LINGUISTIC DIVERSITY AND REDISTRIBUTION

Klaus Desmet
Universidad Carlos III

Ignacio Ortuño-Ortín
Universidad Carlos III

Shlomo Weber
Southern Methodist University

Abstract
This paper investigates the effect of linguistic diversity on redistribution in a broad cross-section of countries. We use the notion of "linguistic distances" and show that the commonly used fractionalization index, which ignores linguistic distances, yields insignificant results. However, once distances between languages are accounted for, linguistic diversity has both a statistically and economically significant effect on redistribution. With an average level of redistribution of 9.5% of GDP in our data set, an increase by one standard deviation in the degree of diversity lowers redistribution by approximately one percentage point. We also demonstrate that other measures, such as polarization and peripheral heterogeneity, provide similar results when linguistic distances are incorporated. (JEL: D6, D74, H5, Z10)

1. Introduction

The literature has long argued that cultural diversity reduces government transfers[1] and that altruistic attitudes are more prevalent within homogenous groups than across ethnically or culturally diverse groups.[2] If, as posited by Becker (1957), individuals have stronger feelings of empathy toward their own group,

The editor in charge of this paper was Roberto Perotti.

Acknowledgments: The authors wish to thank the financial support of the Comunidad de Madrid (grant 06/HSE/0157/2004) and of the Fundación BBVA. We are grateful to Lola Collado, Jim Fearon, Anna Rubinchik, Ekaterina Zhuravskaya, and especially Romain Wacziarg for valuable comments and suggestions. We also thank the editor, Roberto Perotti, for further comments, and Brian Homoleski from SIL International for providing us with the Ethnologue data. Desmet and Weber are members of the CEPR.

E-mail addresses: Desmet: klaus.desmet@uc3m.es; Ortuño-Ortín: iortuno@eco.uc3m.es; Weber: weber@mail.smu.edu

1. See Alesina et al. (2003), La Porta et al. (1999), Easterly and Levine (1997), and Luttmer (2001) for empirical studies; and Caselli and Coleman (2006), Fernández and Levy (2007), Lee and Roemer (2004), and Lind (2007) for theoretical analysis of the relation between diversity and redistribution. For a more general survey on diversity and different measures of economic performance, see Alesina and La Ferrara (2005).

2. See, for example, Vigdor (2004).

Journal of the European Economic Association December 2009 7(6):1291–1318

it is not surprising that the U.S., where there is a strong racial component to the income distribution and the poor tend to be viewed as "other," exhibits lower levels of redistribution than Western European countries, where the poor are often seen as "unlucky" (Alesina, Glaeser, and Sacerdote 2001). One must note, however, that broader cross-country studies have typically failed to pick up a statistically significant relation between cultural diversity and transfers (Alesina et al. 2003).

Most empirical economic studies of diversity use the so-called ELF or fractionalization index, which measures the probability of two randomly selected individuals in society belonging to different ethnolinguistic groups. As with Shannon's diversity index (Shannon, 1949), it fails to take into account the degree of distinctiveness between different groups. Compare, for instance, Andorra, where roughly half of the population speaks Catalan and the other half speaks Spanish (two similar Romance languages), and Belgium, where about 60% of the population speaks Dutch and the other 40% speaks French (a Germanic and a Romance language). Although one would probably think of Belgium being linguistically more diverse than Andorra, according to the fractionalization index Andorra is the more diverse one of the two.

The main contribution of this paper is the incorporation of distances between different groups when measuring diversity. By accounting for distances, we adopt the Becker (1957) view: Not only do individuals prefer their own group, but the degree of their dislike of other groups depends on how different the groups are. An appropriate analysis of ethnolinguistic diversity should therefore, as suggested by Caselli and Coleman (2006) and Alesina and La Ferrara (2005), take into account the dissimilarity between groups.[3] After incorporating distances, we revisit the link between linguistic diversity and redistribution in a wide cross-section of countries.

Naturally, with nearly 7,000 known living languages and about 200 countries,[4] multilingual societies are a common feature across the globe. The underlying motivation for our work is that sorting out linguistic issues can be quite challenging—"language does not lend itself easily to compromise" (Laponce, 1992)—and may lead to various degrees of redistributional tension among groups of citizens. Of course, we do not imply that languages are the only aspect of group dissimilarity. Instead, our measure of linguistic diversity should be viewed as a proxy for the broader notion of ethnolinguistic or cultural diversity.[5] In addition to being an important societal characteristic, an attractive feature of linguistic

3. A similar point has been made in earlier theoretical work by Greenberg (1956) and Rao (1982).

4. The Ethnologue project lists 6,912 known living languages <www.ethnologue.com>. There are 191 member states in the United Nations, and the CIA World Factbook lists 271 nations, which include certain territories and colonies, such as Puerto Rico or Guadeloupe.

5. There are of course examples of countries, such as Rwanda, which are linguistically homogeneous, but ethnically divided.

heterogeneity is that quantifying the degree of dissimilarity between languages is relatively easy.

The relevance of including distances when measuring diversity is ultimately an empirical question. To verify whether this feature improves upon the existing results, we compare the distance-based index with the one that does not include distances. Failing to take into account distances makes diversity statistically insignificant.[6] However, once distances are taken into account, diversity becomes statistically significant at the 1% or 5% level in virtually all specifications. This effect is found to be highly robust, and quantitatively important. Compared to an average level of redistribution of 9.5% of GDP, the model predicts that an increase in diversity by one standard deviation lowers redistribution as a share of GDP by about one percentage point. In other words, an increase by one standard deviation lowers redistribution by about 10%.

To carry out our analysis, we propose a general index of *social effective antagonism* constructed along the lines of the identification-alienation framework of Esteban and Ray (1994). In particular, we consider five special cases: a diversity index without distances (ELF) (Atlas Narodov Mira, 1964); a diversity index with distances (Greenberg, 1956); a polarization index without distances (Reynal-Querol, 2002); a polarization index with distances (Esteban and Ray, 1994); and a peripheral heterogeneity index (Desmet, Ortuño-Ortín, and Weber, 2005). The unified framework allows us to easily compare the aforementioned indices.

The wide variety of indices used in the literature partially stems from the fact that some economic and social outcomes can be explained by societal diversity (Alesina et al., 2003), whereas others are better captured by polarization (Esteban and Ray 1994, 1999). To illustrate the difference between diversity and polarization, compare two countries, A and B. If A consists of two equally sized groups, and B of three equally sized groups, then A is more polarized, but less diverse, than B. When distances between groups are taken into account, the difference between polarization and diversity becomes more subtle. In that case, increasing the number of equally sized groups need no longer imply increasing the level of diversity, since distances between groups also play a role.

Again, the question of which index does a better job at explaining redistribution is an empirical one. The results confirm that the crucial difference is whether distances are taken into account or not. The difference between diversity, peripheral heterogeneity, and polarization turns out to be empirically irrelevant. As long as distances are incorporated, all three indices perform extremely well.[7] This is not surprising, as the correlation between them is high.

6. This negative result is consistent with previous work of Alesina et al. (2003). La Porta et al. (1999), in contrast, do find a statistically significant effect of ethnolinguistic diversity on redistribution, even when distances are not accounted for. However, once they control for per capita income, this negative correlation disappears.

7. Note, however, that this conclusion differs from that of Montalvo and Reynal-Querol (2002, 2005) in their study of civil conflicts. This point is discussed in more detail subsequently.

The rest of the paper is organized as follows. Section 2 presents a general index of social effective antagonism, and shows that several well known indices of diversity and polarization are special cases. Section 3 deals with data and measurement issues. Section 4 computes and discusses the five different diversity and polarization indices for 225 countries. Section 5 shows the strong link between diversity and redistribution, once distances between languages are taken into account. Section 6 considers polarization and peripheral heterogeneity. Section 7 concludes.

2. Indices of Diversity and Polarization

In this section we present a general index of *social effective antagonism* that contains as special cases several indices of diversity and polarization widely used in the literature. To do so, we follow, with some minor differences, the identification-alienation framework of Esteban and Ray (1994).

Consider a country with a population of N individuals, partitioned into K distinct groups, indexed by $j = 1, \ldots, K$. The population of group j is denoted by N_j. We impose no conditions on the geographical distribution of the groups. Thus, individuals from a group can either live in the same region, or be dispersed across different regions. Each individual belongs to one and only one group, so that

$$N = \sum_{j=1}^{K} N_j.$$

We denote the population share of group j by

$$s_j = N_j/N,$$

so that $\sum_{j=1}^{K} s_j = 1$. The population shares, rather than their absolute sizes, is what will matter for our analysis.

A crucial element is the introduction of distances between groups. There is a matrix T that assigns a distance τ_{jk} between groups j and k. This distance is a standardized metric, that is, all values τ_{jk} lie between 0 and 1; $\tau_{jj} = 0$ for all j; $\tau_{jk} = \tau_{kj}$ for all j and k; and $\tau_{ij} \leq \tau_{ik} + \tau_{kj}$ for all i, j and k. Our empirical investigation deals with linguistic distances, where groups are formed by individuals who speak the same language, and τ_{jk} is the linguistic distance between the language spoken by group j and group k. Even though our model deals with a general concept of distance we adopt the linguistic terminology hereafter.

In order to define the notion of *social effective antagonism*, we follow Esteban and Ray (1994) and first introduce the concepts of *alienation* and *identification*. An individual of group j feels identified with other individuals in the same group, for example, those who speak the same language. The degree of identification depends on the size of the group, s_j, and is given by the value s_j^{α}. In Esteban and Ray α is a positive number, implying that the sense of identification is stronger in a larger group. Instead, we also allow for $\alpha = 0$, which captures the possibility of identification being independent of the size of the group.

The alienation felt by an individual of group j toward an individual of group k is increasing in the distance τ_{jk}. The sense of identification toward the own group may affect an individual's alienation toward another group. This interaction between alienation and identification yields antagonism. As defined in Esteban and Ray (1994), the antagonism between an individual of group j and an individual of group k is given by $s_j^{\alpha}\tau_{jk}$. Because there is a fraction s_k of individuals who speak language k, the *effective antagonism* of an individual of group j toward group k is $s_k s_j^{\alpha}\tau_{jk}$.[8] Given that a share s_j of the population speaks language j, the total effective antagonism of group j toward group k is $s_k s_j^{1+\alpha}\tau_{jk}$.

