical common ancestor language of all families, referred to by linguists as Proto-Human. The number below each living language at level 2 indicates the assumed shares of the population speaking the corresponding language. The numbers below the (extinguished) mother languages at level 1 are the aggregated population shares of their corresponding daughter languages.

To compute diversity at different levels, we require that the tree be rooted, and that the number of *branches* (or *edges*) between any leaf and the root be identical. In this subsection, we focus on the widely used index of ethnolinguistic fractionalization (or ELF), the probability that two randomly picked individuals belong to different groups (in our empirical work we also consider measures of polarization). The diversity measure at a given level of aggregation is the ELF index for the linguistic groups as they appear at that level. For example, diversity at level 2 is given by the ELF index, taking the seven living languages as the relevant groups. Thus, $ELF(2) = 1-3\times\left(0.2^2\right)-4\times\left(0.1^2\right) = 0.84$. To calculate diversity at level 1, the seven living languages are aggregated into 3 distinct groups *A*, *B* and *C*, resulting in an ELF index $ELF(1) = 1-0.4^2-2\times\left(0.1^2\right) = 0.66$.

One difficulty remains. The linguistic tree from *Ethnologue* is a rooted tree, but the number of branches varies among linguistic families and subfamilies. Fig. 3 depicts a generic language tree such as the one from Ethnologue. If we look at the proto-languages of the

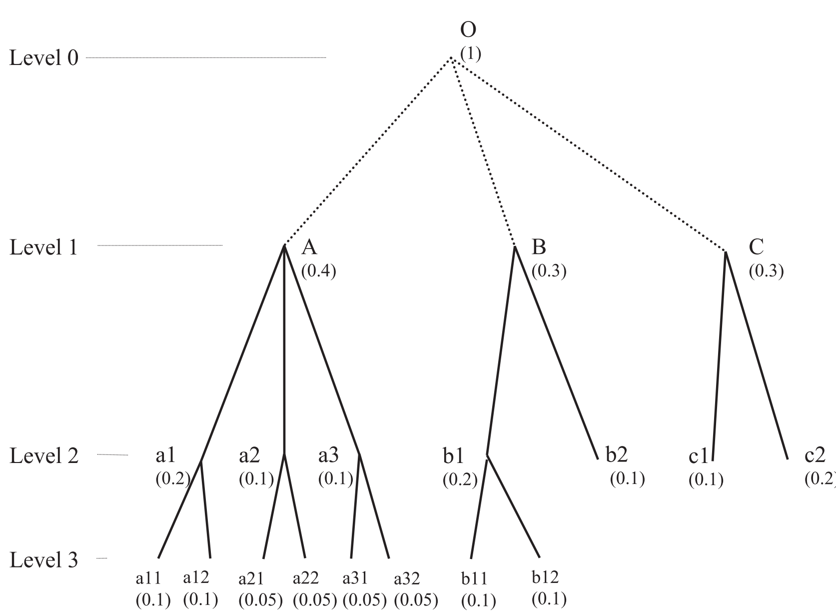


**Fig. 3.** Typical language tree from Ethnologue.

different families (Level 1), we can see that *A* has more descendent generations than *B* or *C*. As before, the leafs of the tree represent the existing languages. They are denoted by the letters *a11*, *a12*, *a21*, *a22*, *a31*, *a32*, *b11*, *b12*, *b2*, *c1* and *c2*. It is clear that for this type of tree we cannot use the method applied in Fig. 2, because at level 3 we would be ignoring 3 of the 11 languages. The branches in the tree need to be extended, and there are two main ways to do this, as displayed in the two panels of Fig. 4. This ensures that all the existing languages are represented as leafs at the lowest level of aggregation.

The first approach, displayed in Panel I of Fig. 4, assumes that all living languages are equally distant from the proto-languages of their respective families, where the distance between languages is defined by the number of branches or nodes separating them (in technical terms, this assumes that the tree is ultrametric). Take, for example, language *c1*. We insert a fictitious language, *c1**, at level 2, so that the total number of branches between *c1* and the origin language of the family, *C*, is the same as for all other leafs. The second approach, displayed in Panel II of Fig. 4, assumes that *c1* is only one branch removed from its origin language *C*. In this case, Fig. 4 shows that to have all living languages at the same level, we move *c1* down to level 3, but assume that its mother and grandmother have all remained the same as the origin language *C*.

In our empirical work, we favor measures based on the first approach, because it is reasonable to assume that languages went through intermediate states between the proto-languages of their respective families and their current form. The second approach, in contrast, assumes that some origin languages remained unchanged until recently. A further advantage of the first approach is that it does not change the "family relations" of the original Ethnologue trees. For example, in Fig. 3 language *b1* has a sister *b2* and a mother *B*; this is still so under the first approach, but not under the second approach, where the mother and the sister of *b1* are now the same language.

Although there are good reasons to prefer the first approach, for the sake of robustness we also computed and used measures based on the second approach. Using either approach did not make much difference for our empirical results. Based on either approach, our empirical results show that diversity measures based on either the highest level of aggregation or the lowest level of aggregation matter most. From Fig. 4 it is easy to see that, at the two extremes, both approaches are identical, so

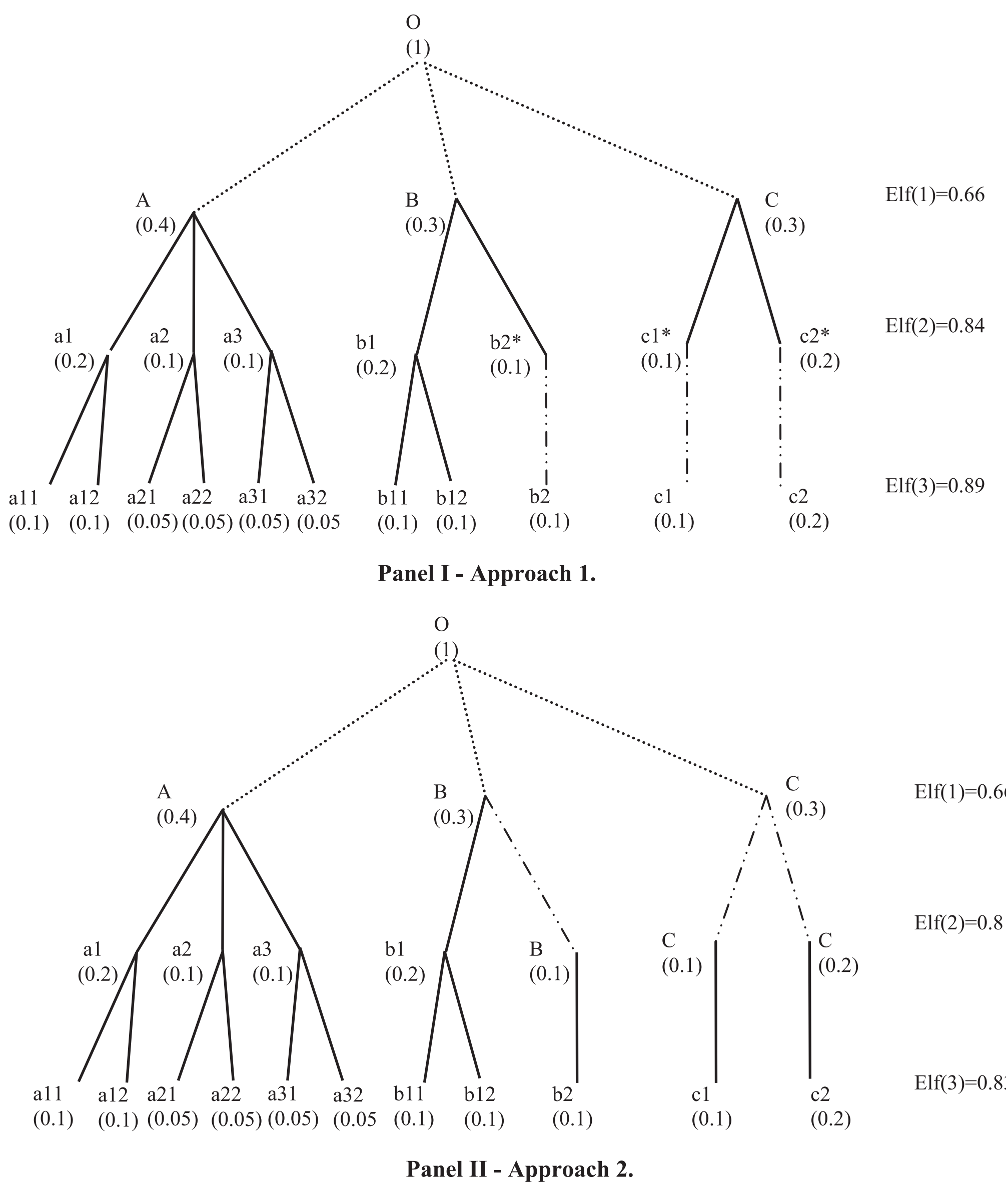


**Fig. 4.** Two different approaches.

**Table 2**
Summary statistics for ethnolinguistic diversity measures.

Panel A. Means and standard deviations

| Variable | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| ELF(1) | 0.156 | 0.180 | 0.000 | 0.647 |
| ELF(3) | 0.241 | 0.221 | 0.000 | 0.818 |
| ELF(6) | 0.328 | 0.272 | 0.000 | 0.941 |
| ELF(10) | 0.394 | 0.301 | 0.000 | 0.989 |
| ELF(15) | 0.412 | 0.308 | 0.000 | 0.990 |
| POL(1) | 0.283 | 0.314 | 0.000 | 0.999 |
| POL(3) | 0.384 | 0.316 | 0.000 | 0.998 |
| POL(6) | 0.423 | 0.297 | 0.000 | 0.996 |
| POL(10) | 0.435 | 0.279 | 0.000 | 0.996 |
| POL(15) | 0.432 | 0.278 | 0.000 | 0.996 |

Panel B. Correlations

| | ELF(1) | ELF(3) | ELF(6) | ELF(10) | ELF(15) | POL(1) | POL(3) | POL(6) | POL(10) |
|---|---|---|---|---|---|---|---|---|---|
| ELF(3) | 0.770 | 1 | | | | | | | |
| ELF(6) | 0.579 | 0.826 | 1 | | | | | | |
| ELF(10) | 0.544 | 0.708 | 0.848 | 1 | | | | | |
| ELF(15) | 0.526 | 0.672 | 0.798 | 0.977 | 1 | | | | |
| POL(1) | 0.988 | 0.754 | 0.565 | 0.530 | 0.514 | 1 | | | |
| POL(3) | 0.720 | 0.939 | 0.788 | 0.683 | 0.651 | 0.737 | 1 | | |
| POL(6) | 0.545 | 0.691 | 0.821 | 0.697 | 0.654 | 0.563 | 0.763 | 1 | |
| POL(10) | 0.444 | 0.568 | 0.643 | 0.664 | 0.638 | 0.466 | 0.637 | 0.838 | 1 |
| POL(15) | 0.391 | 0.513 | 0.595 | 0.542 | 0.555 | 0.408 | 0.572 | 0.777 | 0.925 |

226 observations.

that using one or the other should make no difference (we report on these empirical results in greater details in Section 5).

