DISCUSSION PAPER SERIES

No. 5112

PERIPHERAL DIVERSITY AND REDISTRIBUTION

Klaus Desmet, Ignacio Ortuño-Ortín
and Shlomo Weber

PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP5112.asp

ISSN 0265-8003

PERIPHERAL DIVERSITY AND REDISTRIBUTION

Klaus Desmet, Universidad Carlos III de Madrid and CEPR
Ignacio Ortuño-Ortín, Universidad de Alicante
Shlomo Weber, CORE, UCL and CEPR

Discussion Paper No. 5112
June 2005

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5112

June 2005

ABSTRACT

Peripheral Diversity and Redistribution*

This paper investigates the effect of ethnolinguistic conflict on redistribution. The analysis focuses on the conflict arising between 'peripheral' minority groups and a dominant 'centre'. We propose an index of linguistic conflict that (i) encompasses both diversity and polarization, and (ii) accounts for the distance between languages. Our results suggest that linguistic diversity is a better predictor of redistribution than linguistic polarization. We also find that incorporating linguistic distances improves the predictive power of our conclusions.

JEL Classification: D60, D74, H50 and Z10
Keywords: linguistic diversity, polarization and redistribution

Klaus Desmet
Departamento de Economía
Universidad Carlos III
28903 Getafe, (Madrid)
SPAIN
Tel: (34 91) 624 9845
Fax: (34 91) 624 9875
Email: klaus.desmet@uc3m.es

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=144849

Ignacio Ortuño-Ortín
Department of Economics
University de Alicante
Campus de Sant Vicent
03071 Alicante
SPAIN
Tel: (34 96) 590 34 00 ext 3223
Fax: (34 96) 590 38 98
Email: ortin@merlin.fae.ua.es

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=154176

Shlomo Weber
CORE - UCL
34, voie du Roman Pays
B-1348 Louvain la Neuve
Brussels
BELGIUM
Email: weber@core.ucl.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=161025

*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, Anna Rubinchik-Pessach and Ekaterina Zhuravskaya for useful comments.

Submitted 07 June 2005

Peripheral Diversity and Redistribution*

Klaus Desmet[†], Ignacio Ortuño-Ortín[‡] and Shlomo Weber[§]

April 2005

Abstract

This paper investigates the effect of ethnolinguistic conflict on redistribution. The analysis focuses on the conflict arising between "peripheral" minority groups and a dominant "center". We propose an index of linguistic conflict that (i) encompasses both diversity and polarization, and (ii) accounts for the distance between lanuguages. Our results suggest that linguistic diversity is a better predictor of redistribution than linguistic polarization. We also find that incorporating linguistic distances leads to better predictions.

JEL: D6; D74; H5; Z10
Key words: Linguistic diversity; redistribution; polarization.

1 Introduction

This paper explores how ethnolinguistic diversity between a central dominant group and minority groups affects redistribution. The literature has long argued that ethnolinguistic diversity reduces government transfers.[1] Altruistic attitudes seem to be more prevalent within homogenous groups than across ethnically or culturally diverse groups.[2] In the

*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, Anna Rubinchik-Pessach and Ekaterina Zhuravskaya for useful comments.

†Universidad Carlos III and CEPR. Email: `klaus.desmet@uc3m.es`

‡Universidad de Alicante. Email: `ortin@merlin.fae.ua.es`

§CORE, SMU and CEPR. Email : `weber@core.ucl.ac.be`

[1] For empirical studies, see, for instance, Alesina et al. (2003), La Porta et al. (1999), Easterly and Levine (1997) and Luttmer (2001). Recent theoretical treatments of the relation between diversity and redistribution can be found in, for instance, Caselli (2002), Fernandez and Levy (2004), and Lee and Roemer (2004). For a more general survey on diversity and different measures of economic performance, see Alesina and La Ferrara (2004).

[2] See, for example, Vidgor (2004).

United States, where there is a strong racial component to the income distribution, the poor are considered as "other" (Alesina et al., 2001). In Western Europe the poor are instead viewed as "unlucky". If, as posited by Becker (1957), individuals have stronger feelings of empathy towards their own group, it is not surprising that the U.S. exhibits lower levels of redistribution than Western European countries.

Compared to the existing literature on ethnolinguistic diversity and redistribution, we introduce three novelties. First, we focus on ethnolinguistic conflict that arises between a dominant "central" group and "peripheral" minority groups. Second, we explicitly allow for distances between groups. Third, we design an index able to capture both the notion of ethnolinguistic diversity and ethnolinguistic polarization. We now discuss each one of these contributions in detail.

A first novelty is to analyze ethnolinguistic conflict that originates between the "central" dominant group and several "peripheral" minorities. The standard approach to measuring diversity treats all groups in a symmetric way, and thus exclusively focuses on the number and the relative sizes of groups. However, in many societies conflict arises from the antagonism that minorities feel towards the dominant group, rather than from tension between all groups. To illustrate our approach, take the example of Spain. We will assume that linguistic conflict arises mainly between speakers of Spanish and the different minority languages (Catalan, Basque and Galician), rather than between the minority groups themselves. Our goal, then, is to construct an index of *peripheral ethnolinguistic diversity*. We will focus on linguistic, rather than ethnical, diversity, where the dominant group either speaks the official language or is the largest group.

A second novelty is to include the distance between groups into our measure of diversity. The most commonly used measure in the literature is the probability that two randomly selected individuals belong to different groups. As with Shannon's diversity index (Shannon, 1949), this so-called fractionalization index fails to take into account the degree of distinctiveness between different groups. However, one would probably agree that a country (Andorra) where roughly a half of the population speaks Catalan and the other half speaks Spanish (two similar Romance languages) is less diverse than a country

(Belgium) where about 60% of the population speaks Dutch and the other 40% speak French (a Germanic and a Romance language, respectively). In spite of that, in Alesina et al. (2003) Andorra is ranked as more diverse than either Belgium or Switzerland that contains three sizeable linguistic groups speaking German, French and Italian. Our prior is that the scope for conflict expands as the linguistic distance between groups grows. In the specific case of redistribution, we would expect altruism to decline as ethnolinguistic distance becomes larger. This is in line with the Becker (1957) view: individuals like their own group more, and how much they dislike other groups depends on how different they are. An appropriate analysis of ethnolinguistic diversity should therefore take into account the dissimilarity between groups. This point has previously been made by Caselli and Coleman (2002) and Alesina and La Ferrara (2004).

In spite of the importance of distance between various ethnic groups, a theoretical and empirical investigation of the issue has been lagging behind. Moreover, most of the diversity indices that incorporate distances fail to satisfy our purposes. For instance, Weitzman (1992) and Nehring and Puppe (2002) overlook the importance of the relative sizes of the different groups, and exclusively focus on the dissimilarities between groups. A more relevant approach is the one proposed by Greenberg (1956), and more recently, by Laitin (2000) and Fearon (2003), that generalizes the existing fractionalization index by weighting the product of the sizes of any two groups by their linguistic distances. Our index differs from theirs in a number of respects. First, we analyze the peripheral diversity, so that the nature of the index is different. We do not treat diversity in a symmetric way, but rather focus on the diversity that emanates from the tension between the central dominant group and the peripheral minority groups. Second, our measure of distance is based on the proportion of cognates between languages elaborated by Dyen, Kruskal and Black (1992). In contrast, Fearon (2003) utilizes language trees to compute distances. As will be discussed in the data section, we opt for the Dyen et al. (1992) distance matrix, because it provides more detailed information than language trees. Although the Dyen measure has been used previously in economics by Ginsburgh, Ortuño-Ortín and Weber (2005), it has not been applied to study linguistic diversity and polarization.

The relevance of linguistic distances is an empirical question and a test should be conducted whether this feature improves upon the existing empirical results. To do so, we compare the index that includes distances to the index that fails to do so, and search for the one that provides a better explanation of the level of redistribution. We find that the explanatory power of the index with linguistic distances is superior. This result is shown to be robust to a number of different specifications. Quantitatively the effect of peripheral diversity is important. With an average level of redistribution of 14%, an increase of one standard deviation in the degree of diversity lowers redistribution by around 1.6 percentage points. When comparing highly diverse countries to less diverse countries, the effect is obviously bigger. For instance, the model predicts redistribution in Canada to be 5.9 percentage points lower than in Denmark.

A third novelty of our paper, compared to existing work, is that our index encompasses both the notions of diversity and polarization. By designing such an index, we bring together two branches of the literature. On the one hand, the work on *diversity* — or fractionalization — has argued that the degree of social conflict is positively correlated with the level of diversity. Alesina et al. (2003), for instance, look at the impact of ethnolinguistic diversity on different economic variables, such as economic growth and the size of the welfare state. On the other hand, social conflict may also depend on a society's degree of *polarization* (Esteban and Ray, (1994) and (2004)). To give some intuition about the difference between polarization and diversity, suppose country A is composed of two equally sized linguistic groups and country B of three equally sized linguistic groups. In that case, country A would be more polarized, but less diverse, than country B.[3] Although our main focus is on diversity, our index is flexible enough to account for both diversity and polarization.

Again, the question whether diversity or polarization is a better predictor of redistribution is an empirical one. In the empirical section we state our belief that a higher degree of linguistic conflict lowers redistribution. We then let the data determine

[3] A third view, in between diversity and polarization, claims that social conflict may depend on the degree of *disenfranchisement* (Ginsburgh et al., 2005), where the linguistic disenfranchisement can be viewed as the cost incurred by minority language speakers to learn the majority language.

which measure of peripheral linguistic conflict — diversity or polarization — provides a better explanation of redistribution. We find that diversity tends to be a better predictor, though there is substantial overlap between both notions of linguistic conflict.

2 The General Model

Consider a country whose population consists of N citizens and is partitioned in $K+1$ distinct groups, labeled $0, 1, \ldots, K$. One group, 0, called "center" or "dominant group", consists of N_0 individuals, whereas the other K groups, called "minorities", consist of N_k individuals each. We impose no condition on the geographical distribution of the groups. Thus, it might be the case that all individuals from a group live in the same region or that individuals from different groups live in the same region. Each citizen of the country belongs to one and only one group, i.e.,

$$N = \sum_{k=0}^{K} N_k.$$

It will be convenient to deal with the population share of each group rather than its absolute size, and for every $k = 0, \ldots, K$ we denote

$$s_k = \frac{N_k}{N}$$

That is,

$$\sum_{k=0}^{K} s_k = 1,$$

where the vector $(s_0, s_1, \ldots, s_K)$ belongs to the $k+1$-dimensional simplex Δ. Note that a country's population size will not matter in our analysis.