Similar to Esteban and Ray (1994), we can now define the country's level of *social effective antagonism* as the sum of the effective antagonisms between all pairs of groups:

$$A(\alpha, \tau) = \sum_{j=1}^{K}\sum_{k=1}^{K} s_k s_j^{1+\alpha}\tau_{jk}. \tag{1}$$

Depending on the values of α and the distance matrix τ, the index in equation (1), henceforth referred to as A-index, can be shown to generate as special cases different indices of diversity and polarization.

It is useful to distinguish between three different distance matrices used in our analysis. The first, denoted by T, allows for a continuous measure of distance τ_{jk} on the interval (0, 1] between any two groups j and k, where $j \neq k$. The second matrix, denoted by T^d, is of a dichotomous nature, where $\tau_{jk} = 1$ for all $j \neq k$. Here the distance between any two distinct groups is 1, so that the alienation experienced by an individual speaking language j toward an individual speaking any other language does not depend on the distance between the two languages. The third matrix, denoted by T^c, assumes there is a center group c, such that $\tau_{jk} = 0$ if $j \neq c$ and $k \neq c$. This implies that only the distances between the central group and the other (peripheral) groups matter.

We now consider five special cases of the general A-index, each one of which has been described in the literature.[9] Some are indices of diversity and others are indices of polarization.

8. This assumes that population shares, rather than absolute population sizes, matter.

9. See Esteban and Ray (2006) for a similar discussion in the cases of polarization and ELF.

1. ELF (Atlas Narodov Mira, 1964): $\alpha = 0$ and the distance matrix is T^d.

   The A-index (1) can be written as

$$A(0, T^d) = 1 - \sum_{j=1}^{K} s_j^2. \tag{2}$$

   This is the well known ethnolinguistic fractionalization (ELF) index (see, e.g., Atlas Narodov Mira (1964), Easterly and Levine (1997), and Alesina, Baqir, and Easterly (1999)), known elsewhere as the Gini-Simpson index. The ELF index measures the probability of two randomly chosen individuals being from different groups and does not take into account the distances between the different groups. It satisfies the fundamental requirements of diversity (Shannon, 1949):
   (i) for a given number of groups, the index reaches its maximum when all groups are of the same size;
   (ii) and if all groups are of equal size, then the society with a larger number of groups possesses a higher index of diversity.
2. GI (Greenberg, 1956): $\alpha = 0$ and the distance matrix is T. The A-index (1) becomes

$$A(0, T) = \sum_{j=1}^{K} \sum_{k=1}^{K} s_k s_j \tau_{jk}. \tag{3}$$

   This index was proposed by Greenberg (1956) and was examined (as quadratic entropy) in Rao (1982).[10] GI computes the population weighted total distances between all groups[11] and can be interpreted as the expected distance between two randomly selected individuals.[12] GI is essentially a generalization of ELF, whereby distances between different groups are taken into account. Naturally, GI does not satisfy the requirements of a diversity index mentioned previously and the maximal diversity need not be attained when all groups are of the same size. In fact, one can find distance matrices for which the maximum value of the index is obtained when at least one group i has population share $s_i = 0$, even though the distance from that group to any other group is strictly positive, namely, $\tau_{ij} > 0$ for all $j \neq i$.[13] This explains why some authors refer to GI as a "weak diversity" index (Ricotta, 2005).

10. See also Nei and Li (1979), Fearon (2003), and Bossert, D'Ambrosio, and La Ferrara (2006).

11. If τ_{jk} were the income difference between group j and group k, GI would coincide with the Gini index.

12. Ricotta and Szeidl (2006) argue that this index can be viewed as the expected conflict among species in a given environment.

13. Pavoine, Ollier, and Pontier (2005) show that in the case of ultrametric distances all groups must have a strictly positive population share for the index to be maximized.

3. RQ (Reynal-Querol, 2002): $\alpha = 1$ and the distance matrix is T^d. The A-index (1) is

$$A(1, T^d) = \sum_{j=1}^{K} s_j^2(1 - s_j), \tag{4}$$

which is the polarization index proposed by Reynal-Querol (2002).[14] Similarly to ELF, RQ does not take into account distances between groups. It attains its maximum when there are two groups of equal size.

4. ER (Esteban and Ray, 1994): $\alpha = 1$ and the distance matrix is T. The A-index becomes

$$A(1, T) = \sum_{j=1}^{K} \sum_{k=1}^{K} s_k s_j^2 \tau_{jk}. \tag{5}$$

In fact, this is a special case of the polarization index in Esteban and Ray (1994), who allow for α to be in the range of $[1, 1.6]$.[15] As with GI, this index controls for distances between groups. If distances between all groups are the same, it is perfectly correlated with RQ.

5. PH (peripheral heterogeneity in Desmet, Ortuño-Ortín, and Weber 2005): $\alpha = 0$ and the distance matrix is T^c. The A-index is

$$A(0, T^c) = 2 \sum_{j=1}^{K} s_j s_c \tau_{cj}, \tag{6}$$

where the central group c is the largest. PH is a special case of the diversity index studied in Desmet, Ortuño-Ortín, and Weber (2005). It is important to point out that PH is a variant of GI, with the difference that it takes into account the alienation between the center and the peripheral groups, but not between the peripheral groups themselves. Note that PH can be viewed as an intermediate index between diversity and polarization. If all distances to the center are equal, then any vector of population shares, including the one with all peripheral groups having equal shares, attains the maximum. However, if distances to the center are not equal, then $A(0, T^c)$ attains a maximum when only the center and the peripheral group with the highest value of τ_{cj} have strictly positive population shares.

The ELF index, which has been widely used to study the effects of diversity on different economic outcomes, is based on a dichotomous 0-1 distance measure.

14. Montalvo and Reynal-Querol (2002) show that when the distance matrix is T^d, the parameter α has to be equal to 1 for the index to satisfy the properties of polarization.

15. Actually, to satisfy all the polarization properties in Esteban and Ray (1994), the matrix distance should be additive, that is, $\tau_{ik} = \tau_{ij} + \tau_{jk}$ (Duclos, Esteban, and Ray, 2004). We do not impose this condition, since the language distances in the multidimensional space do not satisfy it.

As soon as two linguistic groups are different, they are assigned a distance of 1. However, in practice defining when a group is distinct from another can be difficult. For example, should we consider speakers of Venetian and Italian as members of different groups in the same way as speakers of Greek and Turkish? These two pairs are drastically different: Venetian is a dialect of Italian, whereas Greek and Turkish pertain to entirely different families, Greek to Indo-European and Turkish to Altaic. In using ELF, one might decide to assign a distance of 0 to Venetian and Italian, and a distance of 1 to Greek and Turkish. This choice implies aggregating speakers of Venetian and Italian into one group. Thus, the problem of assigning distances turns into the problem of identifying groups. GI has a clear advantage in this regard: By shifting from a dichotomous 0-1 measure to a continuous measure of distance, GI avoids the ad hoc group identification problem. Venetian and Italian are considered as two different languages, with a small distance between the two.[16] Even though GI has been examined by several authors, it has not been used to study the effects of within country diversity on different economic variables.[17]

Of course, whether GI improves upon ELF in its capacity to explain redistribution is eventually an empirical question. From a theoretical point of view, it is not entirely obvious whether using a continuous measure of distance should improve the results. One might argue, as we do in this paper, that the degree of conflict depends not only on the number and the sizes of the different groups, but also on how different they are. However, it is also possible that the basis for the alienation experienced by individuals of one ethnic group towards individuals of another ethnic group is simply the fact that they belong to different groups, regardless of their cultural distances. This is the view of Montalvo and Reynal-Querol (2005), who claim that *the dynamics of the "we" versus "you" distinction is more powerful than the antagonism generated by the distance between them*. However, this does not free them altogether from assigning distances. Indeed, Montalvo and Reynal-Querol focus on what they call "relevant" groups. By selecting which groups are "relevant," they implicitly employ a dichotomous 0-1 distance measure between all groups.

Some authors have argued that ethnic conflict is better measured by the degree of polarization than by the degree of diversity (Esteban and Ray 1994, 1999). As mentioned previously, for a given number of groups, in absence of distance considerations, diversity is maximized when all groups are of equal size. In contrast, polarization is maximized when two of the groups consist each of half of the population and all other groups have zero population. If there are only two groups,

16. Note that GI satisfies a "continuity property": When the distance between two groups tends to zero, the diversity approaches the limit where both groups are merged into a single one.

17. For example, Fearon (2003) uses GI to compute a measure of cultural diversity for 160 countries, but he does not analyze its effect on economic variables, whereas Spolaore and Wacziarg (2006) use GI to measure diversity between countries and explore its relation to bilateral income differences.

then polarization and diversity yield identical rankings. Moreover, in practice, there is a significant overlap between both notions, and ranking countries based on either polarization or diversity produces similar results. We will consider two different indices of polarization, with ER using a continuous measure of distance, and RQ a dichotomous measure of distance. Therefore, the link between ER and RQ is the same as between GI and ELF.

In addition to comparing ELF and GI, and ER and RQ, we also analyze the empirical relevance of the PH index, which does not treat group heterogeneity in a symmetric manner. This index focuses on the tension that emanates from the heterogeneity between a central dominant group and the peripheral minority groups. Apart from being theoretically a different concept, this index has the advantage of requiring a smaller set of distance data.

Although from a theoretical point of view all these indices are different, in practice some of them may yield similar conclusions. In fact, as we will see, the correlation between all these indices is relatively high, and the only factor which is important in the empirical analysis is whether we use continuous (GI, PH, and ER) or dichotomous (ELF and RQ) measures of distances. The distinction between GI, PH, and ER does not seem to be empirically relevant to the problem at hand.