### 2.3. Measurement, summary statistics and specification

We consider two sets of commonly used measures of diversity: fractionalization and polarization. For $i(j)=1,\ldots,N(j)$ groups of size $s_{i(j)}$, where $j=1,\ldots,J$ denotes the level of disaggregation at which the group shares are considered, fractionalization is just the probability that two individuals chosen at random, will belong to different groups:

$$ELF(j) = 1-\sum_{i(j)=1}^{N(j)}\left[s_{i(j)}\right]^2 \quad (1)$$

This measure is maximized when each individual belongs to a different group. Polarization, in contrast, is maximized when there are two groups of equal size. We use the polarization measure from Montalvo and Reynal-Querol (2005). This index satisfies the conditions for a desirable index of polarization in the axiomatic approach of Esteban and Ray (1994):

$$POL(j) = 4\sum_{i(j)=1}^{N(j)}\left[s_{i(j)}\right]^2\left[1-s_{i(j)}\right] \quad (2)$$

We compute these measures at each of the 15 levels of aggregation available in the linguistic classification in the 15th edition of *Ethnologue*, the source for our linguistic data (Ethnologue, 2005). The sample contains 226 observations which include countries and their dependencies (due to data availability, our regression results are based on a smaller set of countries). Table 2 presents summary statistics for the diversity measures at 5 levels of aggregation (an online Appendix contains the corresponding data series by country). To facilitate the quantitative assessment of the regression results, Panel A displays means and standard deviations. When measured using the ELF index, the average degree of diversity rises as the level of aggregation falls, as expected. When measured using a polarization index, diversity falls at high levels of aggregation, and plateaus as aggregation falls further. To simplify the presentation of our results, in the empirical sections we focus on only 3 levels of aggregation (levels 1, 6 and 15, with higher numbers denoting a lower degree of aggregation). All our empirical results are also available at the other levels.

Interesting information can also be gleaned from Panel B of Table 2, displaying correlations. First, changing the level of aggregation greatly affects the measures of diversity: the correlation between ELF(1) and ELF(15) is only 0.526. Second, the correlation between polarization and fractionalization, at the same levels of aggregation, rises as the level of aggregation increases (the correlation between POL(15) and ELF(15) is only 0.555, while the correlation between ELF(1) and POL(1) is 0.988). This is intuitive as, when aggregating, fewer groups remain, and the distinction between polarization and fractionalization fades.[9] Third, aggregating up is not the same as switching from a measure of fractionalization to a measure of polarization: the correlation between ELF(1) and POL(15) is only 0.391. This last observation indicates that the issue of aggregation is very different from the choice of functional form to compute diversity measures. In our empirical work, we show that switching from fractionalization to polarization measures has relatively benign effects on the substantive results, while changing the level of aggregation to compute either measure delivers vastly different estimates of the effect of diversity on political economy outcome.

Finally, Figs. 5 and 6 display the full distributions of ELF(1), ELF(15), POL(1) and POL(15). As can be seen, at high levels of aggregation the distributions of both fractionalization (ELF(1)) and polarization (POL(1)) have a strong positive skew. This makes sense: when classifying languages to be different only when they pertain to entirely different families, most countries display low levels of diversity, and only a few exhibit high diversity. In contrast, at low levels of aggregation the distributions of fractionalization (ELF(15)) and polarization (POL(15)) are much more uniform. That is, many of the countries that were not diverse when only looking at language families are now much more diverse.[10] This is the

[9] Although we find that the correlation between fractionalization and polarization increases with aggregation, it is easy to find counterexamples for which this is not true. The intuition for why this result does not always hold is related to the fact that fractionalization decreases with aggregation, whereas polarization could increase or decrease with aggregation. Of course, at the highest level of aggregation, most countries have only one or two groups left, in which case the two indices coincide (up to a constant). This explains the correlation of 0.988 between ELF(1) and POL(1).

[10] For instance, if we consider countries with more than half a million inhabitants, 23 of the 30 most diverse countries in the world are located in Sub-Saharan Africa at the most disaggregated level (ELF(15)). At the least diaggregated level (ELF(1)), only 10 of the 30 most diverse countries are in this region.

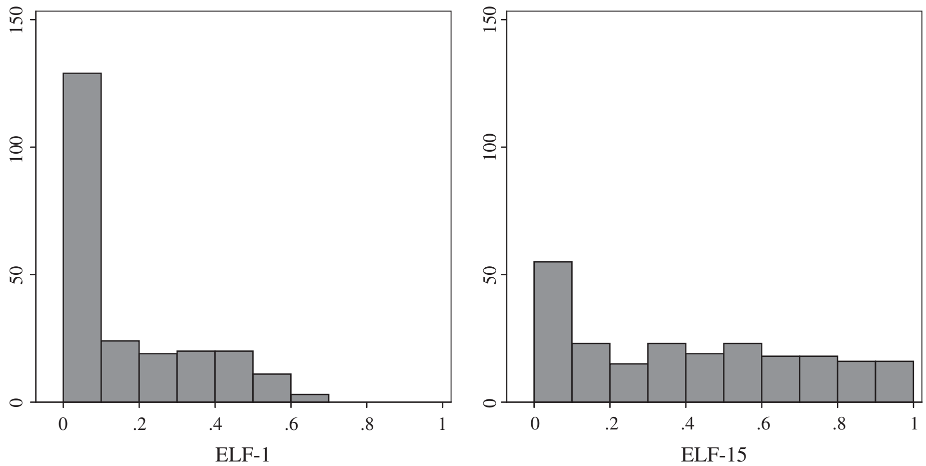


**Fig. 5.** Distributions of ELF-1 and ELF-15.

example of Zambia mentioned above: it is highly diverse if each of the 46 languages are taken to be different, and it is not very diverse when one considers that only 2 out of the 46 languages do not belong to the Niger-Congo family.

Another relevant question is whether linguistic diversity measured at different levels of aggregation proxies for different types of diversity. In particular, we can ask how linguistic diversity measured at varying levels of aggregation relates to religious diversity or ethnic diversity. The data show that the correlation of ELF(15) with ethnic fractionalization from the Atlas Narodov Mira is 0.82 and with the Alesina et al. (2003) measure of ethnic fractionalization it is 0.67. These correlations drop to the 0.35 − 0.40 range when using ELF(1). It therefore does not seem the case that aggregating is equivalent to proxying for ethnic diversity. The same conclusion emerges when analyzing religious diversity. The correlation between ELF(15) and religious fractionalization from the Alesina et al. (2003) dataset is 0.195, and at the level of ELF(1) this correlation drops to 0.098. Obviously it is possible that if one were to aggregate religions or ethnicities, using a procedure similar to the one used here for languages, the correlation with ELF(1) would increase. If so, this would simply reinforce our point: what matters is aggregation and not whether one measures diversity using languages, ethnicities or religions. In the case of religion there is some evidence in favor of this view. Gomes (2010), for example, finds that the correlation between religious diversity and our measure of ELF(1) increases when aggregating religions, although the correlation continues to be low (never surpassing 0.20). Taken together, the evidence suggests that when going from ELF(15) to ELF(1) we are indeed measuring diversity at higher levels of aggregation, rather than proxying for other types of diversity. Our claim that we are capturing deeper cleavages when aggregating is consistent with the fact that the main language families reflect the main demographic divisions in the world (Belle and Barbujani, 2007).