Our model focuses on the frequently observed cases where the "dominant" group contains at least as many individuals as any of the minority groups:

$$s_0 \geq \max_{k=1,\ldots,K} s_k.$$

We examine the subset of vectors $\mathcal{S} \subset \Delta$ given by:

$$\mathcal{S} = \{s = (s_0, s_1, ..., s_K) \in \Delta \mid s_0 \geq \max_{k=1,\ldots,K} s_k\},$$

and denote by $\mathcal{S}^k$ the set of pairs (s_0, s_k) for which there exists $s \in \mathcal{S}$, whose first and $k+1$-coordinates coincide with s_0 and s_k.

A crucial element of our model is the introduction of ethnolinguistic distance between groups. Thus, there is a matrix T that assigns the distance τ_{kl} to each pair of groups k and l. We assume that all values τ_{kl} lie between 0 and 1, and $\tau_{kl} = \tau_{lk}$. The set of such matrices is denoted by $\mathcal{T}$. In this paper we identify such distance between groups with the linguistic distance, i.e., groups are formed by individuals who speak the same language and τ_{kl} is the linguistic distance between the language spoken by group k and the language spoken by group l.[4]

The population shares and linguistic distances will be enough to determine the level of *peripheral diversity.* To do so, we proceed in three steps. First, we define the notion of inter-group alienation. Second, we use this concept to define peripheral alienation. Third, we show that under certain axioms peripheral alienation can be interpreted as peripheral diversity.

We start by defining the notion of inter-group alienation. Formally, we assume there exists an alienation function f such that the value of inter-group alienation between groups k and l is given by the value of the function f,

$$f_k(s_k, s_l, \tau_{kl}),$$

which depends on the size of both groups and their linguistic distance. At this point we do not provide any type of individual foundation of such group functions f_k and simply assume that all functions f_k are identical and are equal to the function f. In Section 3 we discuss some examples in which the function f is constructed from an alienation function at the individual level.[5]

As alluded to previously, we consider the notion of *peripheral alienation*, where the only type of inter-group alienation accounted for is the one from the minority groups to-

[4] This is similar to the resemblance function of Greenberg (1956).

[5] The way our diversity index is constructed is based more on a "primordialist" view of ethnolinguistic conflict than on an "instrumentalist" view. The primordialist view essentially says that the ethnic composition of society enters directly into the utility function of the agents. See Caselli and Coleman (2002) for a discussion of these two categories of ethnolinguistic conflict.

wards the center (centrifugal alienation) and from the center towards the minority groups (centripetal alienation). Thus, the relevant information needed in our analysis consists of the functions $f(s_0, s_k, \tau_{0k})$, that give the centrifugal alienation experienced by each of the $k = 1, ..., K$ minority groups towards the center, and the functions $f_0(s_0, s_k, \tau_{0k})$, that give the centrifugal alienation experienced by the center towards each of the $k = 1, ..., K$ minority groups. Since the the center is the dominant group, it is natural to allow different functional forms of f and f_0 of alienation towards the dominant group and towards the minorities.

For every vector $s = (s_0, ..., s_K) \in \mathcal{S}$, distance matrix $T \in \mathcal{T}$, and the alienation functions f and f_0, we define the *total level of peripheral alienation* $PD(s, T)$ as

$$PD(s,T) = \sum_{k=1}^{K} \left(f(s_0, s_k, \tau_{0k}) + f_0(s_0, s_k, \tau_{0k}) \right). \tag{1}$$

The following conditions introduce some more structure, and will allow us to interpret $PD(s, T)$ as a measure of *peripheral diversity.*

Condition 1 (Continuity): The functions f and f_0 are continuous on $\mathcal{S}^k$.

Condition 2 (Alienation is increasing in distance): For every $(s_0, s_k) \in \mathcal{S}^k$, the functions $f(s_0, s_k, .)$ and $f_0(s_0, s_k, .)$ are strictly increasing on the interval $[0, 1]$.

Condition 3 (Concavity): (i) For every $s_0 \geq \frac{1}{k+1}$ and $\tau \in [0, 1]$, the function $f(s_0, \ldots, \tau)$ is strictly concave on the interval $[0, \min[s_0, 1 - s_0]]$;
(ii) For every $s_k \leq \frac{1}{2}$ and $\tau \in [0, 1]$, the function $f_0(\ldots, s_k, \tau)$ is strictly concave on the interval $[\max[s_k, \frac{1}{k+1}], 1 - s_k]$.

Condition 4 (Supermodularity): Let $s^1 < s^2$, and $\tau^1 < \tau^2$. Then

$$f(s_0, s^2, \tau^1) - f(s_0, s^1, \tau^1) < f(s_0, s^2, \tau^2) - f(s_0, s^1, \tau^2)$$

and

$$f_0(s_0, s^2, \tau^1) - f_0(s_0, s^1, \tau^1) < f_0(s_0, s^2, \tau^2) - f_0(s_0, s^1, \tau^2).$$

Condition 1 and 2 impose continuity and monotonicity. Condition 3 is the key one to obtain an index of diversity. If f is concave in the size of the group, smaller groups experience, in "per capita" terms, more alienation than larger groups. In the case the alienation functions are differentiable, the supermodularity condition implies that $\frac{\partial f(s_0,s,\tau)}{\partial s \partial \tau} > 0$.

The following proposition states that, in order to increase peripheral alienation, the distant minority groups should be large:

Proposition 1: Assume that Conditions 1, 2, 3 and 4 hold. Let the matrix T and the vector $\overline{s} \in \mathcal{S}$ be given and consider the subset $\mathcal{S}_{kl}(\overline{s})$ of population shares in $\mathcal{S}$ such that for all $s \in \mathcal{S}_{kl}(\overline{s})$ the share of the center is $s_0 = \overline{s_0}$. Let two minority groups k, l be such that $\tau_{0k} \geq \tau_{0l}$. Suppose that the maximization problem

$$\max_{s \in \mathcal{S}_{kl}(\overline{s})} PD(s,T)$$

has a unique solution denoted by $s^* \in \mathcal{S}_{kl}(\overline{s})$. Then

$$s_k^* \geq s_l^*.$$

Proof: See Appendix A.

Proposition 1 says that, if $\tau_{0k} \geq \tau_{0l}$, i.e., if group k is more distant from the center than group l, maximum peripheral alienation should satisfy $s_k^* \geq s_l^*$. Note that when $\tau_{0k} = \tau_{0l}$ the proposition implies that $s_k^* = s_l^*$. In this case, the problem resembles the traditional approach to diversity where only the sizes of the groups matter. In that context it is commonly assumed that an index of diversity should satisfy a property similar to the one stated in Proposition 1, namely that diversity is maximized when there is an equal number of individuals in each group.[6] Thus, if Conditions 1, 2, 3 and 4 hold, our index $PD(s,T)$ can be interpreted as an index of *peripheral diversity.*[7]

[6] For example, Shannon's information entropy index satisfies this property (Shannon, 1949).

[7] Notice that Condition 3 requires concavity of the function $f_0(s_0, \ldots, \tau)$. Thus, such concavity, together with the other conditions, is sufficient to obtain that the solution to the maximization problem stated in the proposition is given by $s_k^* \geq s_l^*$. However, concavity of $f_0(s_0, \ldots, \tau)$ is not a necessary condition to obtain the solution. For example, if the function $f(s_0, \ldots, \tau)$ is "sufficiently" concave, the function $f_0(s_0, \ldots, \tau)$ need not be concave.

The proposition clarifies the relationship between diversity and the nature of the inter-group alienation function. Thus, whenever the functions f and f_0 are concave on the size of the group, the index PD can be seen as satisfying a necessary condition to be interpreted as a *peripheral diversity* index.

At this point, one might ask what would happen if instead of Condition 3, we impose the "opposite" condition by taking the functions $f(s_0, \ldots, \tau)$ and $f_0(s_0, \ldots, \tau)$ to be *convex*. This would imply that if groups k and l have the same linguistic distance to the center, the total peripheral alienation increases if members of a smaller group join the larger one. This alternative property could be seen as a necessary condition to obtain an index of *peripheral polarization* instead of an index of peripheral diversity. Thus, depending on whether one believes that inter-group alienation increases in the size of the group in a concave way or in a convex way, the aggregate index PD can be interpreted as satisfying a necessary property of either a measure of diversity or a measure of polarization.

3 A specific index of peripheral alienation

In this section we provide a specific form for the inter-group alienation functions f and f_0. These functions will be the ones used in the empirical part. In contrast to our approach in the previous section, we deduce them from assumptions at the individual level.

To come up with such functions, we follow the identification-alienation framework of Esteban and Ray (1994), though we will allow for a more flexible approach. An individual who speaks language k feels identified with other individuals who speak the same language. This sense of identification is a function of the size of the group, and is represented by s_k^{α}. In Esteban and Ray (1994) α is positive, implying that the sense of identification is stronger the bigger the group. In contrast, we prefer not to restrict the value of α. Indeed, it may very well be that the sense of identification becomes smaller as the group becomes larger, in which case $\alpha < 0$. There are many examples in the world of small linguistic, cultural or religious groups feeling a keener sense of community and a stronger desire to assert their identity.

An agent speaking language k feels more alienated from someone speaking language j the greater the distance τ_{kj}. This alienation (or antagonism) is influenced by the sense of identification. In particular, an individual attaches more weight to the distance τ_{kj} if his sense of identification is stronger. As defined in Esteban and Ray (1994), the alienation between an individual speaking language k and an individual speaking language j is $s_k^\alpha \tau_{kj}$. Since there is a proportion s_0 of individuals speaking the dominant language, the centrifugal alienation of an agent speaking minority language k is $s_0^\beta s_k^\alpha \tau_{0k}$. In Esteban and Ray (1994), $\beta = 1$. In our case, we suppose that an individual's centrifugal alienation only depends on there being an official or dominant language, independently of how many people actually speak that dominant language, so that we set $\beta = 0$. In that case, an individual's centrifugal alienation is $s_k^\alpha \tau_{0k}$. However, the qualitative empirical findings do not change under the alternative specification of $\beta = 1$.