3. Data and Measurement Issues

Our data cover a wide cross-section of countries. The information on how many people speak a given language in a given country comes from the Ethnologue project. What sets the Ethnologue apart from other sources, such as the Britannica Book of the Year, is its detail. For example, in the case of Mexico the Ethnologue lists 291 living languages, much higher than the 21 languages listed in the 1990 edition of the Britannica. Although some aspects of the Ethnologue have been used before (Alesina et al., 2003; Fearon, 2003), its greater detail in terms of languages spoken across countries has so far been left unexploited by economists. One reason could be that when data are highly detailed, the problem of group identification arises. In the absence of information on linguistic distances, one has to assign a distance of either zero or one between Venetian and Italian. This problem does not show up in the less detailed Britannica, which already implicitly aggregates the speakers of Venetian and Italian into a single group.[18]

The introduction of linguistic distances largely resolves the group identification problem. It is no longer necessary to make ad hoc choices about whether Venetian and Italian belong to the same group. By using the detailed Ethnologue, we maintain Venetian and Italian as two distinct groups, but assign a small

18. See, for example, Alesina et al. (2003).

positive distance between the two, based on linguists' assessment of how different these languages are, as we further explain subsequently. In general, once we have detailed information about distances between languages, then more disaggregated data become preferable. In contrast to much of the literature, we therefore make no choices about when a group is a group.[19] Instead, we use the entire Ethnologue database—even if this implies 291 languages in Mexico—but correct for distances.

There are different ways of measuring distances between languages. A first measure is based on linguistic tree diagrams. Using this approach, Fearon (2003) defines the distance between languages j and k to be

$$\tau_{jk} = 1 - \left(\frac{\ell}{m}\right)^{\delta}, \tag{7}$$

where ℓ is the number of shared branches between j and k, m is the maximum number of branches between any two languages, and δ is a parameter that determines how fast the distance declines as the number of shared branches increases. Data on language trees come from the Ethnologue project.[20] A second measure is based on lexicostatistical studies. A prime example is Dyen et al. (1992).[21] They focus on 200 basic meanings, and compute for each pair of 95 Indo-European languages the proportion of cognates.[22] The distance between any two languages j and k can then be defined as one minus the proportion of cognates between j and k. In the empirical analysis the distance measure based on tree diagrams is more useful, because it covers all languages, and not only those from the Indo-European group.

When extracting quantitative distance measures from language trees, one relevant question is: How much more distant should we consider two languages from different families to be relative to languages that belong to the same family? This largely amounts to deciding the value of δ in the distance measure (7). Consider two language pairs: Greek and Italian (both Indo-European, with one shared branch) and Chinese and Italian (with zero shared branches). If, as in Fearon (2003), we take a δ of 0.5, the distance between Greek and Italian is 0.74, whereas the distance between Chinese and Italian is 1. Lowering the value of δ increases this relative distance. For example, if we take a δ of 0.05, the distance

19. See Fearon (2003) for an excellent discussion of this issue.

20. Although Fearon (2003) uses the Ethnologue to compute distances, he does not use the Ethnologue data on the linguistic composition of different countries. Instead, he collects data on ethnic composition from secondary sources.

21. The Dyen measure has previously been used in economics by Ginsburgh, Ortuño-Ortín, and Weber (2005).

22. The term *cognate* applies if the two varieties have an unbroken history of descent from a common ancestral form.

between Greek and Italian is now 0.13, whereas between Chinese and Italian it is still 1.

In our empirical exploration we find that values for δ in the range between 0.04 to 0.10 perform well and give similar results in terms of the statistical significance of the diversity measure.[23] Even though we settle on a value of 0.05, indices that use distance measures based on values of δ outside the [0.04, 0.10] range continue to outperform indices without distances.

Because δ has a useful economic interpretation, there is something to be learned from the relative performance of different values of the parameter. For example, the fact that a δ of 0.05 has high predictive power tells us that Italian and Chinese are perceived to be 6.7 times more distant from each other than Italian and Greek. This result is informative, and very different from the Fearon δ of 0.5, which would have led us to the interpretation that Italian and Chinese are perceived to be only 0.30 times more distant than Italian and Greek.

As for the dependent variable in our regressions, redistribution is measured by government transfers and subsidies as a percentage of GDP. As in La Porta et al. (1999), the data come from the *Economic Freedom Data Network* and we have taken the average for the years 1985–1995. The distinction between transfers and subsidies is based on the target of redistribution: If, as in social welfare programs, the beneficiary is an individual, the redistribution is "transfer"; if the target is a business, for example, state-owned enterprise, it is a "subsidy." The figures aggregate central and local governments.

In analyzing the effect of diversity on redistribution we introduce a number of additional control variables, in line with La Porta et al. (1999) and Alesina et al. (2003): average population 1985–1995, average GDP per capita 1985–1995, legal origin, religious composition, latitude, percentage of population above 65, regional dummies, and a "small island" dummy. Most of these control variables have been taken directly from La Porta et al. (1999), and others come from the World Bank. A detailed description of the data is given in Appendix A.

4. Linguistic Distances

In this section we discuss the different linguistic diversity indices we use in our empirical analysis. The Ethnologue database provides detailed information on the languages spoken in 225 countries. It also provides linguistic trees for each of the 6,912 listed languages. We use these tree diagrams to compute linguistic distances, following the Fearon (2003) formula in equation (7). Based on this information, Table B.1 in Appendix B reports the five different indices discussed

23. In particular, in the eight regressions we report, the diversity index with a δ in this range is statistically significant at the 1% level in at least three of the eight reported regressions, and at the 5% level in the remaining regressions.

in Section 2: GI, ELF, ER, RQ, and PH. Although the data on transfers and subsidies will limit the number of countries in our regressions to 105, we present the different indices for all 225 available countries.

Our main focus is on comparing ELF and GI. As argued before, ELF has been widely used in the literature, partly because of its straightforward interpretation. GI generalizes ELF by controlling for distances between groups. Although the rank correlation between ELF and GI is substantial (0.69), the last column of Table B.1 shows a number of interesting patterns when comparing both rankings.

First, Latin American countries tend to become more diverse when controlling for distance. Their populations are typically made up of a mix of Indo-European and Amerindian speakers. The Amerindian languages spoken in a given country are often quite different. For instance, Bolivia, which tops the GI ranking, moved up 54 positions, compared to ELF. In that country, roughly 40% of the population speaks Spanish, with the remaining 60% Amerindian languages. Within the Amerindian speakers, half speak Quechua and one third Aymara, two languages which belong to entirely different families. Other examples include Mexico and Ecuador, which move up, respectively, 69 and 68 positions.

Second, African countries become less diverse when taking into account distances. Although many languages are spoken in those countries, they tend to be quite similar. The most extreme example is Togo, which drops 126 positions in the ranking. There are 40 languages spoken in that country, nearly all of which belong to the Niger-Congo/Atlantic-Congo/Volta-Congo classification. The same pattern reappears in many other African countries: Ghana drops 119 positions, Benin 114, and Côte d'Ivoire 107. The fact that Africa becomes less diverse, compared to existing diversity measures, is relevant. Conventional wisdom has often considered diversity as acting as a sort of Africa dummy. Indeed, with the traditional ELF index, 15 out of the 20 most diverse countries are African, so that any relation between diversity and economic performance is likely to be strongly influenced by any characteristics specific to Africa. In contrast, with GI only 3 out of the 20 most diverse countries are still African. Our main result will therefore overturn the conventional wisdom: As we will see, taking into account distances, and thus reducing the diversity of Africa, substantially improves the predictive power of diversity.

Third, the picture for Europe is mixed. Some countries become more diverse when controlling for distances. This is the case of Bulgaria, a country with a sizeable Turkish minority, or Estonia, where the majority speaks a non-Indo-European language, but the minority is Russian-speaking. Other European countries move in the opposite direction. Andorra, where half the population speaks Catalan and the other half Spanish, drops 102 positions, on account of the small distance between both languages.

In addition to comparing ELF and GI, we are also interested in contrasting ER and GI. Although polarization and diversity often provide similar rankings—the rank correlation between ER and GI is 0.87—there are some relevant differences. For example, Papua New Guinea is extremely diverse (position 4), but not polarized at all (position 186). That country is made up of 829 different groups, all of which are small. On the contrary, some countries are more polarized than diverse. For instance, Barbados and Guadeloupe are both essentially made of up of two groups, one speaking an Indo-European language (English or French), and the other speaking a Creole language.

Our last index, PH, is closely related to GI: Its rank correlation is 0.93. Given this high correlation, one may conclude that peripheral heterogeneity does not add anything new. However, it also suggests that in the absence of data to compute GI, then PH might be a good proxy. Remember that GI requires data on distances between any two languages spoken in any country, whereas PH only requires data on distances to the dominant language in each country. This is relevant if one wants to use data on linguistic distances based on lexicostatistical studies, such as the Dyen, Kruskal, and Black (1992) database on Indo-European languages.[24]

5. Diversity and Redistribution

Table 1 reports the coefficients of our regressions of redistribution on GI. To make our results comparable to previous work, we have included similar control variables as in La Porta et al. (1999) and Alesina et al. (2003). Robust *t*-ratios are given in brackets. As explained before, the theoretical prior is that the greater the degree of linguistic diversity, the lower the degree of redistribution.

Column 1 reports the most basic specification, only including diversity and a number of exogenous control variables (latitude, regional dummies, and small island dummy). All coefficients are highly significant. The coefficient on diversity is significant at the 5% level. Following La Porta et al. (1999) and Alesina et al. (2003), columns 2 and 3 include legal origin and religious composition as further controls. None of these variables are significant though. Column 4 leaves out the regional dummies, and adds GDP per capita. Column 5 does the same, but keeps the regional dummies. As expected, the level of a country's development increases the degree of redistribution. In contrast to La Porta et al., the effect of diversity on redistribution is robust to the inclusion of GDP per capita. Column 6 adds population, instead of GDP per capita. This is in line with Alesina and Wacziarg (1998), who empirically confirmed that transfers as a share of GDP are unrelated to population size, since this type of government expenditure does not have the nature of a public good. Column 7 leaves out regional dummies, and

24. One advantage of lexicostatistical studies is that the distances are less coarse than those based on language trees. When using trees, the maximum number of nodes—15 in our case—determines the number of possible distances between language pairs. Instead, the Dyen measure is more continuous.