We use these measures to investigate the effects of linguistic diversity at various levels of aggregation on various political and economic outcomes. Our econometric specification builds on the existing literature on the determinants of public goods, civil conflict, redistribution and economic growth, but uses a common set of controls to ensure some consistency across the different dependent variables:

$$y_{it} = \beta X_{it} + \gamma Z_{it} + \delta D_i(j) + \varepsilon_{it} \tag{3}$$

where $y_{it}$ is the dependent variable of interest, $X_{it}$ is a vector of controls specific to outcome $y$, $Z_{it}$ is a vector of controls common to all outcomes $y$, and $D_i(j)$ is either a measure of polarization (POL) or fractionalization (ELF) at aggregation level $j$. The common set of controls $Z_{it}$ includes 1) continent dummy variables for Sub-Saharan Africa, East and Southeast Asia, Latin America and the Caribbean and 2) legal origin dummies from La Porta et al. (1999). The $X_{it}$ controls that are specific to each set of dependent variables are taken from the main contributions from the respective literatures on the determinants of the various outcome

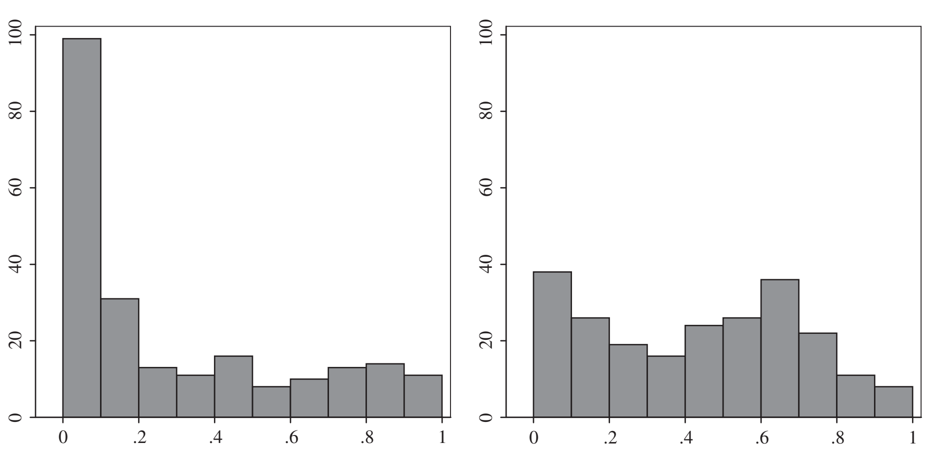


**Fig. 6.** Distributions of POL-1 and POL-15.

**Table 3**
Civil conflict and linguistic diversity (1945–1999). Dependent variable: onset of civil war, logit estimator.

| | (1) ELF(1) | (2) ELF(6) | (3) ELF(15) | (4) POL(1) | (5) POL(6) | (6) POL(15) |
|---|---|---|---|---|---|---|
| ELF (at different levels of aggregation) | 1.157 | 0.040 | −0.165 | | | |
| | [0.545]** | [0.417] | [0.466] | | | |
| POL (at different levels of aggregation) | | | | 0.720 | −0.324 | −0.669 |
| | | | | [0.318]** | [0.427] | [0.483] |
| Lagged civil war | −0.901 | −0.852 | −0.845 | −0.913 | −0.870 | −0.874 |
| | [0.255]*** | [0.257]*** | [0.259]*** | [0.257]*** | [0.265]*** | [0.259]*** |
| Log lagged GDP/cap | −0.616 | −0.613 | −0.627 | −0.615 | −0.619 | −0.610 |
| | [0.140]*** | [0.152]*** | [0.153]*** | [0.140]*** | [0.146]*** | [0.143]*** |
| Log lagged population | 0.311 | 0.293 | 0.300 | 0.310 | 0.300 | 0.298 |
| | [0.068]*** | [0.071]*** | [0.071]*** | [0.067]*** | [0.071]*** | [0.069]*** |
| % mountainous | 0.009 | 0.009 | 0.009 | 0.009 | 0.008 | 0.009 |
| | [0.005]* | [0.005]* | [0.005]* | [0.005]* | [0.005]* | [0.004]** |
| Noncontiguous state dummy | 0.616 | 0.518 | 0.511 | 0.639 | 0.445 | 0.432 |
| | [0.357]* | [0.355] | [0.349] | [0.360]* | [0.365] | [0.361] |
| Oil exporter dummy | 0.621 | 0.724 | 0.746 | 0.601 | 0.742 | 0.783 |
| | [0.246]** | [0.243]*** | [0.245]*** | [0.248]** | [0.241]*** | [0.244]*** |
| New state dummy (1st or 2nd year from independence) | 1.766 | 1.775 | 1.780 | 1.763 | 1.783 | 1.793 |
| | [0.367]*** | [0.369]*** | [0.368]*** | [0.367]*** | [0.370]*** | [0.373]*** |
| Instability dummy (3 years prior) | 0.643 | 0.646 | 0.645 | 0.645 | 0.649 | 0.661 |
| | [0.214]*** | [0.218]*** | [0.217]*** | [0.213]*** | [0.217]*** | [0.217]*** |
| Democracy lagged (Polity 2) | 0.016 | 0.019 | 0.020 | 0.016 | 0.020 | 0.020 |
| | [0.020] | [0.021] | [0.021] | [0.020] | [0.021] | [0.021] |
| French legal origin dummy | 1.383 | 1.453 | 1.553 | 1.396 | 1.571 | 1.686 |
| | [0.648]** | [0.673]** | [0.665]** | [0.646]** | [0.651]** | [0.640]*** |
| UK legal origin dummy | 0.947 | 1.152 | 1.267 | 0.938 | 1.252 | 1.399 |
| | [0.672] | [0.705] | [0.725]* | [0.670] | [0.668]* | [0.648]** |
| Socialist legal origin dummy | 1.126 | 1.237 | 1.282 | 1.133 | 1.299 | 1.408 |
| | [0.694] | [0.707]* | [0.697]* | [0.693] | [0.701]* | [0.692]** |
| Latin America and Caribbean dummy | 0.094 | 0.112 | 0.054 | 0.113 | −0.015 | −0.107 |
| | [0.385] | [0.404] | [0.407] | [0.383] | [0.412] | [0.411] |
| Sub-Saharan Africa dummy | 0.165 | 0.156 | 0.176 | 0.182 | 0.101 | 0.041 |
| | [0.342] | [0.330] | [0.320] | [0.344] | [0.329] | [0.335] |
| East and Southeast Asia dummy | 0.246 | 0.266 | 0.294 | 0.275 | 0.263 | 0.270 |
| | [0.298] | [0.316] | [0.324] | [0.298] | [0.309] | [0.298] |
| Constant | −4.425 | −4.205 | −4.160 | −4.459 | −4.108 | −4.125 |
| | [1.604]*** | [1.627]*** | [1.623]** | [1.599]*** | [1.611]** | [1.604]** |

Robust standard errors, clustered at the level of countries, in parentheses.
All columns involve 5733 observations from 142 countries from 1945 to 1999.
The table reports logit coefficients, not marginal effects.
Results are robust to controlling for the growth of GDP per capita, the growth of GDP per capita lagged, a lagged dichotomous indicator of democracy (instead of the Polity2 index), a squared term for ELF (or POL), a dummy for the new world (and interaction between the new world dummy and diversity), and an anocracy dummy. Results are also robust to introducing ELF(1), ELF(6) and ELF(15) simultaneously (or, similarly, POL(1), POL(6) and POL(15)).
The data is from Fearon and Laitin (2003), except for ELF (authors' calculations from Ethnologue database) and legal origin (from LLSV, 1999).
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

variables under consideration (these are detailed in the following section — for instance, the civil conflict specification includes variables from Fearon and Laitin, 2003).[11] The estimation method also follows these major contributions, though we strive to use cross sectional approaches whenever possible since our measures of linguistic diversity do not vary through time (only in the case of civil conflict onset do we use a pooled panel probit approach as this is the only way to study the determinants of conflict onset).

The main coefficient of interest in our study is $\delta$, the partial correlation of linguistic diversity and each relevant political or economic outcome. Caution should be exercised when interpreting $\delta$ causally. We follow the literature in considering that linguistic diversity is a highly time persistent variable that is likely to be largely historically determined well before our dependent variables are observed. On the other hand, the possibility of reverse causality cannot be entirely ruled out, as discussed in Alesina et al. (2003) and Caselli and Coleman (2010). Thus, we refrain from causal statements, as causality is not the main focus of this paper.

## 3. Linguistic diversity, civil conflict and redistribution

### 3.1. Civil conflict

There is an ongoing academic debate on the relationship between ethnolinguistic diversity and the onset of civil conflict. In a seminal paper, Fearon and Laitin (2003) argued that once measures of income per capita are controlled for, measures of ethnic and religious fractionalization are unrelated to the *onset* of civil conflict. We reexamine this issue using the baseline specification in Fearon and Laitin's study (column 1 of their Table 1, page 84), augmented with our $Z$ controls. Using their data, their estimation method and their dependent variable (the onset of civil conflict), we simply substitute our measures of linguistic heterogeneity for their measure of ethnic fractionalization. Results are presented in Table 3 for selected representative levels of aggregation (1, 6 and 15). The standardized magnitude of the effects of linguistic fractionalization on the probability of conflict onset is displayed graphically *at all levels of aggregation* in Fig. 7.[12]

[11] Further, we use the *broadest* set of controls from these existing studies. Together with the consistent addition of the $Z$ variables in all specifications, this should limit the incidence of omitted variables bias.

[12] The standardized magnitude is computed as the effect of a one standard deviation change in linguistic diversity at each level of aggregation as a percentage of the mean probability of conflict.

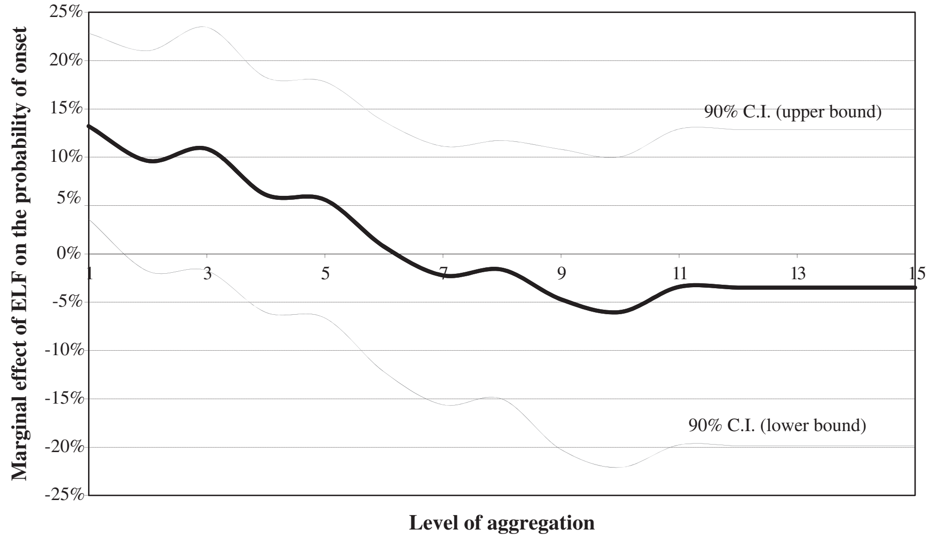


**Fig. 7.** Marginal effect of a one standard deviation increase in ELF (as % of the mean probability of civil conflict onset).