If each individual speaking minority language k feels an alienation $s_k^\alpha \tau_{0k}$, and if a share s_k of the population speaks language k, then the centrifugal alienation of all speakers of language k is $s_k^{1+\alpha} \tau_{0k}$. Thus, the inter-group alienation function f is given by

$$f(s_0, s_k, \tau_{0k}) \equiv s_k^{1+\alpha} \tau_{0k} \tag{2}$$

We assume that individuals of the center have the same type of alienation function as individuals of the minority groups, except for the fact that in this case β is set to 1. Indeed, the centripetal alienation felt by members of the central group should depend on the size of the minorities, so that

$$f_0(s_0, s_k, \tau_{0k}) \equiv s_k s_0^{1+\alpha} \tau_{0k} \tag{3}$$

We can now define the total level of *peripheral alienation* by plugging (2) and (3) into (1):

$$\begin{aligned} PD(s,T) &= \sum_{k=1}^{K} (f(s_0, s_k, \tau_{0k}) + f_0(s_0, s_k, \tau_{0k})) \qquad (4) \\ &= \sum_{k=1}^{K} (s_k^{1+\alpha} \tau_{0k} + s_k s_0^{1+\alpha} \tau_{0k}) \end{aligned}$$

This is the index we will be using in the empirical section of the paper.

Depending on the value of α, (4) can be interpreted as an index of *peripheral diversity* or an index of *peripheral polarization.* In particular, if $\alpha < 0$, this index satisfies Conditions 1, 2, 3 and 4, and can thus be viewed as an index of peripheral diversity. If, in contrast, $\alpha > 0$, the function f would be convex, and (4) could be interpreted as an index of peripheral polarization.[8]

To illustrate the difference between diversity and polarization, consider a country with three linguistic groups. Their respective sizes are s_0, s_1 and s_2. Language 0 is the dominant language, and languages 1 and 2 are the minority languages. Further assume that the distance between each minority language and the dominant language is 1. Index (4) is then equal to $s_1 s_0^{1+\alpha} + s_2 s_0^{1+\alpha} + s_1^{1+\alpha} + s_2^{1+\alpha}$. We can now interpret this example for the two cases we have in mind. If $\alpha < 0$ we get a measure reflecting diversity (or fractionalization). For a given share of the dominant language, the maximum diversity is reached when $s_1 = s_2$. In other words, we face most diversity with two (equally sized) minority languages. If $\alpha > 0$, we obtain a measure reflecting polarization, that attains the highest level if one of the two remaining languages disappears. In other words, the level of polarization is highest if we have only one, rather than two, minority languages. This insight does not change once we allow for different distances between languages.

4 Data and measurement issues

Our data set covers most countries in Europe and the Americas. The information on how many people speak a given language in a given country come from two different data sources, the *Ethnologue* project (Grimes and Grimes, 1996)[9] and the *Britannica Book of the Year 1990.* Of the two, the Ethnologue gives much more disaggregated information, and would therefore seem like a superior option. However, some previous studies, such as Alesina et al. (2003), have preferred to use the Britannica, exactly because the data are less detailed. Without taking into account distances between languages, more detailed

[8] Note that the polarization index in Esteban and Ray (1994) takes into account alienation between all individuals, so that in contrast to our index (4), their index has a double summation.

[9] This data is available at http://www.ethnologue.com.

data are not necessarily better.

This brings us to the question of what defines a group. Should we consider speakers of Dutch and Flemish to belong to different groups in the same way as speakers of Greek and Turkish? In the first example, Flemish is a variation (or a dialect) of Dutch. In the second example, Greek and Turkish pertain to entirely different families, Greek to the Indo-European one and Turkish to the Altaic one. If we do not account for distances between languages, it would make sense to consider speakers of Dutch and Flemish as members of the same group. This is what makes the less detailed *Britannica* more appropriate than the more detailed *Ethnologue* in the analysis of Alesina et al. (2003). However, once we correct for linguistic distances, using more disaggregate data may be preferable. To go back to our previous example, our data on linguistic distances sets the distance between Dutch and Flemish at 0.046, whereas the distance between Dutch and Turkish is 1. Once we take this information into account, maintaining the disaggregation between Dutch and Flemish may be desirable.

While we believe that including linguistic distances is a step in the right direction, the issue of the appropriate level of disaggregation does not disappear altogether. Even after controlling for linguistic distances, the number of groups continues to be important in our diversity index, and in that sense the level of disaggregation still matters. This issue is especially relevant in the case of the indigenous languages of the Americas. For example, the number of languages reported in the *Ethnologue* for Mexico is 295, including 52 types of Mixteco and 27 types of Nahuatl. Clearly, the level of diversity picked up by our index will depend on whether we aggregate the 52 types of Mixteco into a single group or not. We had to make a value judgment in this case, and decided in favor of aggregation. We took this approach for all indigenous languages in the Americas. To aggregate appropriately, we used the information on language families provided by the *Ethnologue*. Doing this in the case of Mexico grouped the indigenous languages into eight language families.

Although we believe we have made substantial progress in solving the problem of aggregation by including linguistic distances, we are aware that our solution is not perfect.

As a robustness check, we therefore compute separate diversity indices, one based on the more detailed *Ethnologue*, and another based on the less detailed *Britannica*. The data for the indices we computed are reported in *Table 1* and *Table 2*.

The distances we use in our diversity index (4) come from Dyen et al. (1992), who conducted an extensive lexicostatistical study of 95 Indo-European languages. They focused on 200 basic meanings, and computed for each pair of languages the proportion of cognates.[10] Dyen et al. provide a 95 by 95 matrix with elements ranging from 0 (if the two languages have no cognates) to 1 (if the two languages are completely identical). For our purposes we have defined the distance τ_{jk} between any two languages j and k to be:

$$\tau_{jk} = 1 - \text{percentage cognate}$$

where the "percentage cognate" is taken from Dyen et al. (1992). To give a couple of examples, the distance between Dutch and Flemish is 0.046; the distance between Spanish and Catalan is 0.270; the distance between English and Danish is 0.407; the distance between German and Spanish is 0.747; the distance between Hindi and English is 0.854; the distance between Albanian and Afghan is 0.930.

Given that the information in Dyen et al. (1992) does not cover all our needs, we have had to make several choices. First, when no information was available on the distance between two varieties of the same language, we set that distance to be 0.05, similarly to the distance between Dutch and Flemish. We applied this to the distance between, for instance, Asturian and Spanish, and Ligurian and Italian. Second, we set the distance between any pair of a non Indo-European language and an Indo-European language to the maximum level of 1. This is the case of the distance between, for instance, Finnish and Swedish, Hungarian and Romanian, or Mayan and Spanish. Given our focus on peripheral diversity, we did not have to compute distances between different non Indo-European languages.[11]

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

[11] In nearly all countries in our data set the dominant language is Indo-European. The only exception is Uruguay, where the dominant language is Tupi. However, since the minority languages are Indo-European (Spanish and Portuguese), we set the distance between Tupi and those languages equal to 1.

We are not the first to introduce distances into indices of diversity. Fearon (2003) and Laitin (2000), based on the work of Greenberg (1956), have proposed a generalization of the fractionalization index by including linguistic distance. They use language trees from the *Ethnologue* project and base the resemblance between two languages on the number of branches they have in common. We prefer to use the Dyen et al. (1992) matrix, mainly because the information provided is finer. If, instead, we had used language trees from the *Ethnologue*, the number of branches in common would have ranged from 0 (if the two languages pertained to different families, such as Turkish and Greek) to 9 (in the case of Franco-Provençal and French). This would have given us 10 possible distances between languages. In contrast, the percentage cognate approach in Dyen et al. (1992) gives a distance between languages which could be anywhere between 0 and 1. Note that in some cases the difference between the two measures is actually quite significant. For instance, in the index proposed by Fearon (2003) the resemblance between Dutch and Flemish is 0.71, between Spanish and Catalan is 0.84, and between Spanish and Portuguese is 0.89.[12] In contrast, Dyen et al. (1992) finds a percentage cognate between Dutch and Flemish of 0.954, between Spanish and Catalan of 0.73, and between Spanish and Portuguese of 0.708. One reason for this difference is that Fearon (2003) uses the number of branches two languages have in common, without taking into account the number of branches that separates them. Using the example above will clarify this point. According to the *Ethnologue*, Dutch and Flemish have 5 branches in common, but only 1 branch that separates them. In contrast, Spanish and Catalan have 7 branches in common, yet 2 branches that separate them. Likewise, Spanish and Portuguese share 8 branches, but have 2 branches that separate them. In this case, it is unclear that Dutch and Flemish should be classified as being less similar than Spanish and Catalan or Spanish and Portuguese. We do not want to claim that our way of measuring distances is necessarily superior to that used by Fearon (2003). It is different, and in some sense is complementary.

In our regressions redistribution is measured by government transfers and subsi-

[12] To compute these distances, we used the formula provided in footnote (23) in Fearon (2003).

dies as a percentage of GDP. The data come from the *Economic Freedom Data Network* and we have taken the average for the years 1985-1995. Although our main focus is on peripheral diversity, we also 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, Latin American dummy, and a "small island" dummy. Most of these control variables have been taken directly from La Porta et al. (1999), while others come from the World Bank. A detailed description of the data is given in Appendix B.

In some cases the exogeneity of the regressors can be put to doubt. For instance, higher GDP per capita may increase the demand for redistribution, but higher redistribution may also affect GDP per capita. However, as argued by La Porta et al. (1999), the explanatory power of GDP per capita is important, so that leaving it out could be problematic. It turns out that our main findings are robust to the inclusion of GDP per capita. The statistical significance and the quantitative importance of our main variable of interest, peripheral diversity, does not change much whether we include GDP per capita or not. As far as population is concerned, a similar endogeneity problem may arise. Not only do people move across borders in function of redistribution, but country borders themselves may be endogenous (Alesina and La Ferrara, 2004). This makes the "small island" dummy attractive. Our preliminary data analysis showed small islands tended to be outliers by exhibiting less redistribution than the model predicted. In as far as the size of small islands is determined by geography, it may be a good exogenous measure of country size. The problem of endogeneity may also afflict the focus of our analysis — the relation between diversity and redistribution — although it is not immediately clear in what way. On the one hand, higher redistribution may attract more immigrants. On the other hand, higher redistribution may lead to more restrictive migration policies (Hanson, Scheve and Slaughter, 2005). The existing literature has largely ignored the endogeneity of diversity. In this paper, we instrument diversity with past diversity. In particular, we use language data of Muller (1964) to compute a diversity measure for the beginning of the 1960s.

5 Empirical analysis

5.1 Linguistic distances

We start by showing how including linguistic distances affects our measure of peripheral diversity (4), $\sum_{k=1}^{K}(s_k^{1+\alpha}\tau_{0k} + s_k s_0^{1+\alpha}\tau_{0k})$, where $-1 < \alpha < 0$. In our empirical work we take $\alpha = -1/2$.[13] Using the *Ethnologue* data, the first column of **Table 3** gives the ranking of countries according to linguistic diversity when we include distances between languages. The second column gives the same ranking but not allowing for different distances between languages. In this case the distance τ_{jk} between any two languages j and k is set equal to 1 if $j \neq k$ and equal to 0 if $j = k$. The third column tells us how many positions a country gains or loses when we allow for different linguistic distances.