Table 1. Diversity (GI) and redistribution.

	Dependent variable: Transfers and subsidies as % of GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GI (Greenberg)	−6.313	−6.663	−3.842	−5.779	−5.071	−4.124	−4.343	−4.042
	[2.60]**	[2.66]***	[1.99]**	[2.77]***	[2.79]***	[2.12]**	[2.13]**	[2.32]**
UK legal origin		2.077	3.212	3.951	3.554	2.239	4.992	4.59
		[0.79]	[1.45]	[1.36]	[1.58]	[0.95]	[2.30]**	[2.39]**
French legal origin		2.246	3.102	3.948	3.853	2.323	4.13	4.011
		[0.80]	[1.25]	[1.22]	[1.52]	[0.87]	[1.57]	[1.72]*
Socialist legal origin		4.949	4.973	9.4	8.441	3.989	9.126	8.666
		[1.47]	[1.72]*	[2.54]**	[2.72]***	[1.33]	[2.92]***	[2.93]***
Scandinavian legal origin		1.985	5.243	6.234	6.104	4.531	6.432	6.626
		[0.50]	[1.17]	[1.39]	[1.44]	[1.00]	[1.70]*	[1.73]*
Catholic 1980			0.029	0.01	0.02	0.024	0.031	0.034
			[1.24]	[0.49]	[1.00]	[1.05]	[1.64]	[1.64]
Muslim 1980			−0.093	−0.05	−0.079	−0.097	0.007	−0.024
			[4.12]***	[2.51]**	[3.93]***	[4.48]***	[0.31]	[1.04]
Protestant 1980			−0.037	−0.045	−0.041	−0.043	−0.009	−0.02
			[1.03]	[1.17]	[1.11]	[1.12]	[0.22]	[0.51]
Population 1985–1995						−0.176	0.063	0.054
						[0.64]	[0.24]	[0.20]
GDP per capita 1985–1995				1.684	1.792		0.621	1.037
				[4.50]***	[3.95]***		[1.48]	[2.05]**
Population above 65							0.872	0.685
							[5.14]***	[4.11]***
Small island	−6.246	−5.811	−5.538	−7.425	−6.672	−6.079	−6.831	−6.363
	[2.29]**	[2.13]**	[2.55]**	[3.01]***	[3.46]***	[2.55]**	[3.28]***	[3.64]***
Latitude	28.638	27.7	21.188	23.328	13.402	20.909	12.059	8.049
	[6.94]***	[5.64]***	[5.02]***	[6.29]***	[2.84]***	[5.04]***	[3.02]***	[1.80]*
Latin America	−1.962	−1.711	−8.098		−7.077	−8.18		−4.691
	[1.20]	[1.05]	[4.75]***		[4.50]***	[4.87]***		[3.06]***
Sub-Saharan Africa	−2.635	−2.332	−5.919		−3.431	−6.15		−1.571
	[1.52]	[1.36]	[3.63]***		[2.16]**	[3.80]***		[1.07]
East Asia & Pacific	−2.71	−1.755	−5.033		−5.021	−4.817		−3.446
	[1.24]	[0.82]	[2.17]**		[2.49]**	[2.07]**		[1.88]*
Constant	3.531	1.268	5.535	−12.842	−7.622	9.737	−11.648	−9.635
	[1.64]	[0.35]	[1.65]	[2.61]**	[1.70]*	[1.71]*	[1.73]*	[1.38]
Observations	105	105	104	103	102	102	103	102
R-squared	0.6569	0.6693	0.7905	0.7787	0.8203	0.7913	0.8225	0.8436

Notes: Robust *t*-statistics in brackets.
*Significant at 10%; ** significant at 5%; *** significant at 1%.

adds the share of population above 65. Column 8 provides the full specification with all regressors. As can be seen, the population above 65 is highly significant.

Based on Table 1, we can see that in all specifications the effect of diversity is robust, both in terms of magnitude and statistical significance. In half of the specifications GI is significant at the 1% level, and in the other half it is significant at the 5% level. The magnitude of its coefficient hovers between −3.8 and −6.6. Taking column 5 as our preferred specification, the model predicts that an increase in diversity by one standard deviation lowers redistribution as a share of GDP by 0.97 percentage points. This effect should be compared to an average level of redistribution of 9.5% of GDP. In other words, an increase by one standard deviation in diversity lowers redistribution by about 10%. An example may help to further illustrate the quantitative importance of diversity. Compare Paraguay, with a level of redistribution of 2.3%, and Uruguay, with a level of redistribution of 13.1%. The model predicts that the greater linguistic diversity of Paraguay should lower redistribution by 5.6 percentage points compared to Uruguay. This implies that about half of the difference in redistribution between Paraguay and Uruguay can be explained by the difference in linguistic diversity.

As for the control variables, neither the legal origin nor the religious composition tends to have a significant impact on redistribution. There are two exceptions to this. Socialist legal origin increases redistribution, although its effect only shows up if we control for GDP per capita. In those regressions, the effect of having a socialist legal origin increases redistribution by 8 to 9 percentage points. The share of Muslims tends to lower redistribution, although its effect vanishes when we control for the share of the population above 65. This suggests that Muslim countries tend to have a young population. The small island dummy[25] is highly significant and its coefficient is robust to all specifications. Being a small island is predicted to reduce redistribution by about 6 percentage points of GDP. This variable was included, because our preliminary data analysis suggested small islands were outliers.

Whether including linguistic distances is relevant for our understanding of redistribution ought to be an empirical question. Table 2 runs the same regressions as Table 1, but uses the standard ELF index, which does not allow for different distances between languages. We will focus on our variable of interest—linguistic diversity—as the coefficients on the control variables are similar to what we found in Table 1. The most obvious result is that ELF loses statistical significance. In six out of the eight specifications the index ceases to be significant at the 10% level. Not surprisingly, all specifications also give lower R^2s, although the differences are small.

25. The small island dummy is defined as an island with a population of less than 0.5 million. See Appendix A.

TABLE 2. Diversity (ELF) and redistribution.

	Dependent variable: Transfers and subsidies as % of GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ELF	−4.763	−4.525	−2.155	0.662	−2.207	−2.057	0.588	−1.666
	[2.41]**	[2.34]**	[1.22]	[0.43]	[1.44]	[1.16]	[0.45]	[1.19]
UK legal origin		1.806	2.992	2.838	3.168	2.098	4.294	4.367
		[0.67]	[1.37]	[1.04]	[1.38]	[0.88]	[2.11]**	[2.19]**
French legal origin		2.268	3.026	3.495	3.715	2.35	3.897	3.965
		[0.77]	[1.20]	[1.12]	[1.42]	[0.86]	[1.53]	[1.66]
Socialist legal origin		4.195	4.49	8.404	7.626	3.674	8.596	8.22
		[1.22]	[1.57]	[2.30]**	[2.43]**	[1.22]	[2.78]***	[2.77]***
Scandinavian legal origin		0.94	4.92	6.266	5.716	4.297	6.568	6.412
		[0.23]	[1.09]	[1.37]	[1.32]	[0.94]	[1.73]*	[1.66]
Catholic 1980			0.03	0.011	0.022	0.026	0.034	0.036
			[1.29]	[0.55]	[1.08]	[1.14]	[1.76]*	[1.73]*
Muslim 1980			−0.093	−0.058	−0.081	−0.097	0.003	−0.024
			[4.12]***	[2.84]***	[3.82]***	[4.42]***	[0.15]	[1.02]
Protestant 1980			−0.041	−0.056	−0.045	−0.045	−0.014	−0.022
			[1.08]	[1.40]	[1.19]	[1.12]	[0.34]	[0.54]
Population 1985–1995						−0.091	0.148	0.125
						[0.33]	[0.56]	[0.46]
GDP per capita 1985–1995				1.537	1.676		0.511	0.968
				[3.94]***	[3.65]***		[1.17]	[1.88]*
Population above 65							0.901	0.7
							[5.23]***	[4.05]***
Small island	−6.632	−5.983	−5.514	−6.585	−6.467	−5.742	−5.92	−5.978
	[2.49]**	[2.40]**	[2.47]**	[2.64]***	[3.37]***	[2.33]**	[2.98]***	[3.53]***
Latitude	27.524	27.788	21.516	26.24	14.6	21.358	13.593	8.746
	[6.79]***	[5.61]***	[5.03]***	[7.34]***	[3.22]***	[5.01]***	[3.51]***	[2.03]**
Latin America	−3.301	−2.984	−8.711		−7.796	−8.736		−5.129
	[2.01]**	[1.81]*	[4.92]***		[4.88]***	[4.97]***		[3.36]***
Sub-Saharan Africa	−1.93	−1.531	−5.457		−3.1	−5.658		−1.183
	[1.10]	[0.89]	[3.10]***		[1.83]*	[3.20]***		[0.76]
East Asia & Pacific	−2.932	−1.899	−5.069		−5.074	−4.872		−3.456
	[1.39]	[0.94]	[2.25]**		[2.58]**	[2.12]**		[1.93]*
Constant	4.986	2.342	5.899	−12.99	−6.614	8.402	−13.473	−10.474
	[2.25]**	[0.61]	[1.73]*	[2.55]**	[1.41]	[1.45]	[1.97]*	[1.49]
Observations	105	105	104	103	102	102	103	102
R-squared	0.6581	0.6676	0.7886	0.7674	0.8147	0.7885	0.8164	0.8401

Notes: Robust t-statistics in brackets.
*Significant at 10%; ** significant at 5%; *** significant at 1%.

Although the specifications we used in Tables 1 and 2 are based on the previous work by La Porta et al. (1999) and Alesina et al. (2003), one could have some doubts about how general the superiority of GI over ELF is. To address this issue, we use all possible combinations of sets of control variables to analyze the effect of diversity on redistribution. In particular, we focus on eight sets of control variables: religion, legal origin, regional dummies, small island dummy, latitude, GDP per capita, population, and population above 65. The different combinations of these eight sets of regressors give us 255 possible regressions. Although not all of these regressions may make theoretical sense, it is still useful to compare the relative performance of GI and ELF. It turns out that in all 255 specifications, the *t*-values are larger for GI than for ELF. Moreover, in 79% of the regressions, GI is significant at least at the 10% level, whereas this figure drops to 32% in the case of ELF. This leads us to conclude that including distances into our diversity index substantially improves our results. The understanding of diversity is therefore enhanced by incorporating linguistic distances in our index.