The first and most important observation is that the effect of fractionalization and the corresponding level of statistical significance both fall dramatically and monotonically when the level of aggregation falls. At level $j = 1$ (the most aggregated level), linguistic fractionalization has a coefficient of 1.157 with a t-statistic of 2.12, and the coefficient falls to −0.165 with a t-statistic of 0.35 at level $j = 15$. This pattern is robust to using polarization instead of fractionalization. The second observation is that the coefficient on linguistic diversity is only positive and significant when considering the most aggregated classification of languages — whether for polarization or for fractionalization. The coefficient remains significant at least at the 10% level for most of the robustness tests we conducted — but since the level of significance sometimes falls below 5% we want to be cautious in claiming that there exists a robust relationship even at this level of aggregation. A conservative reading of our results suggests that, to the extent there is a statistically significant link between diversity and civil conflict, it only appears when the relevant cleavages are the deepest (aggregation level 1). In terms of economic magnitude, the estimated effects are far from trivial at aggregation level 1. When evaluating marginal effects at the mean of all the independent variables, a one standard deviation change in linguistic fractionalization (0.173) is associated with an increase in the probability of conflict equal to roughly 13% of this variable's mean (the mean probability of civil war onset is 1.666% in the sample). This effect quickly fades to zero as the level of aggregation falls, as displayed graphically in Fig. 7. The standardized magnitude is similar for polarization at aggregation level 1, and fades to zero also.

The pattern of coefficients across levels of aggregation is robust to a wide range of modifications of the baseline specification for fractionalization: 1) substituting a dichotomous measure of democracy for the continuous one, 2) controlling for intermediate levels of democracy (anocracy), 3) redefining civil wars to only include "ethnic" civil wars (as defined in Fearon and Laitin, 2003), 4) using the Correlates of War definition of civil wars instead of Fearon and Laitin's, 5) controlling for GDP growth and lagged growth and 6) using the incidence of conflict rather than the onset, as Montalvo and Reynal-Querol (2005) did in their study. These results are available upon request. In addition to these checks, which are specific to civil conflict, we also run a number of further robustness tests that are common across the different dependent variables. We discuss those in Section 5.

As shown in Figs. 5 and 6, most countries in the world appear very homogeneous at level 1. Countries that do feature such cleavages tend to coincide with the geographic breakpoints of major linguistic groups, such as in Chad. Our results indicate that ethnolinguistic divisions of this nature may matter for civil conflict, but that more superficial divisions do not. Since there are few countries that feature high levels of diversity at the very aggregated level of linguistic families, civil conflict related to this type of cleavage must be relatively rare.

Where does this leave us in the debate about the role of ethnolinguistic diversity as a determinant of civil wars? On the one hand, for all but one level of aggregation, ethnic diversity does not matter. As was recognized in the past literature, this does not imply that civil conflicts do not often have an ethnic dimension — conditional on having a civil conflict, it may very well be waged along ethnic or linguistic lines (for instance ethnolinguistic differences may help identify combatants, as in the famous Biblical example of the shibboleth). This is compatible with a finding that linguistic diversity is unrelated to the probability of conflict onset. On the other hand, we did find that the significance and magnitude of diversity rises as the level of aggregation increases. To the extent that civil conflict is associated with the "us" versus "them" divide, this result helps clarify that "us" and "them" need to be separated by deep historical and cultural cleavages for these divides to have any claim of affecting the onset of civil conflict. Caselli and Coleman (2010) provide a possible explanation for this finding. They argue that large differences make groups less porous. Changing sides is difficult and costly, thus giving the winning group exclusive access to the gains from the conflict.

### 3.2. Redistribution

A vast literature examines the role of ethnic and linguistic differences as a determinant of the extent of income redistribution. At the microeconomic level, several authors have examined the propensity to redistribute. For instance, Luttmer and Fong (2009) find in an experimental setting that people donate more money to Hurricane Katrina victims when the victims are perceived to be of the same ethnicity as the donor. In another study, Luttmer (2001) reports that "individuals increase their support for welfare spending as the share of local recipients from their own racial group rises", using data from the United States, also suggesting a preference channel. These results are in line with those of Alesina et al. (2001), as well as Alesina and Glaeser (2004), arguing that the U.S. redistributes less than Europe in part because of its greater degree of ethnic heterogeneity.[13]

[13] Our paper uses linguistic heterogeneity rather than ethnic differences, so by our measures the US would look quite more homogeneous than if we focused on ethnicity. This would affect our results in the direction of making it *less* likely to find any effect of diversity on redistribution.

At the cross-country level, results are more mixed. While the preponderance of evidence points to a negative association between ethnolinguistic fractionalization and redistribution, this finding is not always robust to the use of alternative measures of diversity and to the inclusion of controls. For instance, in Alesina et al. (2003), the effect of ethnolinguistic fractionalization on the share of transfers and subsidies to GDP appears sensitive, in terms of statistical significance, to the inclusion of several control variables. This study measures fractionalization using a rather disaggregated classification of ethnic and linguistic groups. In a broad cross-country sample, Desmet et al. (2009) find that linguistic diversity, measured to account for the distance between groups, is negatively associated with redistribution, measured by the share of transfers and subsidies in GDP. However, this result does not hold when measures of diversity do not account for the degree of linguistic distance between groups, suggesting that the depth of linguistic cleavages matters. In a wide variety of settings, ethnolinguistic diversity seems associated with lower redistribution, but what cleavages are more or less relevant to account for these findings has not been determined.

We start from the specification and data in Desmet et al. (2009) to examine what level of linguistic aggregation matters for redistribution, i.e. what are the relevant cleavages. The dependent variable is the average share of transfers and subsidies in GDP between 1985 and 1995. The specification is the one that involves the broadest set of control variables — including GDP per capita, country size and the percentage of the population over 65 (Table 2, column 8 in Desmet et al., 2009), augmented by our $Z$ controls. Table 4 present the results for fractionalization and polarization. The results for both measures are similar, and reveal a striking pattern: linguistic diversity negatively affects redistribution at high levels of aggregation, but the effect declines in magnitude as the level of aggregation falls, and ceases to be statistically significant at the 5% level after aggregation level 5: Fig. 8 displays this pattern, plotting the standardized beta on fractionalization (i.e. the effect of a one standard deviation increase in fractionalization as a fraction of a one standard deviation change in the dependent variable) against the level of aggregation. The effect of ELF(1) is substantial in magnitude, as it equals −9.6% and is significant at the 5% level. It falls to a statistically significant −7.5% for ELF(5) and ceases to be statistically significant thereafter. These results are robust to considering alternative sets of controls, as in Desmet et al. (2009), with the caveat that with a sufficiently restricted set of control variables, the effect of linguistic diversity remains statistically significant even at low levels of aggregation, although significance is greater for high levels of aggregation.

To summarize, we find that for redistribution, as for conflict, the relevant cleavages are those that capture deep ethnolinguistic splits, rather than divisions that are more recent and superficial. Commentators often point out that solidarity does not travel well across groups. We find that solidarity travels without trouble across groups that are separated by shallow gullies, but not across those separated by deep canyons.

## 4. Linguistic diversity, public goods and growth

### 4.1. Public goods

The effect of ethnolinguistic diversity on the provision of public goods raises interesting conceptual issues. On the one hand, public

**Table 4**
Redistribution and linguistic diversity (1985–1995). Dependent variable: transfers and subsidies as share of GDP, least squares estimator.

| | (1) ELF(1) | (2) ELF(6) | (3) ELF(15) | (4) POL(1) | (5) POL(6) | (6) POL(15) |
|---|---|---|---|---|---|---|
| ELF (at different levels of aggregation) | −4.472 | −1.812 | −1.547 | | | |
| | [2.036]** | [1.364] | [1.493] | | | |
| POL (at different levels of aggregation) | | | | −2.749 | −2.134 | −2.056 |
| | | | | [1.211]** | [1.561] | [1.828] |
| Log GDP per capita 1985–95 | 1.274 | 1.173 | 1.198 | 1.270 | 1.256 | 1.232 |
| | [0.557]** | [0.558]** | [0.562]** | [0.558]** | [0.575]** | [0.576]** |
| Log population 1985–95 | 0.265 | 0.335 | 0.352 | 0.288 | 0.284 | 0.284 |
| | [0.300] | [0.302] | [0.306] | [0.296] | [0.296] | [0.300] |
| Population above 65 | 0.877 | 0.893 | 0.902 | 0.879 | 0.884 | 0.892 |
| | [0.137]*** | [0.147]*** | [0.149]*** | [0.138]*** | [0.156]*** | [0.154]*** |
| Small island dummy | −6.237 | −5.766 | −5.749 | −6.125 | −6.038 | −6.075 |
| | [2.336]*** | [2.123]*** | [2.148]*** | [2.319]*** | [2.130]*** | [2.169]*** |
| Latitude | 4.137 | 5.080 | 4.899 | 3.997 | 5.344 | 5.809 |
| | [4.735] | [4.668] | [4.618] | [4.735] | [4.607] | [4.616] |
| UK legal origin dummy | 3.879 | 3.297 | 3.417 | 3.864 | 2.938 | 3.186 |
| | [2.633] | [2.775] | [2.821] | [2.616] | [2.795] | [2.851] |
| French legal origin dummy | 4.526 | 4.472 | 4.480 | 4.494 | 4.106 | 4.234 |
| | [2.692]* | [2.857] | [2.857] | [2.677]* | [2.887] | [2.895] |
| Socialist legal origin dummy | 8.894 | 8.275 | 8.359 | 8.905 | 8.000 | 8.053 |
| | [3.557]** | [3.617]** | [3.628]** | [3.539]** | [3.595]** | [3.609]** |
| Scandinavian legal origin dummy | 5.548 | 4.751 | 4.838 | 5.570 | 4.102 | 4.166 |
| | [2.915]* | [2.990] | [2.983] | [2.897]* | [2.991] | [2.980] |
| Latin America & Caribbean dummy | −1.845 | −2.309 | −2.242 | −1.866 | −2.347 | −2.206 |
| | [1.102]* | [1.168]* | [1.136]* | [1.102]* | [1.205]* | [1.176]* |
| Sub-Saharan Africa dummy | −0.289 | −0.343 | 0.178 | −0.305 | −0.379 | −0.328 |
| | [1.132] | [1.156] | [1.223] | [1.131] | [1.171] | [1.241] |
| East and Southeast Asia dummy | −4.193 | −4.193 | −4.182 | −4.383 | −4.788 | −4.628 |
| | [1.834]** | [1.845]** | [1.839]** | [1.823]** | [1.799]*** | [1.742]*** |
| Constant | −15.691 | −16.099 | −16.730 | −15.910 | −15.263 | −15.518 |
| | [7.933]* | [7.813]** | [7.876]** | [7.867]** | [7.667]** | [7.816]* |