When introducing distances between languages, the league is headed by Canada, a country with a large French speaking minority. Other countries high up include the United States, some of the Latin American countries with important indigenous languages (Bolivia, Peru, Paraguay, Guatemala), and certain European countries (Moldova, Switzerland, Belgium).

When all languages are taken to be equidistant from each other, Canada is still on top. However, there are quite a few significant changes. The following examples will help to understand the relevance of linguistic distances in our sample of 74 countries:

- Bulgaria moves up 17 positions in the ranking when we allow for different distances between languages. This is because Bulgaria has a sizeable Turkish minority, and because the distance between Turkish and Bulgarian is large.

- Something similar happens to Estonia, that moves up 6 positions, and Finland, that moves up 9 positions. In both of these countries the dominant groups speaks a non Indo-European language, but each has Indo-European speaking minorities: Russian in the case of Estonia and Swedish in the case of Finland.

- Other countries that become more diverse once we introduce different distances

[13] We experimented with different values for $\alpha < 0$. This did not change the results.

are some of the Latin American countries with important indigenous populations: Paraguay and Peru move up 8 positions, Mexico moves up 14 position, and Ecuador moves up 13 positions.

- Spain goes the other way, and drops 16 positions in the ranking, once we allow for different distances between languages. The two large minorities (Galician and Catalan) speak languages very similar to Spanish. Andorra is another interesting case, with 53% of the population Catalan speaking and 42% Spanish speaking, that drops 22 positions when we account for linguistic distances.

- Other countries that move down in the ranking are Ukraine (drops 21 positions) and Belarus (drops 20 positions). The differences between Ukrainian and Russian, and between Byelorussian and Russian, are small. Germany and Italy also lose positions, because the most important minorities are dialects of German and Italian.

5.2 Peripheral diversity and redistribution

Table 4 reports the results of our regressions of redistribution on peripheral diversity, and a number of control variables. Robust t-ratios are given in brackets. As explained before, the theoretical prior is that the greater the degree of peripheral diversity, the lower the degree of redistribution. The data on languages are taken from the *Ethnologue*.

Column (1) reports the most basic specification, only including peripheral diversity and a number of exogenous control variables (latitude, Latin American dummy, and small island dummy). All coefficients are highly significant. In particular, the coefficient on peripheral diversity is statistically significant at the 1% level. Following La Porta et al. (1999) and Alesina et al. (2003), column (2) and (3) include legal origin and religious composition as further controls. None of these variables are significant though. Column (4) takes the basic specification, and adds GDP per capita. As expected, the level of a country's development increases the degree of redistribution. La Porta et al. (1999) have argued that GDP per capita may be endogenous. However, including GDP per capita does not change the coefficient and the statistical significance of peripheral diversity. This suggests that the possible endogeneity of GDP per capita does not affect our variable of interest. This robustness is confirmed by column (5), which takes the previous specification with GDP per capita, and adds legal origin and religious composition. In this case, socialist legal origin becomes statistically significant and gets the expected positive sign. Column (6) throws population in our basic specification. The coefficient is not statistically significant though. Finally, column (7) consists of the full specification, including all control variables. Again, population size does not seem to matter.

Based on Table 4, we can see that in all specifications the effect of peripheral diversity is robust, both in terms of magnitude and statistical significance. In all specifications peripheral diversity is significant at the 1% level, and the magnitude of its coefficient hovers between -3.34 and -4.09. Taking column (5) as our preferred specification, the model predicts that an increase in peripheral diversity by one standard deviation lowers redistribution by 1.6 percentage points. This effect should be compared to an average level of redistribution of 14% of GDP. A couple of other examples may help to further illustrate the quantitative importance of diversity. Compare the United States, with a level of redistribution of 12.3%, and Italy, with redistribution of 24.5%. The model predicts that the greater linguistic diversity in the U.S. lowers redistribution by 4.5 percentage points compared to Italy. This implies that about one third of the difference in redistribution between the U.S. and Italy can be explained by the difference in linguistic diversity. Another example: about two thirds of the gap between Switzerland, with redistribution of 14.6%, and Denmark, with redistribution of 21.3%, can be attributed to Switzerland's greater linguistic diversity.

As for the control variables, neither the legal origin nor the religious composition seems to have a significant impact on redistribution. An exception to this is socialist legal origin, 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 11 percentage points. Latitude is significant and positive in all specifications. This is a common result in the literature. The Latin American dummy seems to be partly picking up other variables. Once we introduce GDP per capita and control for religious composition and legal origin, it loses all significance.

The small island dummy[14] is highly significant and its coefficient is robust to all specifications. Being a small island is predicted to reduce redistribution by 7 to 10 percentage points of GDP. As argued before, the small island dummy may partly be picking up population size. This may explain why population shows up as being statistically insignificant. Indeed, both variables measure the size of the country. However, of the two, the small island dummy is the more exogenous one, and therefore preferable. It is well known that country size is endogenous: borders change and people may migrate because of redistribution. The endogeneity of the border is largely absent in the case of small islands though.

As mentioned in the previous section, diversity may also be endogenous, although it is unclear what bias this potential endogeneity would introduce. More redistribution may attract more immigrants, therefore creating more diversity. However, more redistribution may also imply tighter restrictions on immigration, leading to less diversity. To correct for a possible endogeneity bias, we use an instrumental variable approach. As instrument we compute the level of diversity at the beginning of the 1960s. The rationale is the following. While diversity in the 1960s clearly affects diversity in the 1990s, arguably enough time has passed for it not to influence redistribution in the 1990s. Using even earlier data on diversity would have been desirable. Unfortunately, the earliest quantitative information on the number of speakers of different languages in a large sample of countries seems to have been compiled by Muller (1964). These are the data we use to compute our instrumental variable. The results are reported in **Table 5**. As can be seen, instrumenting does not change the picture. The effect of peripheral diversity on redistribution is unchanged, both in terms of magnitude and statistical significance. To see whether peripheral diversity is indeed endogenous, we perform a standard Hausman test. We take the predicted value of the coefficient on peripheral diversity from the first stage regression, and introduce it as an additional regressor in the original specification. If the coefficient on the predicted value is significantly different from zero, we conclude

[14] The small island dummy is defined as an island with a population of less than 0.5 million. See Appendix B.

that diversity is endogenous. In none of the specifications is the coefficient statistically significant at the 10% level, and we therefore conclude that endogeneity of diversity does not seem to be a serious concern.

Table 6 runs the same regressions as Table 4, but uses a peripheral diversity index which does not allow for different distances between languages. We will focus on our variable of interest — peripheral diversity — as the coefficients on the control variables are similar to what we found in Table 4. The most obvious result is that the peripheral diversity index loses statistical significance. In none of the regressions is the index significant at the 10% level. Not surprisingly, all specifications also give lower R^2s compared to Table 4. 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.

As a further robustness check, we compute our peripheral diversity index using the *Britannica* data. The results reported in **Table 7** confirms our previous findings. Peripheral diversity is statistically significant at the 1% level across all specifications, and its point estimates are fairly stable. In **Table 8** we run the same regressions, but now using the diversity index which does not account for linguistic distances. As before, we find that not including linguistic distances leads to worse results. Both the statistical significance and the explanatory power of the regressions drop across all specifications. However, note that the difference between including distances and not including distances is smaller when using the *Britannica* data than when using the *Ethnologue* data. This is to be expected. As argued before, failing to take into account distances is not as bad when one uses more aggregate data. For instance, the problem of the different Italian or German dialects does not appear in the *Britannica* data set, while it does in the *Ethnologue*.

5.3 Peripheral diversity and peripheral polarization

Some authors have argued that the degree of social conflict has more to do with polarization than with diversity (Esteban and Ray, 1994). As already mentioned in the theoretical section, if $\alpha > 0$, our index (4) can be interpreted as one of peripheral polarization, rather than peripheral diversity. We now explore whether diversity or polarization does a better job at explaining redistribution. To compute our indices of peripheral polarization, we set $\alpha = 1$ in (4).[15]

Table 9 compares the country ranking of polarization and diversity. The correlation between the two measures is 0.79; the rank correlation is even higher and stands at 0.93. In spite of that, there are some differences worth pointing out. Countries with large immigrant populations tend to be diverse but not polarized. The United States, for instance, drops 17 positions when comparing polarization to diversity. European countries with a substantial number of immigrants, such as the United Kingdom, Germany and France, exhibit a similar pattern. Countries that move the other way, and are more polarized than diverse, include Andorra and Trinidad. As already mentioned, Andorra is split up in half between Catalan and Spanish speakers. In Trinidad, for its part, 70% of the population speaks English Creole and the remaining 30% speaks Hindustani.

Table 10 reports the same regressions as before, focusing on peripheral polarization, rather than peripheral diversity. Given that the effect of the control variables is similar, we will exclusively focus on the polarization index. Using column (5) as our preferred specification, an increase in polarization by one standard deviation lowers redistribution by 1.2 percentage points. To further illustrate the quantitative effect of polarization, some other examples may be useful. Guatemala (polarization index 0.282) is predicted to have a level of redistribution which is 3.9 percentage points lower than Costa Rica (polarization index 0.016). Another example: redistribution in Belgium (polarization index 0.184) is predicted to be 2.7 percentage points lower than in Denmark (polarization index 0.003).

We now try to examine whether polarization or diversity does a better job at explaining redistribution. In terms of statistical significance, polarization performs slightly worse. The polarization index is statistically significant at the 5% level in 5 out of the 7

[15] As in the case of diversity, we experimented with different values of $\alpha > 0$. This did not change the results.

specifications, whereas the diversity index is statistically significant at the 1% level in all regressions. The R^2s give a similar picture, with polarization performing somewhat worse than diversity. The fact that polarization and diversity give similar results is not surprising. Many highly diverse societies also tend to be highly polarized. The high correlation between both indices indicate a substantial overlap between both concepts. In fact, the polarization index may be picking up the effect of diversity, or vice versa. One possible way of testing which of the two concepts is more powerful in explaining redistribution is to run regressions that include both the diversity index and the polarization index as explanatory variables. These results are reported in **Table 11**. While we have to be cautious because of possible multicollinearity between both indices, the results we get are quite compelling. The polarization index loses all significance, and its point estimate varies widely depending on the specification. In contrast, the diversity index still exhibits a high degree of significance. In 6 out of the 7 specifications its coefficient is statistically significant at the 10% level. Moreover, its point estimate is relatively stable, and not much changed compared to what we found in Table 4. All of this suggests diversity taking the upper hand when explaining redistribution.