Another possible concern is that we have computed the ELF index using the detailed Ethnologue database. In doing so, we did not aggregate, say, Italian and Venetian speakers into one group. This may bias our results against the ELF index. As mentioned before, previous studies, such as Alesina et al. (2003), have used less-detailed databases, so that de facto Venetian and Italian did not show up as distinct groups. To see whether this is an issue, we re-ran our eight basic regressions using the linguistic fractionalization index of Alesina et al. (2003). In six out of the eight specifications, linguistic fractionalization does not pass the 10% significance threshold. This is similar to our findings when using the ELF index based on the Ethnologue.

6. Robustness: Polarization and Peripheral Heterogeneity

In this section we discuss polarization and peripheral heterogeneity. Once again, the focus is on the importance of including distances. As before, indices with distances perform clearly better. However, among those indices that include distances, all of them—GI, ER, and PH—perform roughly speaking equally well.

Tables 3 and 4 report the same regressions as before, but now use polarization, instead of diversity, as the explanatory variable. In Table 3 polarization is measured by ER, which contains linguistic distances, and in Table 4 it is measured by the RQ index, which does not account for different linguistic distances between groups. The results are similar to the case of GI and ELF. Table 3 shows that ER is significant at the 1% level in three out of the eight specifications, and at the 5% level in the remaining five regressions. When not accounting for distances, Table 4 shows that only two out of the eight specifications are significant at the 5%, with the remaining six regressions not passing the 10% threshold. Taking, as before,

TABLE 3. Polarization (ER) and redistribution.

	Dependent variable: Transfers and subsidies as % of GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ER (Esteban-Ray)	−20.685	−22.504	−16.385	−27.576	−21.352	−18.358	−22.359	−17.812
	[2.47]**	[2.45]**	[2.40]**	[4.25]***	[3.59]***	[2.58]**	[3.15]***	[2.63]**
UK legal origin		1.729	3.139	4.32	3.513	2.148	5.15	4.49
		[0.67]	[1.42]	[1.47]	[1.55]	[0.90]	[2.36]**	[2.33]**
French legal origin		2.162	3.218	4.528	4.082	2.453	4.47	4.141
		[0.78]	[1.30]	[1.40]	[1.61]	[0.92]	[1.71]*	[1.78]*
Socialist legal origin		5.056	5.135	10.261	8.798	4.096	9.583	8.786
		[1.51]	[1.80]*	[2.78]***	[2.82]***	[1.38]	[3.12]***	[2.99]***
Scandinavian legal origin		1.677	5.009	6.478	5.849	4.228	6.521	6.336
		[0.42]	[1.13]	[1.47]	[1.41]	[0.95]	[1.79]*	[1.70]*
Catholic 1980			0.025	0.009	0.016	0.019	0.029	0.029
			[1.10]	[0.47]	[0.80]	[0.87]	[1.56]	[1.45]
Muslim 1980			−0.096	−0.054	−0.083	−0.102	0.001	−0.029
			[4.38]***	[2.73]***	[4.11]***	[4.91]***	[0.03]	[1.28]
Protestant 1980			−0.037	−0.048	−0.041	−0.045	−0.015	−0.021
			[1.10]	[1.38]	[1.24]	[1.22]	[0.40]	[0.59]
Population 1985–1995						−0.256	−0.056	−0.02
						[0.90]	[0.22]	[0.07]
GDP per capita 1985–1995				1.844	1.844		0.751	1.053
				[4.88]***	[4.01]***		[1.77]*	[2.10]**
Population above 65							0.841	0.677
							[5.00]***	[4.06]***
Small island	−6.14	−5.532	−5.48	−7.127	−6.623	−6.293	−6.981	−6.558
	[2.16]**	[2.01]**	[2.45]**	[3.07]***	[3.35]***	[2.58]**	[3.48]***	[3.56]***
Latitude	30.279	29.208	21.772	23.55	13.979	21.532	12.894	8.703
	[7.08]***	[5.77]***	[5.11]***	[6.56]***	[2.94]***	[5.22]***	[3.28]***	[1.93]*
Latin America	−1.285	−1.017	−7.67		−6.499	−7.749		−4.294
	[0.75]	[0.60]	[4.60]***		[4.29]***	[4.74]***		[2.88]***
Sub-Saharan Africa	−2.845	−2.518	−6.24		−3.782	−6.574		−1.989
	[1.55]	[1.37]	[3.83]***		[2.49]**	[4.23]***		[1.45]
East Asia & Pacific	−2.802	−1.788	−5.213		−5.273	−5.029		−3.672
	[1.27]	[0.80]	[2.22]**		[2.56]**	[2.18]**		[1.98]*
Constant	2.764	0.691	5.574	−14.294	−8.045	11.208	−10.505	−8.34
	[1.24]	[0.18]	[1.62]	[2.92]***	[1.74]*	[1.93]*	[1.61]	[1.20]
Observations	105	105	104	103	102	102	103	102
R-squared	0.6545	0.6684	0.7932	0.7923	0.8244	0.7947	0.8313	0.8466

Notes: Robust *t*-statistics in brackets.
*Significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 4. Polarization (RQ) and redistribution.

	Dependent variable: Transfers and subsidies as % of GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RQ (Reynal-Querol)	−17.854	−17.085	−8.419	−7.333	−10.964	−8.711	−2.535	−6.265
	[2.27]**	[2.09]**	[1.17]	[0.93]	[1.60]	[1.13]	[0.34]	[0.87]
UK legal origin		1.245	2.743	3.287	3.141	1.905	4.462	4.186
		[0.45]	[1.24]	[1.08]	[1.30]	[0.78]	[2.05]**	[2.09]**
French legal origin		1.771	2.839	3.711	3.715	2.225	3.953	3.87
		[0.60]	[1.12]	[1.09]	[1.36]	[0.80]	[1.48]	[1.60]
Socialist legal origin		3.695	4.201	8.443	7.573	3.332	8.516	8.004
		[1.07]	[1.47]	[2.19]**	[2.36]**	[1.09]	[2.70]***	[2.72]***
Scandinavian legal origin		−0.291	4.383	5.477	5.048	3.71	6.167	6.026
		[0.07]	[0.97]	[1.18]	[1.18]	[0.81]	[1.64]	[1.58]
Catholic 1980			0.028	0.01	0.02	0.024	0.032	0.034
			[1.19]	[0.47]	[0.98]	[1.04]	[1.68]*	[1.67]*
Muslim 1980			−0.095	−0.054	−0.082	−0.099	0.005	−0.027
			[4.27]***	[2.58]**	[3.83]***	[4.69]***	[0.22]	[1.18]
Protestant 1980			−0.043	−0.052	−0.046	−0.048	−0.013	−0.025
			[1.19]	[1.35]	[1.30]	[1.24]	[0.31]	[0.63]
Population 1985–1995						−0.182	0.117	0.065
						[0.61]	[0.44]	[0.23]
GDP per capita 1985–1995				1.579	1.761		0.51	0.991
				[4.04]***	[3.63]***		[1.15]	[1.86]*
Population above 65							0.897	0.69
							[5.07]***	[3.90]***
Small island	−6.746	−5.906	−5.493	−7.126	−6.609	−6.055	−6.298	−6.158
	[2.38]**	[2.34]**	[2.42]**	[3.05]***	[3.40]***	[2.38]**	[3.17]***	[3.42]***
Latitude	30.362	30.612	22.872	26.515	15.772	22.732	13.741	9.856
	[7.30]***	[6.26]***	[5.27]***	[7.69]***	[3.40]***	[5.27]***	[3.60]***	[2.20]**
Latin America	−2.546	−2.274	−8.381		−7.497	−8.506		−4.939
	[1.53]	[1.37]	[4.82]***		[4.85]***	[5.02]***		[3.30]***
Sub-Saharan Africa	−3.028	−2.537	−5.975		−3.591	−6.241		−1.631
	[1.72]*	[1.42]	[3.58]***		[2.30]**	[3.86]***		[1.15]
East Asia & Pacific	−3.246	−2.29	−5.326		−5.464	−5.172		−3.693
	[1.55]	[1.09]	[2.32]**		[2.83]***	[2.27]**		[2.09]**
Constant	4.047	1.938	5.878	−12.636	−7.219	9.966	−12.505	−9.619
	[1.82]*	[0.50]	[1.67]*	[2.47]**	[1.50]	[1.63]	[1.82]*	[1.36]
Observations	105	105	104	103	102	102	103	102
R-squared	0.66	0.67	0.79	0.77	0.82	0.79	0.82	0.84

Notes: Robust *t*-statistics in brackets.
*Significant at 10%; ** significant at 5%; *** significant at 1%.

column 5 as our preferred specification, the model predicts that an increase in ER by one standard deviation lowers redistribution as a share of GDP by 1.41 percentage points. This effect is somewhat larger than the 0.97 percentage points in the case of GI.

When comparing ER and GI, it is unclear which one prevails. Both indices are similar in their level of statistical significance: three specifications at the 1% level and the remaining five specifications at the 5% level. This suggests that the concepts of polarization and diversity have a significant overlap, and are thus hard to distinguish. This is not surprising, given the correlation of 0.76 between the two indices in the sample of 105 countries included in our regressions. In contrast to Montalvo and Reynal-Querol (2005), who demonstrate the relevance of polarization in the context of ethnic conflict, we show that diversity and polarization are similar in their impact on redistribution.[26] What does matter, instead, is taking into account distances between groups.

Table 5 analyzes the effect of PH on redistribution. Once again, this index, which accounts for distances, performs well. Peripheral heterogeneity is significant at the 1% level in four out of the eight specifications, and at the 5% in the remaining four specifications. Although not reported here, if one were to leave out distances, five out of the eight regressions would not pass the 10% significance level. The economic significance is similar as in the case of GI: an increase in PH by one standard deviation reduces redistribution by 1.09 percentage points. In terms of *t*-statistics, peripheral heterogeneity does better than GI in six out of the eight specifications. When running all 255 regressions, the statistical significance of peripheral heterogeneity is greater in 74% of the cases.