Robust standard errors, in parentheses.
All columns involve 103 country observations.
Results are robust to controlling for religious affiliation (as in Desmet et al., 2009), a squared term for ELF (or POL), a dummy for the new world (and interaction between the new world dummy and diversity). Results are also robust to introducing ELF(1), ELF(6) and ELF(15) simultaneously (or, similarly, POL(1), POL(6) and POL(15)).
The data is from Desmet et al. (2009), except for diversity measures (authors' calculations from Ethnologue database).
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

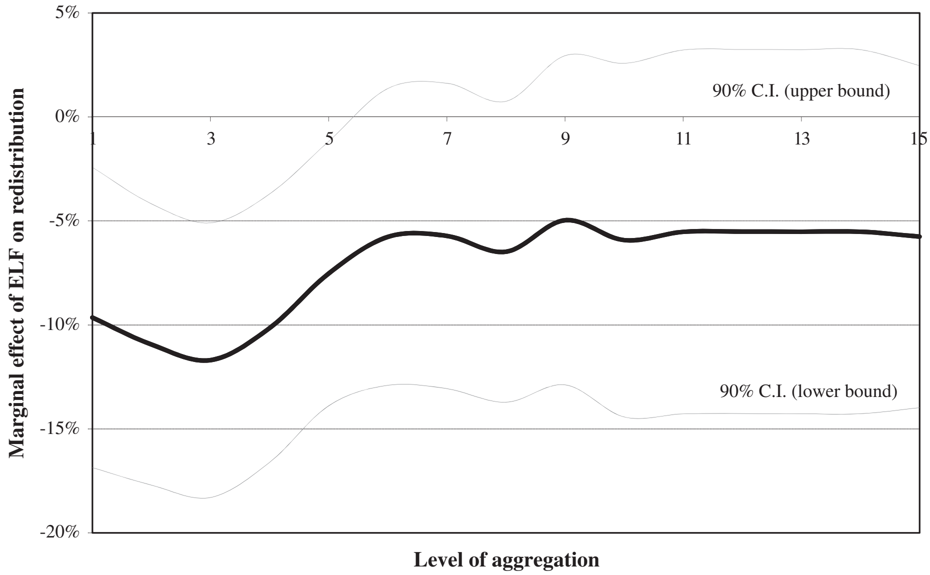


**Fig. 8.** Effect of a one standard deviation increase in ELF on redistribution (as % of standard deviation of redistribution).

goods entail a dimension of redistribution, and differences in preferences may hinder their provision. In this sense, there is an element of conflict of interest when it comes to public goods. On the other hand, free rider problems and coordination failures need to be overcome for the effective provision of public goods. Linguistic diversity may work to affect public goods through both channels.

Several studies have explored the relationship between public goods provision and ethnolinguistic diversity, both across and within countries. In their important study of the cross-national determinants of the quality of government, La Porta et al. (1999, henceforth LLSV) showed that ethnolinguistic fractionalization, measured by an average of five existing indices of fractionalization, generally had a negative impact on several measures of public goods, such as literacy rates, infant mortality, school attainment and infrastructure. Alesina et al. (2003) broadly confirmed these results using new data on ethnic, linguistic and religious fractionalization and polarization, although the results were somewhat sensitive to the chosen measure of diversity and specification. In a within country context, Alesina et al. (1999) showed that across cities, metropolitan areas and urban counties of the United States, greater ethnic diversity was associated with lower provision of education, roads and sewers.

In a more microeconomic context, Habyarimana et al. (2007) report that in a variety of games, co-ethnic participants from a sample of slum dwellers in Kampala, Uganda, play cooperative strategies more so than players from different ethnic groups. This is consistent with findings in Miguel and Gugerty (2005), suggesting that public goods provision is lower in more ethnically diverse locations in Kenya. Other studies include Vigdor (2004) who shows that higher racial, generational and socioeconomic heterogeneity across US counties is associated with lower response rates to the 2000 Census questionnaire, and Banerjee et al. (2005) who, in the context of rural India, find that higher caste and religious fragmentation is associated with lower provision of a wide range of public goods. Although these results are compelling, it is not clear what ethnolinguistic cleavages are most relevant as determinants of public goods provision.

To analyze empirically the relationship between diversity computed at different levels of aggregation and the provision of public goods, we start with the econometric specification and data in LLSV (1999). To minimize the potential for omitted variables bias, we focus on a specification that include a large set of control variables — including legal origins, GNP per capita, latitude and regional dummies. Instead of focusing on a broad set of measures of the quality of government as they did, we focus on the category of dependent variables they label "output of public goods". This includes log infant mortality, log of school attainment, the illiteracy rate, and an index of infrastructure quality.

The results are presented in the top panel of Table 5. For three of the four dependent variables, the statistical significance of the coefficient on ELF rises as the level of aggregation goes from very aggregated ($j=1$) to less aggregated ($j=6$). For two of those the statistical significance further rises when moving to the finest level of aggregation ($j=15$). For illustrative purposes, the evolution of the standardized magnitude of the coefficient on fractionalization as the level of aggregation falls is displayed in Fig. 9 for the illiteracy rate. The standardized magnitude of the coefficient on ELF rises steadily from 5% to 25%. The effects are of the expected signs, namely linguistic fractionalization is negatively associated with school attainment, but positively associated with log infant mortality and the illiteracy rate. There is no significant association with the index of infrastructure quality at any level of aggregation (this was also the case in LLSV). The LLSV measure of ethnolinguistic fractionalization is most highly correlated with ELF(15) — the correlation between the two measures is 0.835, and falls steadily as the level of aggregation rises. Correspondingly, in quantitative terms the magnitude of our estimates is very close to LLSV's when ELF is measured at aggregation level 15. Finally, comparing results on fractionalization and polarization, we see that linguistic fractionalization is a slightly better predictor of public goods than linguistic polarization, but the same pattern emerges with respect to aggregation levels.

In order to investigate whether these results hold up to using a broader set of indicators of public goods provision, the bottom panel of Table 5 considers 6 additional dependent variables, taken from the World Development Indicators (World Bank, 2008). These include measures of health care (hospital beds per person, measles immunization rates for children), measures of access to public services (availability of sanitation services and clean water), and specific measures of infrastructure (road and rail network density).[14] The results show that: 1) for measures of sanitation and clean water, the

[14] We measure the latter as a ratio of kilometers per 1000 inhabitants, but the results are unchanged when using kilometers per square kilometer of land area instead. Results are available upon request.

**Table 5**
Public goods and linguistic diversity, OLS estimates (dependent variable listed in the leftmost column).

| | ELF(1) | ELF(6) | ELF(15) | POL(1) | POL(6) | POL(15) | # of obs. | Adj-R2 min | Adj-R2 max |
|---|---|---|---|---|---|---|---|---|---|
| *Output of public goods (from LLSV, 1999)* | | | | | | | | | |
| Log infant mortality | 0.466<br>[0.199]** | 0.413<br>[0.122]*** | 0.512<br>[0.113]*** | 0.248<br>[0.107]** | 0.379<br>[0.113]*** | 0.501<br>[0.118]*** | 173 | 0.78 | 0.8 |
| Log of school attainment | −0.143<br>[0.215] | −0.261<br>[0.137]* | −0.203<br>[0.142] | −0.050<br>[0.108] | −0.097<br>[0.109] | −0.003<br>[0.145] | 101 | 0.74 | 0.76 |
| Illiteracy rate | 6.891<br>[8.314] | 20.662<br>[5.166]*** | 17.784<br>[5.086]*** | 3.506<br>[4.746] | 18.174<br>[5.316]*** | 12.674<br>[6.344]** | 119 | 0.48 | 0.54 |
| Infrastructure quality index | −0.167<br>[0.819] | −0.095<br>[0.452] | 0.229<br>[0.474] | −0.133<br>[0.496] | −0.077<br>[0.427] | −0.224<br>[0.463] | 59 | 0.77 | 0.78 |
| *Additional measures of public goods* | | | | | | | | | |
| Hospital beds (per 1000 people) | −1.061<br>[1.000] | −0.595<br>[0.608] | 0.058<br>[0.637] | −0.452<br>[0.577] | −0.227<br>[0.760] | −0.759<br>[0.924] | 170 | 0.46 | 0.46 |
| Measles immunization rates (% of children 12–23 months) | −23.203<br>[5.831]*** | −17.138<br>[3.805]*** | −13.152<br>[3.896]*** | −12.194<br>[3.439]*** | −12.545<br>[3.587]*** | −7.613<br>[4.172]* | 169 | 0.5 | 0.56 |
| Improved sanitation facilities (% of population with access) | −12.536<br>[8.487] | −16.374<br>[5.120]*** | −23.398<br>[5.217]*** | −5.481<br>[4.852] | −15.890<br>[5.056]*** | −10.908<br>[5.700]* | 147 | 0.72 | 0.76 |
| Improved water source (% of population with access) | −13.896<br>[7.195]* | −6.985<br>[4.651] | −13.995<br>[4.074]*** | −7.790<br>[4.124]* | −7.135<br>[4.162]* | −3.920<br>[4.726] | 157 | 0.59 | 0.62 |
| Road network density (km per 1000 inhabitants) | −4.426<br>[4.338] | −3.159<br>[1.812]* | −1.870<br>[2.101] | −2.842<br>[2.259] | −1.505<br>[2.353] | −3.386<br>[2.509] | 151 | 0.27 | 0.28 |
| Rail network density (km per 1000 inhabitants) | 0.299<br>[0.311] | −0.055<br>[0.140] | 0.151<br>[0.164] | 0.140<br>[0.144] | 0.085<br>[0.183] | −0.107<br>[0.161] | 89 | 0.33 | 0.35 |

Robust standard errors in brackets.
For all regressions, the specification includes the following controls: Socialist legal origin dummy, French legal origin dummy, German legal origin dummy, Scandinavian legal origin dummy, East and Southeast Asia dummy, Sub-Saharan Africa dummy, Latin America and Caribbean dummy, latitude and log GNP per capita. The table reports coefficient estimates on diversity indices at various levels of aggregation, in regressions where the dependent variable is the one listed in the leftmost column. The data is from LLSV (1999) and the World Bank (2008), except for diversity measures (authors' calculations from Ethnologue database).
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

effect of fractionalization becomes more statistically significant as the level of aggregation falls; 2) for measures of health services, the effect of ELF remains consistently significant for the measles immunization rate across aggregation levels, but is insignificant for hospital beds; 3) infrastructure measures are unaffected by fractionalization whatever the level of aggregation; and 4) fractionalization is usually a better predictor of public goods provision than polarization.