Regarding the importance of distance, our findings in the case of polarization reinforce our previous results. If one does not allow for different distances between languages, the polarization index loses all significance. Those results are reported in **Table 12**. When analyzing social conflict, whether the focus is on diversity or polarization, it seems crucial to consider the distances between groups.

6 Concluding remarks

This paper has studied the effect of linguistic conflict on redistribution. In particular, we have analyzed the conflict arising between peripheral minority groups and a dominant center. Our main contribution is to explicitly introduce linguistic distances into our measure of linguistic conflict. The empirical part of the paper shows that this improves results substantially. Although our focus is on diversity, the index we propose encompasses both diversity and polarization. The advantage of such an index is that we can let the data tell us which notion of linguistic conflict, diversity or polarization, is more appropriate when explaining redistribution. The index we propose could of course be applied to study the effect of diversity (or polarization) on other economic variables, such as economic growth, the quality of the government, or the degree of decentralization.

This paper leaves a number of important questions unaddressed. First, it should be interesting to further explore asymmetries in ethnolinguistic conflict. While the existing literature has tended to treat all individuals symmetrically when measuring ethnolinguistic diversity, we have focused on the conflict originating between the center and the periphery. There are clearly other types of possible asymmetries. This question merits further research. Second, the issue of the right level of aggregation, and what defines a group, is not fully resolved. We believe that introducing distances between groups is a major step in the right direction, but even so, it is still unclear what level of disaggregation is desirable.

References

[1] Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and R. Wacziarg, 2003. "Fractionalization", *Journal of Economic Growth*, 8(2), 155-194.

[2] Alesina, A., Glaeser, E., and B. Sacerdote, 2001. "Why Doesn't the US have a European Style Welfare System?", *NBER Working Paper* #8524.

[3] Alesina, A. and E. La Ferrara, 2004. "Ethnic Diversity and Economic Performance", *NBER Working Paper* #10313.

[4] Becker, G., 1957. *The Economics of Discrimination*, Chicago: University of Chicago Press.

[5] Caselli, F. and W.J. Coleman II, 2002. "On the Theory of Ethnic Conflict", manuscript,
(January 2005) http://post.economics.harvard.edu/faculty/caselli/caselli.html.

[6] Dyen, I., Kruskal, J.B., and P. Black, 1992. "An Indo-European Classification: A Lexicostatistical Experiment," *Transactions of the American Philosophical Society*, 82(5), Philadelphia: American Philosophical Society.

[7] Easterly, W. and R. Levine, 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions", *Quarterly Journal of Economics*, 112(4), 1203-1250.

[8] Esteban, J.-M., and D. Ray, 1994. "On the Measurement of Polarization", *Econometrica*, 62(4), 819-851.

[9] Fearon, J.D., 2003. "Ethnic and Cultural Diversity by Country", *Journal of Economic Growth*, 8, 195-222.

[10] Fernandez, R. and G. Levy , 2004. "Class and Tastes: The Effects of Income and Preference Heterogeneity on Redistribution", mimeo.

[11] Ginsburgh, V., Ortuño-Ortín, I., and S. Weber, 2005. "Disenfranchisement in Linguistically Diverse Societies: The Case of the European Union," forthcoming in *Journal of the European Economic Association.*

[12] Hanson, G.H., Scheve, K. and M. Slaughter, 2005. "Local Public Finance and Individual Preferences over Globalization Strategies", *NBER Working Paper* #11028.

[13] Greenberg, J.H., 1956. "The Measurement of Linguistic Diversity", *Language* 32:109-115

[14] Grimes, J. and B. Grimes, 1996. *Ethnologue: Languages of the World*, 13th edition Dallas, TX: Summer Institute of Linguistics.

[15] La Porta, R., Lopez de Silanes, F., Shleifer, A., and R. Vishny, 1999. "The Quality of Government", *Journal of Law, Economics and Organization*, 15(1), 222-279.

[16] Laitin, D.D., 2000. "What Is a Language Community?", *American Journal of Political Science*, 44, 1, 142-155.

[17] Lee, W. and J.E. Roemer, 2004, "Racism and Redistribution in the United States: A Solution to the Problem of American Exceptionalism", mimeo.

[18] Luttmer, E., 2001. "Group Loyalty and the Taste for Redistribution", *Journal of Political Economy* 109, 500-528.

[19] Montalvo, J., and M. Reynal-Querol, 2002. "Religious Polarization and Economic Development", mimeo.

[20] Muller, S.H., 1964. *The World's Living Languages: Basic Facts of Their Structure, Kinship, Location and Number of Speakers*, New York: Frederick Ungar Publishing Co.

[21] Nehring, K and C. Puppe, 2002. "A Theory of Diversity", *Econometrica* 70, 1155-1198.

[22] Shannon, C.E., 1949. *The Mathematical Theory of Communication*. University of Illinois Press.

[23] Vigdor, J.L., 2004. "Community Heterogeneity and Collective Action: Initial Mail Response to the 2000 Census." *Review of Economics and Statistics* 86: 303-12.

[24] Weitzman, M.L., 1992. "On Diversity", *The Quarterly Journal of Economics*, 107(2), 363-405.

A Appendix A: Proof of Proposition 1

1) First consider the case $\tau_{0k} = \tau_{0l} = \tau$. We have to show that $s_k^* \geq s_l^*$. Suppose, to the contrary, that $s_k^* < s_l^*$. Let $s' \in \mathcal{S}_{kl}(\overline{s})$ be such that $s'_j = s_j^*$ for all $j \neq k$, $j \neq l$ and $s'_k = s'_l = x \equiv \frac{s_k^*+s_l^*}{2}$. Since $PD(s', T) < PD(s^*, T)$, it follows that

$$2f(s_0, x, \tau) + 2f_0(s_0, x, \tau) < f(s_0, s_k^*, \tau) + f(s_0, s_l^*, \tau) + f_0(s_0, s_k^*, \tau) + f_0(s_0, s_l^*, \tau). \quad (5)$$

By Condition 3, functions $f(s_0, \ldots, \tau)$ and $f_0(s_0, \ldots, \tau)$ are concave, which implies

$$f(s_0, x, \tau) \geq \frac{1}{2} f(s_0, s_k^*, \tau) + \frac{1}{2} f(s_0, s_l^*, \tau) \quad (6)$$

and

$$f_0(s_0, x, \tau) \geq \frac{1}{2} f_0(s_0, s_k^*, \tau) + \frac{1}{2} f_0(s_0, s_l^*, \tau). \quad (7)$$

It is straightforward to verify that inequalities (5), (6), (7) and can not hold simultaneously. Thus, we have that $s_k^* \geq s_l^*$. Notice that $\tau_{0l} = \tau_{0k}$ implies $s_l^* \geq s_k^*$ and $s_k^* \geq s_l^*$ so that $s_k^* = s_l^*$.

2) Consider the case $\tau_{0k} > \tau_{0l}$. We shall show that $s_k^* \geq s_l^*$. Suppose, in negation, that $s_k^* < s_l^*$. Let $T' \in \mathcal{T}$, be such that $\tau'_{0j} = \tau_{0j}$ for all $j \neq l$ and $\tau'_{0l} = \tau_{0k}$. Notice that $\tau'_{0j} > \tau_{0l}$. Similarly to the previous examination, let $s' \in \mathcal{S}_{kl}(\overline{s})$ be such that $s'_j = s_j^*$ for all $j \neq k$, $j \neq l$ and $s'_k = s'_l = x \equiv \frac{s_k^*+s_l^*}{2}$. We have

$$PD(s', T) < PD(s^*, T). \quad (8)$$

This implies that

$$\begin{aligned} & f(s_0, x, \tau_k) + f(s_0, x, \tau_l) + f_0(s_0, x, \tau_k) + f_0(s_0, x, \tau_l) \quad (9) \\ < \; & f(s_0, s_k^*, \tau_k) + f(s_0, s_l^*, \tau_l) + f_0(s_0, s_k^*, \tau_k) + f_0(s_0, s_l^*, \tau_l), \end{aligned}$$

which is equivalent to

$$\begin{aligned} & f(s_0, x, \tau_k) - f(s_0, s_k^*, \tau_k) + f_0(s_0, x, \tau_k) - f_0(s_0, s_k^*, \tau_k) \quad (10) \\ < \; & f(s_0, s_l^*, \tau_l) - f(s_0, x, \tau_l) + f_0(s_0, s_l^*, \tau_l) - f_0(s_0, x, \tau_l). \end{aligned}$$

The argument used in case 1 above yields

$$PD(s', T') > PD(s^*, T'),$$

which implies that

$$\begin{aligned} & f(s_0, x, \tau_k) + f(s_0, x, \tau_k) + f_0(s_0, x, \tau_k) + f_0(s_0, x, \tau_k) \\ > \quad & f(s_0, s_k^*, \tau_k) + f(s_0, s_l^*, \tau_k) + f_0(s_0, s_k^*, \tau_k) + f_0(s_0, s_l^*, \tau_k). \end{aligned} \tag{11}$$

By rearranging the terms we obtain

$$\begin{aligned} & f(s_0, x, \tau_k) - f(s_0, s_k^*, \tau_k) + f_0(s_0, x, \tau_k) - f_0(s_0, s_k^*, \tau_k) \\ > \quad & f(s_0, s_l^*, \tau_k) - f(s_0, x, \tau_k) + f_0(s_0, s_l^*, \tau_k) - f_0(s_0, x, \tau_k). \end{aligned} \tag{12}$$

Inequalities (10) and (12) imply

$$\begin{aligned} & f(s_0, s_l^*, \tau_k) - f(s_0, x, \tau_k) + f_0(s_0, s_l^*, \tau_k) - f_0(s_0, x, \tau_k) \\ < \quad & f(s_0, s_l^*, \tau_l) - f(s_0, x, \tau_l) + f_0(s_0, s_l^*, \tau_l) - f_0(s_0, x, \tau_l). \end{aligned} \tag{13}$$

Since $s_l^* > x$ and $\tau_k > \tau_l$, Condition 4 implies that

$$\begin{aligned} f(s_0, s_l^*, \tau_k) - f(s_0, x, \tau_k) \quad &> \quad f(s_0, s_l^*, \tau_l) - f(s_0, x, \tau_l) \text{ and} \\ f_0(s_0, s_l^*, \tau_k) - f_0(s_0, x, \tau_k) \quad &> \quad f_0(s_0, s_l^*, \tau_l) - f_0(s_0, x, \tau_l), \end{aligned} \tag{14}$$

and (14) and (12) do not not hold simultaneously. Hence we conclude that $s_k^* \geq s_l^*$.