This suggests that PH is a reasonable alternative to GI. If our main focus were to be Europe and the Americas, there might be some advantage to using PH, rather than GI, as it would allow us to use alternative distance measures. In all of those countries, with the exception of Paraguay, the dominant language is Indo-European. This being the most widely studied language family, there exist detailed lexicostatistical studies, such as the one by Dyen et al. (1992), that offer alternative measures of distances. In an earlier version of this paper, Desmet, Ortuño-Ortín, and Weber (2005) used the distances in Dyen et al. for a sample of 55 countries in Europe and the Americas, and found, once again, that incorporating distances is crucial for diversity to be statistically significant. In addition, a decrease in diversity by one standard deviation was predicted to lower redistribution by around 11%, similar to the figure of 10% found in the current paper for a larger sample of countries and a different measure of distance.

26. One should be careful when interpreting the differences between diversity and polarization. As mentioned before, GI is not always a pure index of diversity. Thus, the high correlation between GI and ER might be partly due to the fact that they are conceptually more “similar” than ELF and ER.

Table 5. Peripheral diversity and redistribution.

	Dependent variable: Transfers and subsidies as % of GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peripheral diversity	−8.195	−8.605	−5.694	−9.958	−7.94	−6.396	−7.704	−6.417
	[2.41]**	[2.42]**	[2.26]**	[3.81]***	[3.58]***	[2.46]**	[2.79]***	[2.65]***
UK legal origin		1.735	3.092	4.15	3.499	2.094	5.013	4.464
		[0.67]	[1.40]	[1.42]	[1.56]	[0.89]	[2.30]**	[2.33]**
French legal origin		2.238	3.191	4.416	4.074	2.41	4.386	4.132
		[0.81]	[1.29]	[1.37]	[1.61]	[0.91]	[1.67]*	[1.78]*
Socialist legal origin		4.946	5.003	10.016	8.727	3.953	9.406	8.731
		[1.48]	[1.75]*	[2.72]***	[2.81]***	[1.33]	[3.05]***	[2.97]***
Scandinavian legal origin		1.729	5.142	6.479	6.042	4.388	6.538	6.506
		[0.44]	[1.16]	[1.47]	[1.45]	[0.98]	[1.76]*	[1.73]*
Catholic 1980			0.026	0.009	0.017	0.021	0.029	0.031
			[1.15]	[0.46]	[0.86]	[0.92]	[1.56]	[1.50]
Muslim 1980			−0.095	−0.051	−0.081	−0.1	0.003	−0.028
			[4.29]***	[2.63]***	[4.05]***	[4.77]***	[0.14]	[1.21]
Protestant 1980			−0.039	−0.048	−0.042	−0.046	−0.014	−0.023
			[1.10]	[1.33]	[1.25]	[1.23]	[0.37]	[0.61]
Population 1985–1995						−0.244	−0.024	−0.007
						[0.87]	[0.09]	[0.03]
GDP per capita 1985–1995				1.834	1.876		0.741	1.088
				[4.94]***	[4.08]***		[1.76]*	[2.15]**
Population above 65							0.843	0.672
							[4.99]***	[4.02]***
Small island	−6.231	−5.631	−5.488	−7.301	−6.683	−6.265	−7.01	−6.554
	[2.23]**	[2.06]**	[2.49]**	[3.11]***	[3.44]***	[2.60]**	[3.48]***	[3.66]***
Latitude	29.656	28.713	21.622	23.144	13.577	21.359	12.53	8.435
	[7.05]***	[5.77]***	[5.06]***	[6.39]***	[2.84]***	[5.15]***	[3.16]***	[1.86]*
Latin America	−1.533	−1.307	−7.852		−6.684	−7.939		−4.471
	[0.91]	[0.78]	[4.63]***		[4.39]***	[4.78]***		[3.00]***
Sub-Saharan Africa	−2.852	−2.499	−6.111		−3.585	−6.412		−1.815
	[1.58]	[1.39]	[3.75]***		[2.37]**	[4.10]***		[1.32]
East Asia & Pacific	−2.857	−1.819	−5.156		−5.207	−4.961		−3.626
	[1.30]	[0.83]	[2.19]**		[2.55]**	[2.13]**		[1.96]*
Constant	3.055	0.882	5.547	−14.068	−8.258	10.99	−10.91	−8.783
	[1.40]	[0.24]	[1.61]	[2.90]***	[1.82]*	[1.91]*	[1.66]	[1.27]
Observations	105	105	104	103	102	102	103	102
R-squared	0.6558	0.669	0.7921	0.7885	0.8241	0.7934	0.828	0.8459

Notes: Robust t-statistics in brackets.
*Significant at 10%; ** significant at 5%; *** significant at 1%.

7. Concluding Remarks

This paper has studied the effect of ethnolinguistic diversity on redistribution in a cross-section of countries. The main focus has been on the explicit introduction of linguistic distances into the measure of diversity. Our empirical investigation shows that when we control for distances, the effect of diversity on redistribution becomes highly significant, both statistically and economically. At the same time, ignoring distances typically leads to insignificant effects. Although the focus of this paper has been on diversity, we have also studied alternative measures of social tension, such as polarization and peripheral heterogeneity, and again, the crucial element that determines the significance of the results is whether linguistic distances are taken into account or not. The type of analysis conducted in this paper could be applied to study the effect of ethnolinguistic diversity (or polarization) on other economic variables, such as economic growth, the quality of government, civil conflicts, or the degree of decentralization. This is left for future research.

Appendix A: Description of the Data

- Languages spoken in each country and language trees. *Source: Ethnologue: Languages of the World, 15th Edition, SIL International, 2005, <www.ethnologue.com>.*
- Transfers and subsidies as percent of GDP: Average for 1985, 1990, and 1995. *Source: Economic Freedom Data Network.*
- GDP 1985–1995: Log GDP per capita (in constant 1995 dollars), average for the years between 1985 and 1995. *Source: World Bank.*
- Population 1985–1995: Log total population, average for the years between 1985 and 1995. *Source: World Bank.*
- Legal origin: Identifies the legal origin of the company law or the commercial code for each country. There are five possible origins: (1) English common law; (2) French commercial code; (3) German commercial code; (4) Scandinavian commercial code; and (5) Socialist/communist laws. *Source: La Porta et al. (1999).*
- Catholic 1980: The percentage of the population that is Catholic in 1980. *Source: La Porta et al. (1999).*
- Muslim 1980: The percentage of the population that is Muslim in 1980. *Source: La Porta et al. (1999).*
- Protestant 1980: The percentage of the population that is Protestant in 1980. *Source: La Porta et al. (1999).*
- Latitude: The absolute value of the latitude of the country, scaled to be between 0 and 1. *Source: La Porta et al. (1999).*
- Population above 65: Percentage of the population above 65. *Source: World Bank.*

- Small island dummy: Islands with a population of less than 0.5 million in 1990.

Appendix B: Indices of Diversity and Redistribution

Table B.1. Indices of diversity and polarization.

	Country	GI	ELF	ER	RQ	PH	Diff GI-ELF
1	Bolivia	0.650	0.680	0.201	0.207	0.463	54
2	Belize	0.624	0.693	0.172	0.188	0.433	51
3	United Arab Emirates	0.623	0.777	0.117	0.143	0.382	34
4	Papua New Guinea	0.598	0.990	0.005	0.009	0.038	−4
5	Suriname	0.595	0.788	0.114	0.158	0.301	29
6	Chad	0.591	0.950	0.029	0.045	0.177	−1
7	Mauritius	0.564	0.641	0.184	0.195	0.499	57
8	Qatar	0.545	0.608	0.203	0.217	0.500	59
9	New Caledonia	0.542	0.834	0.100	0.116	0.462	18
10	East Timor	0.540	0.897	0.055	0.091	0.119	7
11	Niger	0.540	0.646	0.157	0.183	0.420	51
12	Malaysia	0.525	0.758	0.101	0.143	0.341	28
13	Kazakhstan	0.521	0.701	0.152	0.187	0.369	36
14	Guatemala	0.518	0.691	0.153	0.166	0.500	39
15	Singapore	0.515	0.748	0.102	0.157	0.298	30
16	Iran	0.512	0.797	0.107	0.143	0.403	16
17	Congo	0.511	0.820	0.103	0.117	0.465	11
18	Cayman Islands	0.505	0.547	0.228	0.239	0.499	65
19	Fiji	0.503	0.607	0.198	0.221	0.493	49
20	Gibraltar	0.498	0.498	0.249	0.249	0.498	72
21	Namibia	0.488	0.808	0.081	0.134	0.239	9
22	Trinidad and Tobago	0.487	0.696	0.125	0.178	0.338	28
23	Kyrgyzstan	0.481	0.670	0.144	0.193	0.349	33
24	Kenya	0.472	0.901	0.043	0.084	0.135	−10
25	Bahrain	0.467	0.663	0.108	0.175	0.314	34
26	Laos	0.466	0.678	0.102	0.149	0.336	29
27	Nigeria	0.463	0.870	0.061	0.104	0.137	−8
28	Uganda	0.461	0.928	0.029	0.064	0.119	−18
29	Guam	0.458	0.640	0.139	0.204	0.317	36
30	Estonia	0.457	0.476	0.214	0.217	0.451	71
31	Sudan	0.457	0.587	0.129	0.153	0.393	44
32	Georgia	0.453	0.576	0.130	0.165	0.366	45
33	Bhutan	0.442	0.846	0.073	0.123	0.335	−8
34	Mayotte	0.433	0.459	0.197	0.207	0.410	69
35	India	0.427	0.930	0.027	0.061	0.111	−27
36	Israel	0.407	0.665	0.094	0.158	0.296	22
37	South Africa	0.394	0.869	0.048	0.109	0.118	−17
38	French Polynesia	0.385	0.596	0.112	0.176	0.284	32
39	Central African Republic	0.385	0.960	0.017	0.038	0.183	−35
40	Dem Rep of Congo	0.376	0.948	0.021	0.046	0.077	−34
41	Cyprus	0.361	0.366	0.178	0.180	0.358	79
42	Tajikistan	0.360	0.482	0.138	0.177	0.320	57
43	Aruba	0.359	0.387	0.153	0.156	0.355	71
44	Peru	0.350	0.376	0.137	0.138	0.336	74
45	Azerbaijan	0.349	0.373	0.142	0.147	0.332	74