To summarize, across a wide range of measures of public goods, we broadly confirm results from the literature referenced above: ethnolinguistic diversity is negatively related to public goods provision. Public goods share both dimensions of conflict of interest (different tastes for public goods) and coordination (need to overcome free rider problems). Correspondingly we find that for several of the measures of public goods provision (infant mortality, measles immunization rates), diversity at all levels matters statistically. However, we also find that, broadly speaking, measures of fractionalization based on finer classifications of linguistic groups tend to matter more than those based on deep cleavages only. In contrast with redistribution, for which only deep splits were important, even relatively recent and shallow linguistic cleavages seem sufficient to hinder the provision of public goods.

### 4.2. Growth

In recent years, scholars have focused on ethnolinguistic diversity as a determinant of economic performance. Easterly and Levine (1997) argue that ethnic diversity, measured by an index of fractionalization, may account for much of Africa's growth tragedy. These cross-country results were reinforced and extended in Alesina et al. (2003). In particular, the latter paper showed that linguistic diversity per se, not just ethnic diversity, has a significantly negative effect on per capita income growth in a panel of countries, so that both ethnic and linguistic diversity are alternative ways to capture a broader concept of cultural heterogeneity. In addition, the paper found that fractionalization measures were more robust predictors of growth than polarization measures, an issue we revisit below.[15]

To examine the impact of linguistic cleavages at various levels of aggregation, we start from a growth specification derived from the augmented Solow model, which includes the investment rate, a measure of human capital (the number of years of schooling in the adult population aged 25 and over — results do not change when using a flow measure such as the secondary school enrollment rate), and a measure of population growth. In addition, we include measures of market size used in Ades and Glaeser (1999) and Alesina et al. (2000), namely the ratio of imports plus exports to GDP, the log of population, and the interaction between these two variables. Finally, the regression includes our full set of regional and legal origins dummy variables. The timespan extends from 1970 to 2004, and the regressions are run on a single cross-section of 100 countries.[16] Coefficient estimates are shown in Tables 6. The standardized betas on fractionalization are displayed in Fig. 10. The effect of ELF becomes greater in magnitude and more significant when the level of aggregation falls, with the standardized beta equal to 2.46% for $j=1$,

[15] For a survey of the empirical literature on ethnolinguistic diversity and economic performance at the level of countries, cities and villages in developing countries, see Alesina and La Ferrara (2005). This is related to a more microeconomic approach highlighting the costs and benefits of cultural and linguistic diversity within teams or organizations. See for instance Cremer et al. (2007), Lazear (1999), and Prat (2002). While at the cross-country level the empirical results point to a negative relationship between ethnolinguistic diversity and growth, the findings are more contrasted at the within-firm level, with some studies pointing to positive effects of diversity. At the cross-city level in the U.S., Ottaviano and Peri (2006) also point to a positive effect of cultural diversity on the productivity of U.S. natives.

[16] Hauk and Wacziarg (2009), using simulations based on the Solow model, show that running growth regressions using OLS on a single cross-section of countries provides the least biased coefficients on the determinants of growth in the presence of a multiplicity of data problems, such as regressor endogeneity, cross-country heterogeneity and measurement error.

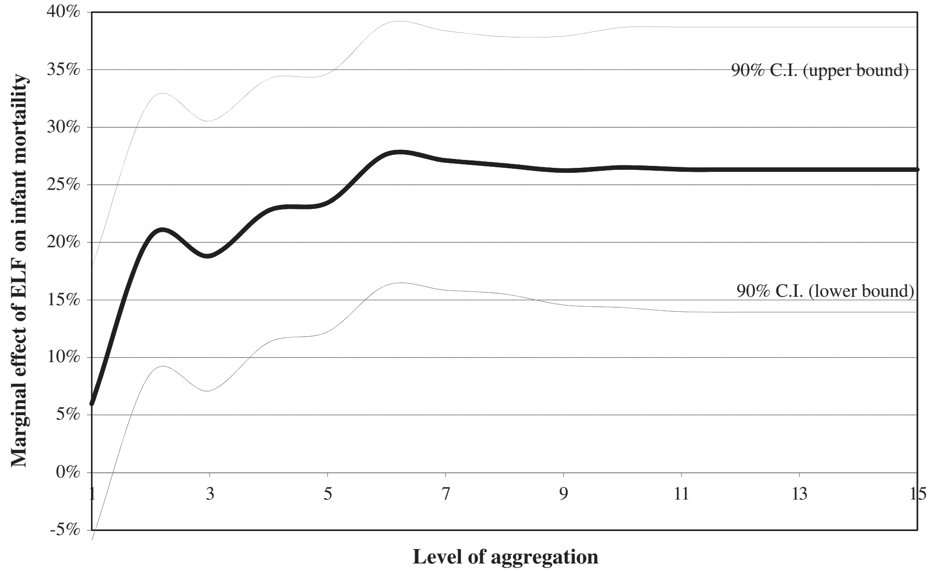


**Fig. 9.** Effect of a one standard deviation increase in ELF on illiteracy rate (as % of standard deviation of illiteracy rate).

rising steadily as the level of aggregation falls, and settling around 22% at levels $j=11$ and higher. The level of statistical significance also rises with $j$, with ELF becoming significant at the 95% confidence level at aggregation level $j=5$. Consistent with findings in Alesina et al. (2003), polarization measures appear largely unrelated to growth.

To illustrate the quantitative importance of ethnolinguistic diversity for economic growth, we analyze the case of the world's two most populous countries, China and India. Both have experienced high growth rate in recent decades, although India continues to lag behind its East Asian neighbor. According to the Penn World Tables (version 6.2), over the period 1970–2004 China averaged an annual growth rate in real GDP per capita of 6.93%, compared to 2.71% in India. China is also much less linguistically diverse than India: at aggregation level $j=15$, India's ELF index is 0.93, while China's is 0.49. Fitting the regression model in Column 3 of Table 6 to these two datapoints, we calculate that 14.72% of the growth difference between India and China over 1970–2004 is accounted for by differences between these two countries in ELF(15). Thus, taken at face value linguistic fractionalization at a high level of disaggregation can account for a substantial portion of the growth differential between India and China. In contrast, the share of the growth differential accounted for by differences in ELF(1) is only 2.11%, and is not statistically significant.

In additional empirical work available upon request, we extended these results in various directions. First, we showed that the same pattern of increasing growth effects of fractionalization as aggregation falls are obtained when using exactly the specification, estimation method and data in Easterly and Levine (1997).[17] Second, we showed that, in the augmented Solow model specification above, the same pattern obtains in a random effects or SUR panel context, instead of a single cross-section of time averages.

To summarize, these results show that to capture the relevant cleavages that affect economic growth, focusing only on deep cleavages is not sufficient. Instead, one needs to take into account finer distinctions across linguistic groups. This does not imply that deep cleavages do not contribute to negatively affecting growth, as these deep cleavages do contribute to diversity at lower levels of aggregation: fractionalization measured at low levels of aggregation is affected by both deep and shallow cleavages. Fractionalization measured at high levels of aggregation ignores many of the shallower, yet relevant, cleavages, and therefore amounts to a noisy measure to predict the effect of diversity on growth.

## 5. Robustness checks

In this section we conduct a series of robustness checks common to each set of dependent variables considered above. The corresponding econometric estimates are available upon request.

In a first exercise, we examine whether the effects of linguistic diversity might be nonlinear, by adding a quadratic term in diversity to the regressions. Ashraf and Galor (2010) have argued that the relationship between genetic diversity and development may be hump-shaped: too little or too much diversity may be detrimental to growth. Consistent with their findings, in the case of growth there is indeed some evidence of such a hump-shaped relationship, and consistent with our results this is true only at low levels of aggregation. In the case of child mortality there is also a slight nonlinearity, although the overall effect is always negative. For none of the other dependent variables do we find any evidence of a nonlinear relationship. In any case, none of these results change the basic picture of which levels of aggregation matter most for different dependent variables.

In a second robustness check, we examine whether the results differ between the New World and the Old World. This is important because while linguistic cleavages largely reflect historical population splits of varying degrees of depth, in the New World language replacement may have severed or weakened the link between linguistic cleavages and historic divisions. For instance, in the US both Blacks and Whites speak English, and in Latin America people of Amerindian descent often (though by no means always) speak European languages. Consistent with this fact, when splitting up the sample into the Old World and the New World – defined as the Americas and Oceania – we find the relationship between diversity and the different dependent variables to be less robust for the New World. This may partly reflect the substantially smaller sample of countries in the New World. An alternative specification to test this hypothesis is to keep the whole sample, but to include a dummy

[17] In addition, using the Easterly and Levine (1997) setup, the basic pattern holds, or is even reinforced, when: 1) controlling for political assassinations, 2) controlling for political assassinations plus financial depth, the black market premium, and the fiscal surplus to GDP ratio, and 3) controlling for all of the above plus the number of coups d'état, the number of revolutions, a dummy for civil wars and a measure of political rights (Gastil's index of democracy).