B Appendix B: Description of the data

- Transfers and subsidies as % of GDP: Average for 1985, 1990, and 1995. *Source: Economic Freedom Data Network.*
- GDP 85-95: Log GDP per capita (in constant 1995 dollars), average for the years between 1985 and 1995. *Source: World Bank.*
- Population 85-95: 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 commerical code; (4) Scandinavian commercial code; and (5) Socialist/communist laws. *Source: La Porta et al. (1999).*
- Catholic 80: the percentage of the population that is catholic in 1980. *Source: La Porta et al. (1999).*
- Muslim 80: the percentage of the population that is muslim in 1980. *Source: La Porta et al. (1999).*
- Protestant 80: 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).*
- Latin American dummy: dummy for all countries in Latin America and the Caribbean.
- Small island dummy: islands with a population of less than 0.5 million in 1990.

	Peripheral diversity	Peripheral diversity (no dist)	Peripheral polarization	Peripheral polarization (no dist)
Albania	0.474	0.538	0.047	0.053
Andorra	0.326	1.195	0.085	0.315
Argentina	0.719	0.944	0.063	0.092
Austria	0.220	0.517	0.016	0.053
Bahamas	0.025	0.493	0.006	0.119
Barbados	0.013	0.265	0.002	0.046
Belarus	0.271	0.950	0.033	0.131
Belgium	1.224	1.907	0.184	0.272
Belize	1.639	1.912	0.180	0.219
Bolivia	1.684	1.807	0.260	0.262
Bosnia-Herzegovina	0.000	0.000	0.000	0.000
Brazil	0.196	0.276	0.011	0.016
Bulgaria	0.983	0.787	0.364	0.103
Canada	1.814	2.284	0.159	0.207
Chile	0.234	0.247	0.029	0.030
Colombia	0.375	0.385	0.013	0.014
Costa Rica	0.202	0.236	0.016	0.020
Croatia	0.222	0.294	0.025	0.033
Cuba	0.000	0.000	0.000	0.000
Czech Republic	0.214	0.316	0.022	0.030
Denmark	0.044	0.177	0.003	0.014
Dominican Republic	0.177	0.304	0.014	0.029
Ecuador	0.724	0.765	0.122	0.124
El Salvador	0.048	0.048	0.002	0.002
Estonia	1.182	1.241	0.249	0.251
Finland	0.335	0.369	0.057	0.058
France	0.771	1.289	0.068	0.105
Germany	0.558	1.317	0.034	0.135
Greece	0.572	0.784	0.043	0.064
Guatemala	0.947	0.947	0.282	0.282
Guyana	0.347	0.347	0.036	0.036
Haiti	0.000	0.000	0.000	0.000
Honduras	0.421	0.443	0.031	0.032
Hungary	0.656	0.656	0.066	0.066
Iceland	0.000	0.000	0.000	0.000
Ireland	0.319	0.388	0.068	0.083
Italy	0.210	2.170	0.011	0.169
Jamaica	0.229	0.254	0.016	0.017
Latvia	0.962	1.546	0.156	0.246
Liechtenstein	0.148	0.430	0.019	0.061
Lithuania	0.621	0.988	0.088	0.140
Luxembourg	1.056	1.360	0.197	0.216
Macedonia	1.082	1.210	0.163	0.182
Malta	0.097	0.097	0.008	0.008
Mexico	0.750	0.771	0.061	0.062
Moldova	1.350	1.639	0.152	0.188
Netherlands Antilles	0.244	0.321	0.047	0.062
Netherlands	0.509	0.866	0.041	0.085
Nicaragua	0.290	0.311	0.038	0.040
Norway	0.192	0.302	0.023	0.029
Panama	0.817	0.902	0.114	0.129
Paraguay	1.221	1.221	0.136	0.136
Peru	1.027	1.027	0.156	0.156
Poland	0.209	0.499	0.019	0.055
Portugal	0.055	0.055	0.003	0.003
Puerto Rico	0.134	0.176	0.017	0.023
Romania	0.744	0.805	0.104	0.108
Slovakia	0.476	0.678	0.095	0.112
Slovenia	0.190	0.439	0.028	0.079
Spain	0.408	1.173	0.049	0.179
St Kitts	0.000	0.000	0.000	0.000
St Vincent	0.000	0.000	0.000	0.000
Suriname	1.221	1.762	0.152	0.246
Sweden	0.507	1.175	0.033	0.154
Switzerland	1.336	1.740	0.145	0.191
Trinidad	0.686	0.803	0.204	0.239
Turks&Caicos	0.000	0.000	0.000	0.000
United Kingdom	0.275	0.374	0.017	0.022
Ukraine	0.476	1.331	0.050	0.204
Uruguay	0.129	0.390	0.012	0.040
United States	1.564	2.047	0.100	0.133
US Virgin Islands	0.025	0.501	0.006	0.122
Venezuela	0.411	0.411	0.031	0.031
Yugoslavia	0.567	0.814	0.101	0.136

Table 1: Peripheral diversity and peripheral polarization indices (Ethnologue)

	Peripheral diversity	Peripheral diversity (no dist)
Albania	0.171	0.195
Andorra	0.443	1.152
Argentina	0.149	0.268
Austria	0.087	0.109
Bahamas	0.419	0.523
Barbados	0.018	0.357
Belarus	0.261	1.028
Belgium	0.959	1.322
Belize	1.603	1.605
Bolivia	1.234	1.234
Bosnia-Herzegovina	0.055	1.038
Brazil	0.194	0.267
Bulgaria	0.935	0.845
Canada	1.557	1.999
Chile	0.331	0.331
Colombia	0.210	0.219
Costa Rica	0.218	0.256
Croatia	0.222	0.294
Cuba	0.000	0.000
Czech Republic	0.229	0.564
Denmark	0.155	0.308
Dominican Republic	0.058	0.161
Ecuador	0.331	0.331
El Salvador	0.000	0.000
Estonia	1.088	1.186
Finland	0.306	0.306
France	0.658	1.208
Germany	0.412	0.525
Greece	0.290	0.325
Guatemala	0.905	0.905
Guyana	0.740	0.832
Haiti	0.074	0.482
Honduras	0.316	0.330
Hungary	0.407	0.407
Iceland	0.000	0.000
Ireland	0.227	0.276
Italy	0.237	0.615
Jamaica	0.315	1.054
Latvia	0.950	1.526
Liechtenstein	0.100	0.000
Lithuania	0.588	0.934
Luxembourg	0.818	1.137
Macedonia	1.015	1.311
Malta	0.164	0.164
Mexico	0.673	0.673
Moldova	1.122	1.366
Netherlands Antilles	0.288	0.379
Netherlands	0.258	0.409
Nicaragua	0.329	0.355
Norway	0.056	0.205
Panama	0.756	0.876
Paraguay	0.627	0.627
Peru	0.947	0.947
Poland	0.036	0.160
Portugal	0.000	0.000
Puerto Rico	0.051	0.067
Romania	0.748	0.829
Slovakia	0.600	0.824
Slovenia	0.189	0.438
Spain	0.364	1.075
St Kitts	0.000	0.000
St Vincent	0.000	0.000
Suriname	0.966	0.936
Sweden	0.214	0.214
Switzerland	0.902	1.114
Trinidad	0.282	0.418
Turks&Caicos	0.583	0.626
United Kingdom	0.094	0.146
Ukraine	0.348	1.130
Uruguay	0.000	0.000
United States	1.071	1.359
US Virgin Islands	0.606	0.686
Venezuela	0.180	0.180
Yugoslavia	0.920	1.174

Table 2: Peripheral diversity indices (Britannica)

C Appendix C: Tables

	diversity (incl distance)	diversity (exl distance)	difference
Canada	1	1	0
Bolivia	2	6	4
Belize	3	4	1
United States	4	3	-1
Moldova	5	9	4
Switzerland	6	8	2
Belgium	7	5	-2
Paraguay	8	16	8
Suriname	9	7	-2
Estonia	10	15	5
Macedonia	11	17	6
Luxembourg	12	11	-1
Peru	13	21	8
Bulgaria	14	31	17
Latvia	15	10	-5
Guatemala	16	24	8
Panama	17	26	9
France	18	14	-4
Mexico	19	33	14
Romania	20	29	9
Ecuador	21	34	13
Argentina	22	25	3
Trinidad	23	30	7
Hungary	24	36	12
Lithuania	25	22	-3
Greece	26	32	6
Yugoslavia	27	28	1
Germany	28	13	-15
Netherlands	29	27	-2
Sweden	30	19	-11
Slovakia	31	35	4
Ukraine	32	12	-20
Albania	33	37	4
Honduras	34	42	8
Venezuela	35	45	10
Spain	36	20	-16
Colombia	37	48	11
Guyana	38	51	13
Finland	39	50	11
Andorra	40	18	-22
Ireland	41	47	6
Nicaragua	42	54	12
United Kingdom	43	49	6
Belarus	44	23	-21
Netherlands Antilles	45	52	7
Chile	46	61	15
Jamaica	47	60	13
Croatia	48	57	9
Austria	49	38	-11
Czech Republic	50	53	3
Italy	51	2	-49
Poland	52	40	-12
Costa Rica	53	62	9
Brazil	54	58	4
Norway	55	56	1
Slovenia	56	43	-13
Dominican Republic	57	55	-2
Liechtenstein	58	44	-14
Puerto Rico	59	64	5
Uruguay	60	46	-14
Malta	61	65	4
Portugal	62	66	4
El Salvador	63	67	4
Denmark	64	63	-1
US Virgin Islands	65	39	-26
Bahamas	66	41	-25
Barbados	67	59	-8
Bosnia-Herzegovina	68	68	0
Iceland	69	69	0
Turks&Caicos	70	70	0
St Vincent	71	71	0
St Kitts	72	72	0
Haiti	73	73	0
Cuba	74	74	0