Table B.1. (*Continued*)

	Country	GI	ELF	ER	RQ	PH	Diff GI-ELF
46	Northern Mariana Islands	0.341	0.642	0.102	0.196	0.234	17
47	Nepal	0.333	0.742	0.057	0.144	0.201	−1
48	Iraq	0.328	0.666	0.079	0.187	0.205	9
49	Afghanistan	0.325	0.732	0.064	0.160	0.186	−1
50	Paraguay	0.322	0.347	0.137	0.141	0.314	74
51	Panama	0.322	0.324	0.136	0.136	0.303	78
52	U.S. Virgin Islands	0.319	0.339	0.144	0.147	0.316	74
53	Sao Tome e Principe	0.311	0.389	0.128	0.161	0.282	59
54	Dominica	0.308	0.313	0.153	0.155	0.306	76
55	Sri Lanka	0.306	0.313	0.150	0.153	0.303	77
56	French Guiana	0.304	0.480	0.095	0.162	0.244	44
57	Mali	0.303	0.876	0.027	0.095	0.113	−39
58	Solomon Islands	0.297	0.965	0.011	0.033	0.033	−56
59	Bahamas	0.295	0.386	0.128	0.166	0.272	56
60	Guinea-Bissau	0.278	0.853	0.039	0.119	0.090	−38
61	Oman	0.269	0.693	0.056	0.184	0.149	−10
62	Brunei	0.265	0.456	0.085	0.152	0.221	42
63	Uzbekistan	0.263	0.428	0.093	0.150	0.228	45
64	Turkey	0.258	0.289	0.108	0.119	0.245	70
65	Thailand	0.254	0.753	0.040	0.168	0.111	−24
66	Myanmar	0.254	0.521	0.072	0.155	0.196	22
67	Nauru	0.254	0.596	0.066	0.182	0.172	2
68	Cameroon	0.248	0.942	0.009	0.049	0.062	−61
69	Sierra Leone	0.245	0.817	0.038	0.135	0.101	−40
70	Ecuador	0.243	0.264	0.109	0.115	0.238	68
71	Greenland	0.242	0.242	0.121	0.121	0.242	70
72	Vanuatu	0.238	0.972	0.008	0.027	0.016	−71
73	Turkmenistan	0.221	0.386	0.082	0.146	0.194	43
74	Macedonia	0.212	0.566	0.064	0.203	0.152	5
75	Syria	0.203	0.503	0.062	0.172	0.157	16
76	Slovakia	0.196	0.307	0.084	0.130	0.180	57
77	Saudi Arabia	0.191	0.609	0.047	0.215	0.111	−11
78	Russia	0.183	0.283	0.071	0.112	0.165	57
79	Bulgaria	0.178	0.224	0.083	0.103	0.172	65
80	Eritrea	0.178	0.749	0.030	0.164	0.088	−38
81	Mauritania	0.170	0.172	0.079	0.079	0.167	72
82	British Virgin Islands	0.167	0.167	0.084	0.084	0.167	74
83	Viet Nam	0.161	0.234	0.067	0.097	0.150	59
84	Hungary	0.153	0.158	0.071	0.071	0.153	74
85	Tanzania	0.149	0.965	0.004	0.033	0.022	−82
86	Turks and Caicos Islands	0.145	0.145	0.073	0.073	0.145	74
87	Lebanon	0.144	0.161	0.068	0.075	0.142	70
88	Liberia	0.142	0.912	0.011	0.076	0.043	−75
89	Anguilla	0.140	0.140	0.070	0.070	0.140	72
90	Indonesia	0.138	0.846	0.015	0.110	0.064	−66
91	Netherlands	0.137	0.389	0.045	0.142	0.113	22
92	Gabon	0.137	0.919	0.010	0.070	0.032	−81
93	Romania	0.134	0.168	0.063	0.079	0.130	61
94	Canada	0.129	0.549	0.038	0.193	0.097	−12
95	Argentina	0.127	0.213	0.054	0.093	0.118	51
96	Mexico	0.127	0.135	0.057	0.061	0.123	69
97	Micronesia	0.124	0.792	0.017	0.146	0.049	−64

TABLE B.1. (*Continued*)

	Country	GI	ELF	ER	RQ	PH	Diff GI-ELF
98	Moldova	0.122	0.589	0.034	0.183	0.091	−24
99	Finland	0.121	0.140	0.059	0.067	0.120	63
100	Equatorial Guinea	0.112	0.453	0.034	0.159	0.086	5
101	Serbia and Montenegro	0.112	0.359	0.042	0.150	0.096	21
102	Belgium	0.110	0.734	0.024	0.177	0.065	−55
103	Ethiopia	0.109	0.843	0.014	0.116	0.057	−77
104	China	0.107	0.491	0.031	0.155	0.083	−8
105	Djibouti	0.099	0.592	0.030	0.228	0.064	−32
106	France	0.097	0.272	0.038	0.109	0.087	30
107	Netherlands Antilles	0.097	0.266	0.039	0.115	0.088	30
108	Armenia	0.096	0.174	0.043	0.079	0.091	43
109	New Zealand	0.095	0.102	0.045	0.048	0.094	61
110	Cambodia	0.094	0.157	0.042	0.072	0.090	49
111	Gambia	0.094	0.748	0.021	0.167	0.064	−68
112	USA	0.092	0.353	0.033	0.136	0.078	11
113	Burkina Faso	0.091	0.773	0.014	0.133	0.051	−76
114	Pakistan	0.091	0.762	0.015	0.148	0.051	−75
115	Barbados	0.091	0.091	0.046	0.046	0.091	57
116	Philippines	0.090	0.849	0.011	0.116	0.036	−93
117	American Samoa	0.088	0.116	0.041	0.055	0.086	52
118	Sweden	0.088	0.167	0.039	0.074	0.083	37
119	Cote d'Ivoire	0.085	0.917	0.006	0.069	0.029	−107
120	Nicaragua	0.081	0.081	0.040	0.040	0.081	55
121	Switzerland	0.079	0.547	0.026	0.182	0.067	−37
122	Saint Pierre and Miquelon	0.078	0.134	0.037	0.063	0.076	44
123	Guinea	0.078	0.748	0.019	0.170	0.058	−79
124	Guyana	0.077	0.078	0.037	0.037	0.076	52
125	Latvia	0.077	0.595	0.027	0.211	0.065	−54
126	Botswana	0.076	0.444	0.023	0.154	0.059	−20
127	Taiwan	0.076	0.488	0.026	0.189	0.062	−30
128	Senegal	0.076	0.772	0.013	0.156	0.039	−90
129	Benin	0.075	0.901	0.006	0.079	0.029	−114
130	Niue	0.071	0.071	0.036	0.036	0.071	48
131	Cape Verde Islands	0.070	0.070	0.035	0.035	0.070	48
132	Luxembourg	0.069	0.498	0.026	0.190	0.061	−39
133	Germany	0.067	0.189	0.029	0.082	0.063	15
134	Malawi	0.067	0.519	0.018	0.167	0.048	−45
135	Seychelles	0.066	0.067	0.033	0.033	0.066	46
136	Reunion	0.066	0.066	0.032	0.032	0.065	46
137	Ukraine	0.066	0.492	0.022	0.194	0.053	−42
138	Yemen	0.065	0.579	0.020	0.229	0.044	−62
139	Cook Islands	0.064	0.379	0.022	0.150	0.052	−22
140	Morocco	0.062	0.466	0.023	0.172	0.057	−38
141	Zimbabwe	0.060	0.526	0.016	0.165	0.043	−55
142	Togo	0.060	0.897	0.006	0.086	0.023	−126
143	Antigua and Barbuda	0.057	0.057	0.028	0.028	0.056	42
144	Mongolia	0.054	0.331	0.020	0.133	0.047	−16
145	Tokelau	0.054	0.054	0.027	0.027	0.054	42
146	Bosnia and Herzegovina	0.053	0.416	0.018	0.158	0.044	−37
147	Dominican Republic	0.051	0.053	0.025	0.026	0.051	41
148	Jordan	0.051	0.484	0.016	0.175	0.040	−50
149	Madagascar	0.051	0.656	0.011	0.164	0.035	−88

TABLE B.1. (*Continued*)