**Table 6**
Growth and linguistic diversity. (Augmented Solow specification, OLS estimator, 1970–2004 panel).

| | (1) ELF(1) | (2) ELF(6) | (3) ELF(15) | (4) POL(1) | (5) POL(6) | (6) POL(15) |
|---|---|---|---|---|---|---|
| ELF (at various levels of aggregation) | −0.279 | −1.056 | −1.412 | | | |
| | [0.933] | [0.576]* | [0.473]*** | | | |
| POL (at various levels of aggregation) | | | | −0.205 | −0.667 | 0.250 |
| | | | | [0.537] | [0.531] | [0.573] |
| Log initial real per capita GDP | −1.209 | −1.153 | −1.143 | −1.214 | −1.208 | −1.203 |
| | [0.234]*** | [0.219]*** | [0.230]*** | [0.234]*** | [0.234]*** | [0.231]*** |
| Investment share of GDP | 0.043 | 0.032 | 0.032 | 0.044 | 0.038 | 0.043 |
| | [0.034] | [0.031] | [0.031] | [0.034] | [0.032] | [0.034] |
| Avg. schooling years in total population aged 25+ | 0.211 | 0.193 | 0.186 | 0.212 | 0.233 | 0.208 |
| | [0.098]** | [0.092]** | [0.095]* | [0.097]** | [0.095]** | [0.096]** |
| Growth of population | −0.599 | −0.524 | −0.508 | −0.595 | −0.526 | −0.634 |
| | [0.196]*** | [0.204]** | [0.184]*** | [0.198]*** | [0.214]** | [0.194]*** |
| Log population | 0.197 | 0.220 | 0.225 | 0.197 | 0.194 | 0.198 |
| | [0.151] | [0.147] | [0.145] | [0.150] | [0.150] | [0.151] |
| Interaction between openness and log population | −0.005 | −0.005 | −0.004 | −0.005 | −0.005 | −0.005 |
| | [0.002]** | [0.002]** | [0.002]** | [0.002]** | [0.002]** | [0.002]** |
| Openness (imports + exports over GDP) | 0.046 | 0.045 | 0.041 | 0.046 | 0.044 | 0.046 |
| | [0.017]*** | [0.017]** | [0.017]** | [0.017]*** | [0.018]** | [0.017]*** |
| Latin America and Caribbean dummy | −0.725 | −0.990 | −1.090 | −0.726 | −0.909 | −0.673 |
| | [0.309]** | [0.289]*** | [0.278]*** | [0.306]** | [0.286]*** | [0.311]** |
| Sub-Saharan Africa dummy | −1.531 | −1.656 | −1.291 | −1.536 | −1.585 | −1.491 |
| | [0.528]*** | [0.542]*** | [0.516]** | [0.529]*** | [0.523]*** | [0.526]*** |
| East and Southeast Asia dummy | 1.329 | 1.461 | 1.524 | 1.316 | 1.195 | 1.373 |
| | [0.639]** | [0.623]** | [0.645]** | [0.643]** | [0.660]* | [0.651]** |
| British legal origin dummy | 0.286 | 0.377 | 0.491 | 0.291 | 0.367 | 0.214 |
| | [0.402] | [0.415] | [0.417] | [0.401] | [0.418] | [0.384] |
| French legal origin dummy | 0.448 | 0.504 | 0.557 | 0.447 | 0.583 | 0.407 |
| | [0.441] | [0.434] | [0.439] | [0.440] | [0.433] | [0.388] |
| Socialist legal origin dummy | 0.290 | 0.312 | 0.241 | 0.286 | 0.401 | 0.269 |
| | [0.439] | [0.469] | [0.481] | [0.440] | [0.529] | [0.416] |
| German legal origin dummy | −0.066 | 0.016 | −0.021 | −0.070 | 0.181 | −0.116 |
| | [0.570] | [0.565] | [0.566] | [0.570] | [0.609] | [0.574] |
| Constant | 8.911 | 8.623 | 8.595 | 8.951 | 8.889 | 8.844 |
| | [2.410]*** | [2.336]*** | [2.347]*** | [2.421]*** | [2.391]*** | [2.389]*** |
| R-squared | 0.54 | 0.56 | 0.57 | 0.54 | 0.55 | 0.54 |

Robust standard errors in parentheses.
All columns involve 100 countries.
Investment, schooling, population growth, log population and openness are entered as period averages; log initial per capita income is for 1970. The data on income per capita, income growth, population, population growth, openness and investment are from the Penn World Tables, version 6.2 (Heston et al., 2006). The data on human capital is from Barro and Lee (2000). The diversity measures are based on the authors' calculations using the Ethnologue database.

* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

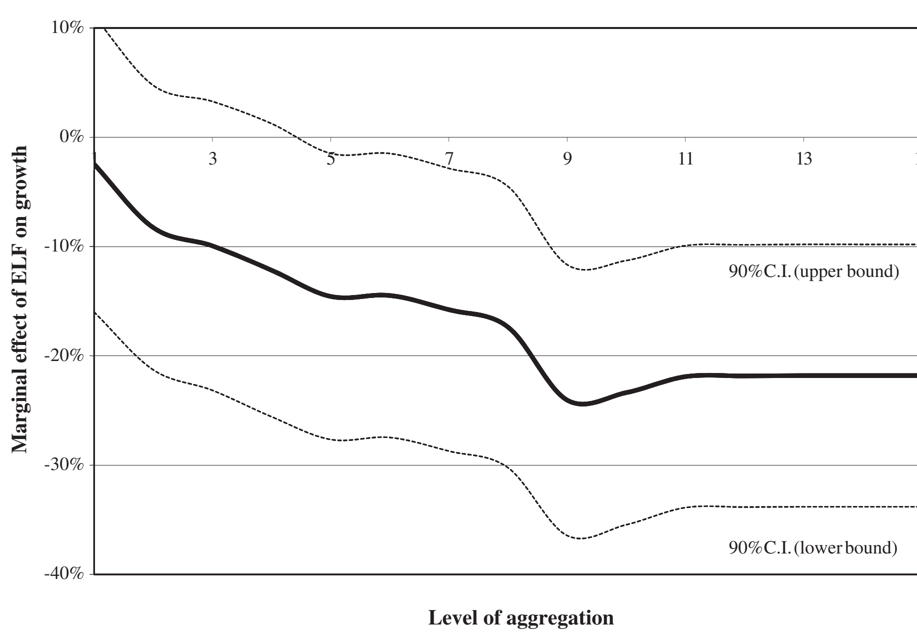


**Fig. 10.** Marginal effect of a standard deviation increase in ELF as a % of a standard deviation of growth.

variable for the New World and an interaction term between that dummy and linguistic diversity. When doing so, we find that the interaction term is not significantly different from zero.

In a third exercise, we run a horserace between measures of linguistic diversity at various levels of aggregation, by entering levels $j = 1, 6, 15$ all at once in the regressions. Doing so confirms our results. In the case of civil conflict and redistribution, only ELF(1) and POL(1) continue to be statistically significant at the 5% level, with the measures at lower levels of aggregation being statistically insignificant. In contrast, in the case of growth and public goods the lower levels of aggregation tend to matter most, except for the case of illiteracy where diversity measured at an intermediate level of aggregation is most significant.

In a fourth robustness check, we ran horseraces between fractionalization and polarization. It is worth pointing out that these two measures of linguistic diversity tend to be highly correlated, particularly at very high levels of aggregation ($j = 1$).[18] Consistent with this observation, we find that for outcomes for which diversity at high levels of aggregation (deep cleavages) matter most (conflict and redistribution), multicollinearity is severe, and both ELF and POL become insignificant. In contrast, for outcomes where finer linguistic disaggregation matters most (growth and public goods) and where ELF and POL are less collinear, we find that ELF wins out in a horserace with POL.

Finally, we reran all the regressions using Approach 2 for the construction of linguistic cleavages (see Fig. 4 and the conceptual discussion in Section 2.2.2). The results are virtually unchanged. The reason is straightforward. As discussed in the theoretical section, Approach 1 and Approach 2 give identical measures of diversity at both extremes, i.e., for $j = 1$ and $j = 15$. Since in the cases of civil conflict and redistribution diversity measured at one extreme (the highest levels of aggregation) matters most and in the cases of public goods and economic growth diversity measured at the other extreme (the lowest levels of aggregation) matters most, the results are bound not to change substantially.

To summarize, these various robustness checks do not modify our basic result that deep cleavages matter most for conflict and redistribution, while more superficial linguistic cleavages are sufficient to hinder growth and the provision of public goods.

## 6. Conclusion

In this paper, we have uncovered new evidence on the relationship between ethnolinguistic diversity and a range of political economy outcomes, such as the onset of civil wars, the extent of redistribution, the provision of public goods, and economic growth. We sought to identify the relevant linguistic cleavages to explain variation in these outcomes. We let the data tell us whether deep cleavages, originating at an earlier time in history, are more or less important than more superficial cleavages that have arisen more recently. Doing so, we departed from the common approach relying on arbitrary definitions of what constitutes a relevant ethnolinguistic group. Our results carry several lessons. When it comes to civil conflict and redistribution, deeper cleavages tend to matter more. In contrast, for economic growth and public goods, we found that diversity measured using only deep cleavages is not sufficient to predict significant differences in growth. Instead, measures based on more disaggregated classifications of linguistic groups, capturing finer distinctions between languages, are important correlates of growth and public goods provision both in terms of statistical significance and in terms of economic magnitude.