Source: Ethnologue, own calculations

Table 3: Ranking diversity, with and without distance

Dependent variable: Transfers and subsidies as % of GDP (average 1985-1995)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peripheral diversity	-3.947***	-3.961***	-3.634***	-4.093***	-3.344***	-3.958***	-3.437***
	(3.29)	(2.88)	(2.65)	(3.29)	(2.68)	(3.25)	(2.68)
GDP 85-95				1.835***	4.678***		4.712***
				(2.80)	(4.12)		(4.19)
Population 85-95						-0.047	-0.305
						(0.08)	(0.93)
UK legal origin		-0.349	-0.705		0.607		0.526
		(0.16)	(0.29)		(0.31)		(0.26)
French legal origin		1.873	0.302		3.718		3.847
		(0.93)	(0.10)		(1.42)		(1.43)
Socialist legal origin		0.412	0.461		10.876***		10.672***
		(0.16)	(0.14)		(2.69)		(2.62)
Scandinavian legal origin		-2.007	1.057		4.313		3.984
		(0.66)	(0.20)		(0.93)		(0.82)
Catholic 80			0.004		0.002		0
			(0.10)		(0.08)		(0.01)
Muslim 80			-0.004		0.161		0.138
			(0.01)		(0.81)		(0.71)
Protestants 80			-0.061		-0.069		-0.069
			(0.77)		(1.08)		(1.06)
Latitude	14.623*	20.224**	24.841***	9.412	13.358**	14.453*	12.381*
	(1.83)	(2.24)	(2.75)	(1.20)	(1.97)	(1.82)	(1.83)
Latin American dummy	-9.76***	-8.764**	-7.237**	-8.857***	-1.973	-9.846***	-2.553
	(2.98)	(2.50)	(2.08)	(2.74)	(0.64)	(2.95)	(0.83)
Small island dummy	-9.036***	-8.004***	-7.179***	-10.685***	-8.779***	-9.197***	-9.884***
	(4.04)	(3.27)	(2.77)	(4.47)	(3.34)	(3.00)	(3.50)
Intercept	15.079***	11.614**	10.632	1.139	-31.608***	15.934	-26.15**
	(3.21)	(2.27)	(1.62)	(0.16)	(2.66)	(1.49)	(2.04)
Number of observations	56	56	55	56	55	56	55
Adjusted R2	0.7153	0.7255	0.7633	0.7467	0.8462	0.7153	0.8467

Absolute value of robust t statistics in parenthesis
*significant at 10% **significant at 5% ***significant at 1%

Table 4: Peripheral diversity and redistribution

Dependent variable: Transfers and subsidies as % of GDP (average 1985-1995)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peripheral diversity	-3.808***	-3.94***	-3.956***	-3.642***	-3.877***	-3.687***	-3.631**
	(2.84)	(2.84)	(2.85)	(2.58)	(2.64)	(2.65)	(2.46)
GDP 85-95				1.45**	4.401***		4.476***
				(2.17)	(3.51)		(3.60)
Population 85-95						-0.202	-0.484
						(0.32)	(1.16)
UK legal origin		-1.136	-0.72		1.107		1.154
		(0.46)	(0.29)		(0.51)		(0.52)
French legal origin		2.864	3.107		5.403*		5.574*
		(1.33)	(0.92)		(1.83)		(1.86)
Socialist legal origin		1.671	1.988		12.912***		13.294***
		(0.61)	(0.42)		(2.78)		(2.90)
Scandinavian legal origin		-2.584	-2.754		0.276		-0.055
		(0.84)	(0.46)		(0.05)		(0.01)
Catholic 80			0.02		0.018		0.015
			(0.62)		(0.68)		(0.57)
Muslim 80			0.147		0.217		0.148
			(0.42)		(0.90)		(0.62)
Protestants 80			0.022		0.003		0.003
			(0.25)		(0.04)		(0.04)
Latitude	18.622**	27.373***	28.264***	13.406*	16.464**	18.073**	14.846**
	(2.33)	(3.06)	(3.05)	(1.69)	(2.32)	(2.30)	(2.09)
Latin American dummy	-8.741***	-7.287**	-7.208*	-7.903**	-1.893	-9.05***	-2.633
	(2.65)	(2.05)	(1.89)	(2.43)	(0.55)	(2.77)	(0.80)
Small island dummy	-10.741**	-10.386***	-10.569***	-10.595**	-8.766**	-11.436**	-10.405**
	(2.46)	(2.74)	(2.64)	(2.37)	(2.00)	(2.35)	(2.37)
Intercept	13.128***	7.754	5.356	2.126	-33.283***	16.667	-25.135*
	(2.75)	(1.49)	(0.75)	(0.30)	(2.59)	(1.50)	(1.78)
Number of observations	47	47	47	47	47	47	47
Adjusted R2	0.7464	0.7753	0.7774	0.767	0.8529	0.7477	0.857

Absolute value of robust t statistics in parenthesis
*significant at 10% **significant at 5% ***significant at 1%

Table 5: Peripheral diversity and redistribution (IV estimation)

Dependent variable: Transfers and subsidies as % of GDP (average 1985-1995)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peripheral diversity (no distance)	-1.189	-1.168	-0.746	-1.977	-1.406	-1.205	-1.411
	(0.91)	(0.88)	(0.54)	(1.54)	(1.19)	(0.93)	(1.17)
GDP 85-95				2.046***	4.924***		4.947***
				(3.07)	(4.00)		(4.00)
Population 85-95						0.104	-0.184
						(0.17)	(0.51)
UK legal origin		-1.055	-1.617		0.006		-0.058
		(0.45)	(0.65)		(0.00)		(0.02)
French legal origin		1.942	-0.119		3.575		3.647
		(0.82)	(0.03)		(1.19)		(1.18)
Socialist legal origin		0.348	0.149		10.869**		10.748**
		(0.11)	(0.04)		(2.51)		(2.45)
Scandinavian legal origin		-1.306	2.903		5.387		5.225
		(0.36)	(0.50)		(1.02)		(0.95)
Catholic 80			-0.001		-0.002		-0.004
			(0.03)		(0.07)		(0.13)
Muslim 80			-0.061		0.122		0.108
			(0.20)		(0.61)		(0.55)
Protestants 80			-0.083		-0.086		-0.086
			(0.97)		(1.24)		(1.22)
Latitude	15.148*	20.083**	24.464**	9.417	13.054*	15.524*	12.449*
	(1.76)	(2.06)	(2.53)	(1.11)	(1.76)	(1.83)	(1.67)
Latin American dummy	-9.709***	-8.837**	-7.183*	-8.862**	-1.781	-9.525***	-2.125
	(2.75)	(2.30)	(1.89)	(2.57)	(0.54)	(2.68)	(0.65)
Small island dummy	-7.512***	-6.314**	-5.334*	-9.722***	-7.655***	-7.175**	-8.288***
	(3.40)	(2.49)	(1.92)	(4.11)	(3.01)	(2.30)	(2.90)
Intercept	13.45***	10.292*	10.18	-1.502	-33.884***	11.575	-30.629**
	(2.66)	(1.78)	(1.40)	(0.21)	(2.63)	(1.04)	(2.18)
Number of observations	56	56	55	56	55	56	55
Adjusted R2	0.6854	0.6982	0.7394	0.7233	0.8302	0.6856	0.8307

Absolute value of robust t statistics in parenthesis
*significant at 10% **significant at 5% ***significant at 1%

Table 6: Peripheral diversity (without distance) and redistribution

Dependent variable: Transfers and subsidies as % of GDP (average 1985-1995)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peripheral diversity (Britannica)	-4.891***	-4.936***	-4.421***	-4.706***	-3.651***	-4.97***	-3.863***
	(3.62)	(2.93)	(2.61)	(3.50)	(2.59)	(3.59)	(2.71)
GDP 85-95				1.663**	4.55***		4.581***
				(2.50)	(3.86)		(3.95)
Population 85-95						-0.162	-0.375
						(0.29)	(1.04)
UK legal origin		0.117	-0.219		0.868		0.807
		(0.05)	(0.09)		(0.40)		(0.37)
French legal origin		2.109	0.654		3.869		4.044
		(1.05)	(0.22)		(1.42)		(1.46)
Socialist legal origin		1.069	1.315		11.294***		11.062***
		(0.41)	(0.41)		(2.74)		(2.67)
Scandinavian legal origin		-1.767	1.084		4.504		4.05
		(0.61)	(0.22)		(0.99)		(0.85)
Catholic 80			0.004		0.002		-0.001
			(0.11)		(0.07)		(0.03)
Muslim 80			-0.05		0.112		0.083
			(0.16)		(0.56)		(0.42)
Protestants 80			-0.055		-0.066		-0.066
			(0.72)		(1.05)		(1.01)
Latitude	14.524*	20.031**	23.991***	9.838	12.869*	13.931*	11.67*
	(1.87)	(2.27)	(2.66)	(1.26)	(1.76)	(1.82)	(1.61)
Latin American dummy	-9.82***	-8.762**	-7.448**	-8.978***	-2.273	-10.119***	-3.009
	(3.06)	(2.55)	(2.13)	(2.81)	(0.69)	(3.10)	(0.91)
Small island dummy	-8.304***	-7.294***	-6.583**	-9.667***	-8.017***	-8.867***	-9.38***
	(3.68)	(3.05)	(2.57)	(4.16)	(3.18)	(2.86)	(3.45)
Intercept	15.016***	11.233**	10.395	2.199	-30.728**	17.999*	-23.932*
	(3.32)	(2.28)	(1.61)	(0.32)	(2.50)	(1.67)	(1.72)
Number of observations	56	56	55	56	56	56	55
Adjusted R2	0.7209	0.7308	0.7658	0.7467	0.8435	0.7214	0.8457

Absolute value of robust t statistics in parenthesis
*significant at 10% **significant at 5% ***significant at 1%

Table 7: Peripheral diversity and redistribution (Britannica)

Dependent variable: Transfers and subsidies as % of GDP (average 1985-1995)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peripheral diversity no distance (Britannica)	-3.203**	-3.344**	-2.524	-3.235***	-2.198*	-3.222**	-2.347*
	(2.57)	(2.19)	(1.45)	(2.70)	(1.76)	(2.52)	(1.78)
GDP 85-95				1.772***	4.715***		4.747***
				(2.59)	(3.70)		(3.74)
Population 85-95						-0.061	-0.302
						(0.10)	(0.79)
UK legal origin		0.372	-0.304		0.914		0.894
		(0.15)	(0.10)		(0.36)		(0.35)
French legal origin		2.466	0.798		4.154		4.325
		(1.17)	(0.24)		(1.37)		(1.39)
Socialist legal origin		1.165	1.215		11.614***		11.454**
		(0.42)	(0.34)		(2.62)		(2.57)
Scandinavian legal origin		-1.657	1.565		4.912		4.525
		(0.55)	(0.28)		(1.00)		(0.87)
Catholic 80			0.001		0		-0.002
			(0.04)		(0.00)		(0.08)
Muslim 80			-0.071		0.101		0.077
			(0.23)		(0.50)		(0.39)
Protestants 80			-0.061		-0.07		-0.07
			(0.71)		(0.99)		(0.96)
Latitude	15.085*	20.952**	24.413**	10.07	12.843*	14.865*	11.89
	(1.84)	(2.26)	(2.56)	(1.22)	(1.68)	(1.84)	(1.56)
Latin American dummy	-9.805***	-8.776**	-7.462**	-8.925***	-2.118	-9.917***	-2.713
	(2.91)	(2.40)	(2.00)	(2.69)	(0.63)	(2.91)	(0.79)
Small island dummy	-7.888***	-6.901***	-6.04*	-9.41***	-7.686***	-8.097***	-8.791***
	(3.48)	(2.76)	(2.18)	(4.06)	(2.99)	(2.59)	(3.03)
Intercept	14.478***	10.381**	9.931	0.946	-32.596**	15.6	-27.192*
	(3.03)	(1.98)	(1.42)	(0.13)	(2.49)	(1.41)	(1.87)
Number of observations	56	56	55	56	55	56	55
Adjusted R2	0.704	0.7151	0.7486	0.7334	0.8328	0.7041	0.8342