	Country	GI	ELF	ER	RQ	PH	Diff GI-ELF
150	Ghana	0.050	0.805	0.007	0.123	0.029	−119
151	Costa Rica	0.049	0.050	0.024	0.024	0.049	39
152	Egypt	0.048	0.509	0.016	0.205	0.038	−62
153	Spain	0.046	0.438	0.015	0.174	0.037	−46
154	Honduras	0.046	0.056	0.022	0.027	0.046	32
155	Algeria	0.046	0.313	0.019	0.133	0.043	−24
156	Lithuania	0.044	0.339	0.018	0.140	0.042	−31
157	United Kingdom	0.044	0.139	0.020	0.063	0.042	6
158	Martinique	0.043	0.043	0.021	0.021	0.043	34
159	Austria	0.042	0.540	0.016	0.239	0.033	−74
160	Greece	0.041	0.175	0.018	0.078	0.039	−10
161	Australia	0.039	0.126	0.018	0.058	0.038	7
162	Italy	0.039	0.593	0.011	0.167	0.031	−90
163	Angola	0.038	0.785	0.006	0.147	0.019	−128
164	Libya	0.038	0.362	0.015	0.163	0.033	−43
165	Guadeloupe	0.038	0.084	0.018	0.041	0.037	9
166	Belarus	0.037	0.397	0.014	0.156	0.033	−55
167	Zambia	0.035	0.855	0.004	0.110	0.015	−146
168	Bangladesh	0.034	0.332	0.013	0.140	0.030	−41
169	San Marino	0.032	0.494	0.016	0.247	0.032	−75
170	Kuwait	0.031	0.556	0.011	0.201	0.026	−90
171	Chile	0.030	0.034	0.015	0.017	0.030	22
172	Mozambique	0.029	0.929	0.002	0.064	0.008	−163
173	Ireland	0.028	0.223	0.013	0.103	0.027	−28
174	Albania	0.028	0.257	0.012	0.113	0.026	−34
175	Marshall Islands	0.027	0.027	0.013	0.013	0.027	23
176	Colombia	0.026	0.030	0.013	0.015	0.026	20
177	Montserrat	0.026	0.026	0.013	0.013	0.026	23
178	Somalia	0.025	0.179	0.011	0.083	0.023	−29
179	Venezuela	0.024	0.026	0.012	0.013	0.024	20
180	Andorra	0.024	0.574	0.009	0.227	0.020	−102
181	Grenada	0.022	0.064	0.011	0.032	0.022	2
182	Wallis and Futuna	0.022	0.407	0.008	0.198	0.017	−72
183	Comoros	0.022	0.551	0.007	0.236	0.016	−102
184	Monaco	0.022	0.521	0.008	0.200	0.019	−97
185	Slovenia	0.021	0.174	0.009	0.084	0.020	−33
186	Saint Lucia	0.020	0.020	0.010	0.010	0.020	16
187	Denmark	0.018	0.051	0.009	0.025	0.017	2
188	Kiribati	0.017	0.033	0.008	0.016	0.017	6
189	Malta	0.016	0.016	0.008	0.008	0.016	15
190	Palestinian West Bank and Gaza	0.015	0.208	0.007	0.102	0.015	−43
191	Japan	0.014	0.028	0.007	0.014	0.014	6
192	Portugal	0.011	0.022	0.006	0.011	0.011	9
193	Jamaica	0.011	0.011	0.006	0.006	0.011	14
194	Brazil	0.011	0.032	0.005	0.016	0.011	1
195	Croatia	0.010	0.087	0.005	0.042	0.010	−22
196	Saint Kitts and Nevis	0.010	0.010	0.005	0.005	0.010	12
197	Liechtenstein	0.010	0.128	0.005	0.061	0.010	−30
198	Saint Vincent and the Grenadines	0.009	0.009	0.004	0.004	0.009	12
199	Czech Republic	0.008	0.069	0.004	0.033	0.008	−19
200	Palau	0.007	0.077	0.004	0.039	0.007	−23
201	Uruguay	0.007	0.092	0.003	0.044	0.007	−30
202	Lesotho	0.007	0.260	0.003	0.121	0.006	−63

TABLE B.1. (*Continued*)

	Country	GI	ELF	ER	RQ	PH	Diff GI-ELF
203	Puerto Rico	0.006	0.049	0.003	0.024	0.006	−12
204	Tuvalu	0.006	0.139	0.003	0.070	0.006	−40
205	Poland	0.006	0.060	0.003	0.029	0.006	−21
206	Tunisia	0.005	0.012	0.003	0.006	0.005	0
207	Swaziland	0.005	0.228	0.002	0.107	0.005	−64
208	El Salvador	0.004	0.004	0.002	0.002	0.004	3
209	Korea, South	0.003	0.003	0.002	0.002	0.003	5
210	Samoa	0.002	0.002	0.001	0.001	0.002	5
211	Iceland	0.001	0.019	0.001	0.010	0.001	−8
212	Burundi	0.001	0.004	0.001	0.002	0.001	0
213	Rwanda	0.001	0.004	0.000	0.002	0.001	0
214	Maldives	0.001	0.010	0.000	0.005	0.001	−5
215	Tonga	0.000	0.015	0.000	0.008	0.000	−10
216	Haiti	0.000	0.000	0.000	0.000	0.000	1
217	Cuba	0.000	0.001	0.000	0.000	0.000	−1
218	Bermuda	0.000	0.000	0.000	0.000	0.000	0
219	British Indian Ocean Territory	0.000	0.000	0.000	0.000	0.000	0
220	Falkland Islands	0.000	0.000	0.000	0.000	0.000	0
221	Korea, North	0.000	0.000	0.000	0.000	0.000	0
222	Norfolk Island	0.000	0.000	0.000	0.000	0.000	0
223	Pitcairn	0.000	0.000	0.000	0.000	0.000	0
224	Saint Helena	0.000	0.000	0.000	0.000	0.000	0
225	Vatican State	0.000	0.000	0.000	0.000	0.000	0

References

Alesina, Alberto, Reza Baqir, and William Easterly (1999). “Public Goods and Ethnic Divisions.” *Quarterly Journal of Economics*, 114, 1243–1284.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003). “Fractionalization.” *Journal of Economic Growth*, 8(2), 155–194.

Alesina, Alberto, Edward Glaeser, and Bruce Sacerdote (2001). “Why Doesn’t the US Have a European Style Welfare System?” *Brookings Papers on Economic Activity*, Fall, 187–278.

Alesina, Alberto, and Eliana La Ferrara (2005).“Ethnic Diversity and Economic Performance.” *Journal of Economic Literature*, 43(3), 762–800.

Alesina, Alberto, and Romain Wacziarg (1998). “Openness, Country Size and Government.” *Journal of Public Economics*, 69, 305–321.

Atlas Narodov Mira (1964). Moscow: Miklukho-Maklai Ethnological Institute at the Department of Geodesy and Cartography of the State Geological Committee of the Soviet Union.

Becker, Gary (1957). *The Economics of Discrimination*. University of Chicago Press.

Bossert, Walter, Conchita D’Ambrosio, and Eliana La Ferrara (2006). “A Generalized Index of Ethno-Linguistic Fractionalization.” Working Paper No. 313, IGIER, Università Bocconi.

Caselli, Francesco, and Wilbur John Coleman II (2006). “On the Theory of Ethnic Conflict.” NBER Working Paper No. 12125.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Shlomo Weber (2005). “Peripheral Diversity and Redistribution.” CEPR Discussion Paper No. 5112.

Duclos, Jean-Yves, Joan Maria Esteban, and Debraj Ray (2004). “ Polarization: Concepts, Measurement, Estimation.” *Econometrica*, 72, 1737–1772.

Dyen, Isidore, Joseph B. Kruskal, and Paul Black (1992). “An Indo-European Classification: A Lexicostatistical Experiment.” *Transactions of the American Philosophical Society*, 82(5).

Easterly, William, and Ross Levine (1997). “Africa’s Growth Tragedy: Policies and Ethnic Divisions.” *Quarterly Journal of Economics*, 112(4), 1203–1250.
Esteban, Joan Maria, and Debraj Ray (1994). “On the Measurement of Polarization.” *Econometrica*, 62(4), 819–851.
Esteban, Joan Maria, and Debraj Ray (1999). “Conflict and Distribution.” *Journal of Economic Theory*, 87, 379–415.
Esteban, Joan Maria, and Debraj Ray, (2006). “Polarization, Fractionalization and Confict.” Working paper.
Fearon, James D. (2003). “Ethnic and Cultural Diversity by Country.” *Journal of Economic Growth*, 8, 195–222.
Fernández, Raquel, and Gilat Levy (2007). “Diversity and Redistribution.” *Journal of Public Economics*, forthcoming.
Ginsburgh, Victor, Ignacio Ortuño-Ortín, and Shlomo Weber (2005). “Disenfranchisement in Linguistically Diverse Societies: The Case of the European Union.” *Journal of the European Economic Association*, 3(4), 946–965.
Greenberg, Joseph H. (1956). “The Measurement of Linguistic Diversity.” *Language* 32, 109–115.
Laponce, Jean (1992). “Language and Politics.” In *Encyclopedia of Government and Politics* 1, edited by Hawkesworth, M and Hogan, M. Routledge.
La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert Vishny (1999). “The Quality of Government.” *Journal of Law, Economics and Organization*, 15(1), 222–279.
Lee, Woojin, and John E. Roemer (2004). “Racism and Redistribution in the United States: A Solution to the Problem of American Exceptionalism.” Working paper.
Lind, Jo Thori (2007). “Fractionalization and the Size of the Government.” *Journal of Public Economics*, 91, 51–76.
Luttmer, Erzo (2001). “Group Loyalty and the Taste for Redistribution.” *Journal of Political Economy* 109, 500–528.
Montalvo, Jose G., and Marta Reynal-Querol (2002). “Why Ethnic Fractionalization? Polarization, Ethnic Conflict and Growth.” Working Paper No. 660, Universitat Pompeu Fabra.
Montalvo, Jose G., and Marta Reynal-Querol (2005). “Ethnic Polarization, Potential Conflict, and Civil Wars.” *American Economic Review*, 95, 796–815.
Nei, Masatoshi, and Wen-Hsiung Li (1979). “Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases.” *Proceedings of the National Academy of Science of the United States of America*, 76, 5269–5273.
Pavoine, Sandrine, Sébastien Ollier, and Dominique Pontier (2005). “Measuring Diversity from Dissimilarities with Rao’s Quadratic Entropy: Are Any Dissimilarities Suitable?” *Theoretical Population Biology*, 67, 231–239.
Rao, C. R. (1982). “Diversity and Dissimilarity Coefficients: A Unified Approach.” *Theoretical Population Biology*, 21, 24–43.
Reynal-Querol, Marta (2002). “Ethnicity, Political Systems, and Civil Wars.” *Journal of Conflict Resolution*, 46(1), 29–54.
Ricotta, Carlo (2005). “Through the Jungle of Biological Diversity.” *Acta Biotheoretica*, 53, 29–38.
Ricotta, Carlo, and Laszlo Szeidl (2006). “Towards a Unifying Approach to Diversity Measures: Bridging the Gap between Shannon Entropy and Rao’s Quadratic Index.” *Theoretical Population Biology*, 70, 237–243.
Shannon, Claude E. (1949). *The Mathematical Theory of Communication*. University of Illinois Press.
Spolaore, Enrico, and Romain Wacziarg (2006). “The Diffusion of Development.” NBER Working Paper No. 12153.
Vigdor, Jacob L. (2004). “Community Heterogeneity and Collective Action: Initial Mail Response to the 2000 Census.” *Review of Economics and Statistics*, 86, 303–312.