How should we interpret these results? We have shown that the type of cultural diversity that matters for outcomes involving conflicts of interest – civil wars, redistribution – is different from the type of diversity that matters for outcomes that entail issues of efficiency and coordination, such as growth. When it comes to conflict and redistribution, preferences are of the essence. The willingness to settle disputes or to transfer resources across a cultural divide depends on how deep the divide happens to be. Deep cleavages that go back thousands of years appear to be related with more conflicts of interest, compared to more superficial cleavages.

In contrast, economic growth requires that groups have the ability to coordinate, interact and organize in networks of production, knowledge and trade. This ability is affected by linguistic divisions. In India, for instance, the degree of integration between regions is likely hindered by linguistic barriers — even linguistic barriers separate relatively similar linguistic groups such as Hindi and Gujarati speakers. Coordination, integration and more generally the ability to form knowledge, production and trading networks is hampered as soon as linguistic differentiation prevents interactions between groups, and this can occur between groups that are relatively similar linguistically.

The case of public goods shares characteristics of both types of outcomes: public goods are inherently redistributive in nature, and their provision depends on differences in preferences among participants. At the same time, the provision of public goods requires coordination and interactions, that even superficial cleavages might hamper. We found that, much as in the case of growth, for a wide array of measures of public goods, fine distinctions between linguistic groups matter to hinder their provision. Even when cleavages are shallow, a country may fail to have well-functioning public services, not necessarily because people are unwilling to redistribute, but because of coordination failures.

Future work should seek to better understand the theoretical mechanisms that account for the contrasting findings between conflict and redistribution on the one hand, and growth and public goods on the other hand. In particular, clarifying the differing effects of diversity on efficiency and coordination (where fine distinctions seem to matter more) and preferences (where coarse distinctions seem of the essence) may help account for our results.

Finally, we have focused on linguistic diversity, as a measure of a broader concept of ethnolinguistic heterogeneity, and even more broadly as a proxy for cultural diversity. One advantage of focusing on languages is that linguistic distinctions are quite objective: it is easier to judge whether two populations speak different languages than to decide whether two populations belong to different ethnicities, a more amorphous concept (precisely for this reason, ethnic categorizations are often based on linguistic divisions, particularly for Africa). Another advantage is that data on linguistic divisions, particularly in the form of trees, is more readily available than data on the genealogical structure of ethnic groups within countries. There are, however, drawbacks to focusing on languages: to the extent that linguistic divisions are imperfect measures of the source of diversity that matters most, this should lead to downward bias on the estimates of the effect of diversity on political economy outcomes. In principle, the methodology we have developed for linguistic trees should be applicable to other kinds of differences between populations. With advances in population genetics, population phylogenies have become more widely available. Although this data is not yet available in a single format such as the *Ethnologue* for languages, applying our method to genetic data could lead to fruitful advances in the study of the political economy of cultural diversity.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.jdeveco.2011.02.003.

[18] Table 2 shows that ELF(1) and POL(1) bear a 0.988 correlation with each other, falling to 0.555 for ELF(15) and POL(15).

## References


Ades, A., Glaeser, E.L., 1999. Evidence on growth, increasing returns, and the extent of the market. Quarterly Journal of Economics vol. 114 (no. 3), 1025–1045 August.

Alesina, A., Glaeser, E.L., 2004. Fighting Poverty in the US and Europe: A World of Difference. Oxford University Press, Oxford.

Alesina, A., La Ferrara, E., 2005. Ethnic diversity and economic performance. Journal of Economic Literature vol. 43, 762–800 September.

Alesina, A., Baqir, R., Easterly, W., 1999. Public goods and ethnic divisions. Quarterly Journal of Economics vol. 114 (no. 4), 1243–1284 November.

Alesina, A., Spolaore, E., Wacziarg, R., 2000. Economic integration and political disintegration. The American Economic Review vol. 90 (no. 5), 1276–1296 December.

Alesina, A., Glaeser, E., Sacerdote, B., 2001. Why doesn't the United States have a European-style welfare state? Brookings Papers on Economic Activity 2001 (2), 187–254.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R., 2003. Fractionalization. Journal of Economic Growth vol. 8 (no. 2), 155–194 June.

Ashraf, Q., Galor, O., 2010. The out of Africa hypothesis, human genetic diversity and comparative development. Working Paper. Brown University.

Banerjee, A., Iyer, L., Somanathan, R., 2005. History, social divisions, and public goods in rural India. Journal of the European Economic Association 3 (2–3), 639–647.

Barro, R.J., Lee, J.W., 2000. International data on educational attainment: updates and implications. Harvard CID Working Paper No. 42. . April.

Belle, E.M.S., Barbujani, G., 2007. Worldwide analysis of multiple microsatellites: language diversity has a detectable influence on DNA diversity. American Journal of Physical Anthropology 133 (4), 1137–1146.

Bossert, W., D'Ambrosio, C., La Ferrara, E., 2011. A Generalized Index of Fractionalization. Economica 78 (312), 723–750.

Caselli, F., Coleman II, W.J., 2010. On the theory of ethnic conflict. Working Paper. Duke University.

Cavalli-Sforza, L.L., Piazza, A., Menozzi, P., Mountain, J., 1988. Reconstruction of human evolution: bringing together genetic, archaeological and linguistic data. Proceedings of the National Academy of Sciences of the United States of America 85, 6002–6006.

Cremer, J., Garicano, L., Prat, A., 2007. Language and the theory of the firm. Quarterly Journal of Economics vol. 122 (no. 1), 373–407 February.

Darwin, C., 1902. On the Origin of Species by Means of Natural Selection, or, The Preservation of Favoured Races in the Struggle for Life, 6th edition. Grant Richards, London.

Desmet, K., Ortuño Ortín, I., Weber, S., 2009. Linguistic diversity and redistribution. Journal of the European Economic Association 7 (6), 1291–1318.

Dixon, R.M.W., 1997. The Rise and Fall of Languages. Cambridge University Press, Cambridge.

Easterly, W., Levine, R., 1997. Africa's growth tragedy: policies and ethnic divisions. Quarterly Journal of Economics vol. 112 (no. 4), 1203–1250 November.

Emeneau, M., Anwar, D., 1980. Language and Linguistic Area: Essays by Murray B. Emeneau. Stanford University Press, Palo Alto.

Esteban, J.M., Ray, D., 1994. On the measurement of polarization. Econometrica 62 (4), 819–851.

Ethnologue, 2005. Ethnologue: Languages of the World, 15th Edition. SIL International. www.ethnologue.com.

Fearon, J., 2003. Ethnic and cultural diversity by country. Journal of Economic Growth vol. 8 (no. 2), 195–222 June.

Fearon, J., Laitin, D., 2003. Ethnicity, insurgency, and civil war. American Political Science Review vol. 97 (no. 1), 75–90 February.

Gomes, Joseph F., 2010. The political economy of religious diversity. Working Paper. Universidad Carlos III.

Gray, R.D., Atkinson, Q.D., 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature vol. 426 (27), 435–439 November.

Greenberg, J.H., 1956. The measurement of linguistic diversity. Language vol. 32 (no. 1), 109–115 Jamuary.

Greenberg, J.H., 1987. Language in the Americas. Stanford University Press, Palo Alto.

Habyarimana, J., Humphreys, M., Posner, D., Weinstein, J., 2007. Why does ethnic diversity undermine public goods provision? American Political Science Review vol. 101 (no. 4), 709–725 November.

Hauk, W., Wacziarg, R., 2009. A Monte Carlo study of growth regressions. Journal of Economic Growth vol. 14 (no. 2), 103–147 June.

Heston, A., Summers, R., Aten, B., 2006. Penn World Table Version 6.2. Center for International Comparisons of Production, Income and Prices. University of Pennsylvania. September.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R., 1999. The quality of government. Journal of Law, Economics, and Organization 15 (1), 222–279.

Lazear, E., 1999. Culture and language. Journal of Political Economy vol. 107 (no. 6), S95–S126 December.

Lithgow, D., 1973. Language change on Woodlark Island. Oceania vol. 44 (no. 2), 101–108 December.

Luttmer, E.F.P., 2001. Group loyalty and the taste for redistribution. Journal of Political Economy vol. 109 (no. 3), 500–528 June.

Luttmer, E.F.P., Fong, C., 2009. What determines giving to Hurricane Katrina victims? Experimental evidence on racial group loyalty. American Economic Journal: Applied Economics vol. 1 (no. 2), 64–87 April.

Miguel, E., Gugerty, M.K., 2005. Ethnic diversity, social sanctions, and public goods in Kenya. Journal of Public Economics 89 (11), 2325–2368.

Montalvo, J.G., Reynal-Querol, M., 2005. Ethnic polarization, potential conflict and civil war. The American Economic Review vol. 95 (no. 3), 796–816 June.

Ottaviano, G.M., Peri, G., 2006. The economic value of cultural diversity: evidence from U.S. cities. Journal of Economic Geography vol. 6 (no. 1), 9–44 January.

Posner, D., 2003. The colonial origins of ethnic cleavages: the case of linguistic divisions in Zambia. Comparative Politics 35 (2), 127–146.

Posner, D., 2005. Institutions and Ethnic Politics in Africa. Cambridge University Press, Cambridge (UK).

Prat, A., 2002. Should a team be homogeneous? European Economic Review 46 (7), 1187–1207.

Scarritt, J.R., Mozaffar, S., 1999. The specification of ethnic cleavages and ethnopolitical groups for the analysis of democratic competition in contemporary Africa. Nationalism and Ethnic Politics vol. 5 (no. 1), 82–117 March.

Spolaore, E., Wacziarg, R., 2009. The diffusion of development. Quarterly Journal of Economics vol. 124 (no. 2), 469–529 May.

Vigdor, J.L., 2004. Community composition and collective action: analyzing initial mail response to the 2000 census. The Review of Economics and Statistics 86 (1), 303–312.

World Bank, 2008. World Development Indicators 2008. World Bank, Washington, DC.

Young, C., 1976. The Politics of Cultural Pluralism. University of Wisconsin Press, Madison.