Absolute value of robust t statistics in parenthesis
*significant at 10% **significant at 5% ***significant at 1%

Table 8: Peripheral diversity (without distance) and redistribution (Britannica)

Country	Polarization	Diversity	Difference
Bulgaria	1	14	13
Guatemala	2	16	14
Bolivia	3	2	-1
Estonia	4	10	6
Trinidad	5	23	18
Luxembourg	6	12	6
Belgium	7	7	0
Belize	8	3	-5
Macedonia	9	11	2
Canada	10	1	-9
Latvia	11	15	4
Peru	12	13	1
Moldova	13	5	-8
Suriname	14	9	-5
Switzerland	15	6	-9
Paraguay	16	8	-8
Ecuador	17	21	4
Panama	18	17	-1
Romania	19	20	1
Yugoslavia	20	27	7
United States	21	4	-17
Slovakia	22	31	9
Lithuania	23	25	2
Andorra	24	40	16
Ireland	25	41	16
France	26	18	-8
Hungary	27	24	-3
Argentina	28	22	-6
Mexico	29	19	-10
Finland	30	39	9
Ukraine	31	32	1
Spain	32	36	4
Netherlands Antilles	33	45	12
Albania	34	33	-1
Greece	35	26	-9
Netherlands	36	29	-7
Nicaragua	37	42	5
Guyana	38	38	0
Germany	39	28	-11
Belarus	40	44	4
Sweden	41	30	-11
Venezuela	42	35	-7
Honduras	43	34	-9
Chile	44	46	2
Slovenia	45	56	11
Croatia	46	48	2
Norway	47	55	8
Czech Rep	48	50	2
Poland	49	52	3
Liechtenstein	50	58	8
Puerto Rico	51	59	8
United Kingdom	52	43	-9
Jamaica	53	47	-6
Costa Rica	54	53	-1
Austria	55	49	-6
Dominican Republic	56	57	1
Colombia	57	37	-20
Uruguay	58	60	2
Italy	59	51	-8
Brazil	60	54	-6
Malta	61	61	0
US Virgin Islands	62	65	3
Bahamas	63	66	3
Denmark	64	64	0
Portugal	65	62	-3
Barbados	66	67	1
El Salvador	67	63	-4
Bosnia-Herzegovina	68	68	0
Iceland	69	69	0
Turks&Caicos	70	70	0
St Vincent	71	71	0
St Kitts	72	72	0
Haiti	73	73	0
Cuba	74	74	0

Source: Ethnologue, own calculations

Table 9: Ranking of polarization versus diversity

Dependent variable: Transfers and subsidies as % of GDP (average 1985-1995)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peripheral polarization	-16.421**	-16.521*	-16.475**	-13.549	-14.541**	-17.547**	-16.326**
	(2.08)	(1.85)	(2.05)	(1.63)	(2.24)	(2.14)	(2.28)
GDP 85-95				1.565**	4.68***		4.721***
				(2.06)	(3.87)		(4.00)
Population 85-95						-0.251	-0.447
						(0.41)	(1.23)
UK legal origin		-0.754	-1.24		0.094		-0.008
		(0.32)	(0.54)		(0.04)		(0.00)
French legal origin		2.123	0.282		3.684		3.898
		(0.92)	(0.09)		(1.29)		(1.35)
Socialist legal origin		1.566	1.145		11.483***		11.241***
		(0.52)	(0.35)		(2.75)		(2.68)
Scandinavian legal origin		-0.882	2.467		5.66		5.135
		(0.28)	(0.50)		(1.28)		(1.11)
Catholic 80			0		-0.002		-0.005
			(0.01)		(0.05)		(0.18)
Muslim 80			0.024		0.184		0.158
			(0.07)		(0.77)		(0.66)
Protestants 80			-0.074		-0.082		-0.082
			(0.97)		(1.31)		(1.29)
Latitude	14.918*	19.427**	24.837***	10.514	13.311*	14.003*	11.972*
	(1.77)	(1.97)	(2.60)	(1.23)	(1.76)	(1.69)	(1.64)
Latin American dummy	-9.464***	-8.4**	-6.626*	-8.682**	-1.425	-9.919***	-2.231
	(2.71)	(2.24)	(1.82)	(2.48)	(0.43)	(2.83)	(0.69)
Small island dummy	-8.078***	-6.801***	-6.067**	-9.192***	-7.711***	-9***	-9.373***
	(3.76)	(2.87)	(2.44)	(4.05)	(3.01)	(2.83)	(3.36)
Intercept	13.793***	10.325*	9.799	1.582	-32.391**	18.447	-24.343*
	(2.77)	(1.85)	(1.41)	(0.20)	(2.55)	(1.57)	(1.77)
Number of observations	56	56	55	56	55	56	55
Adjusted R2	0.7	0.7128	0.7555	0.7223	0.8384	0.701	0.8413

Absolute value of robust t statistics in parenthesis
*significant at 10% **significant at 5% ***significant at 1%

Table 10: Peripheral polarization and redistribution

Dependent variable: Transfers and subsidies as % of GDP (average 1985-1995)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peripheral diversity	-4.191***	-4.216*	-3.375	-5.968***	-3.363*	-4.165***	-3.099*
	(2.74)	(1.86)	(1.38)	(3.55)	(1.84)	(2.95)	(1.75)
Peripheral polarization	1.731	1.729	-1.758	13.209	0.124	1.506	-2.35
	(0.18)	(0.13)	(0.13)	(1.40)	(0.01)	(0.17)	(0.26)
GDP 85-95				2.059***	4.678***		4.709***
				(2.95)	(4.07)		(4.14)
Population 85-95						-0.025	-0.331
						(0.04)	(1.00)
UK legal origin		-0.344	-0.724		0.609		0.492
		(0.16)	(0.29)		(0.30)		(0.24)
French legal origin		1.852	0.317		3.717		3.874
		(0.91)	(0.10)		(1.40)		(1.43)
Socialist legal origin		0.31	0.535		10.872***		10.741**
		(0.10)	(0.16)		(2.61)		(2.57)
Scandinavian legal origin		-2.058	1.113		4.31		4.027
		(0.65)	(0.21)		(0.91)		(0.82)
Catholic 80			0.003		0.002		-0.001
			(0.09)		(0.08)		(0.03)
Muslim 80			0.001		0.16		0.143
			(0.00)		(0.82)		(0.72)
Protestants 80			-0.061		-0.069		-0.07
			(0.77)		(1.06)		(1.06)
Latitude	14.611*	20.252**	24.877***	8.685	13.355*	14.52*	12.36*
	(1.82)	(2.22)	(2.73)	(1.14)	(1.94)	(1.79)	(1.80)
Latin American dummy	-9.778***	-8.8**	-7.178**	-8.881***	-1.977	-9.822***	-2.531
	(2.99)	(2.51)	(2.05)	(2.81)	(0.63)	(2.89)	(0.81)
Small island dummy	-9.04***	-8.032***	-7.143***	-10.914***	-8.782***	-9.127***	-9.929***
	(4.00)	(3.21)	(2.66)	(4.45)	(3.28)	(2.91)	(3.49)
Intercept	15.097***	11.666**	10.572	-0.42	-31.606***	15.558	-25.707**
	(3.21)	(2.29)	(1.60)	(0.06)	(2.63)	(1.41)	(1.98)
Number of observations	56	56	55	56	55	56	55
Adjusted R2	0.7154	0.7256	0.7633	0.7512	0.8462	0.7154	0.8477

Absolute value of robust t statistics in parenthesis
*significant at 10% **significant at 5% ***significant at 10%

Table 11: Peripheral diversity, peripheral polarization, and redistribution

Dependent variable: Transfers and subsidies as % of GDP (average 1985-1995)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peripheral polarization (no distance)	-8.905	-8.425	-6.211	-10.928	-9.967	-9.044	-11.344
	(1.05)	(0.90)	(0.67)	(1.48)	(1.36)	(1.05)	(1.44)
GDP 85-95				1.848***	4.911***		4.971***
				(2.65)	(3.96)		(4.05)
Population 85-95						-0.044	-0.355
						(0.07)	(0.90)
UK legal origin		-1.105	-1.606		-0.028		-0.108
		(0.45)	(0.65)		(0.01)		(0.05)
French legal origin		1.944	-0.126		3.535		3.698
		(0.80)	(0.04)		(1.17)		(1.20)
Socialist legal origin		0.701	0.379		11.287***		11.085**
		(0.23)	(0.11)		(2.59)		(2.53)
Scandinavian legal origin		-1.104	2.97		5.714		5.257
		(0.31)	(0.53)		(1.13)		(0.98)
Catholic 80			0		0		-0.003
			(0.01)		(0.01)		(0.11)
Muslim 80			-0.055		0.13		0.105
			(0.18)		(0.64)		(0.53)
Protestants 80			-0.082		-0.085		-0.085
			(0.98)		(1.28)		(1.26)
Latitude	15.665*	20.29**	24.79**	10.571	13.422*	15.514*	12.423
	(1.83)	(2.03)	(2.51)	(1.22)	(1.71)	(1.83)	(1.62)
Latin American dummy	-9.33***	-8.419**	-6.881*	-8.381**	-1.272	-9.408***	-1.88
	(2.66)	(2.21)	(1.84)	(2.42)	(0.38)	(2.67)	(0.57)
Small island dummy	-7.342***	-6.077**	-5.221**	-9.039***	-7.331***	-7.498**	-8.639***
	(3.41)	(2.49)	(2.00)	(3.99)	(2.96)	(2.42)	(3.13)
Intercept	13.022***	9.854*	9.81	-0.974	-34.465***	13.838	-28.333**
	(2.60)	(1.73)	(1.35)	(0.14)	(2.65)	(1.19)	(2.00)
Number of observations	56	56	55	56	55	56	55
Adjusted R2	0.6861	0.6988	0.7403	0.7177	0.8309	0.6861	0.8328

Absolute value of robust t statistics in parenthesis
*significant at 5% **significant at 1%

Table 12: Peripheral diversity (without distance) and redistribution