Soc Choice Welf (2017) 49:787–823
DOI 10.1007/s00355-016-1017-5

ORIGINAL PAPER

Peripheral diversity: transfers versus public goods

Klaus Desmet[1,2] · **Ignacio Ortuño-Ortín**[3] · **Shlomo Weber**[4,5]

Received: 21 April 2016 / Accepted: 15 December 2016 / Published online: 9 January 2017

Abstract This paper advances the hypothesis that in societies that suffer from ethnolinguistic center-periphery tension it is harder to agree on public goods than on transfers. After micro-founding a new peripheral diversity index, it puts forth a simple theory in which the cost of public goods increases with peripheral ethnolinguistic diversity and tax compliance decreases with overall ethnolinguistic diversity. It then empirically explores the relation between public goods provision, transfers, peripheral diversity and overall diversity. Consistent with the theory, we find that higher levels of peripheral diversity are associated with less provision of public goods, but more transfers, whereas higher levels of overall diversity have a negative association with

We thank two anonymous referees for thoughtful comments. I. Ortuño-Ortín acknowledges the financial support of the Spanish Ministry of Science and Innovation, Project ECO-2013-42710-P, and S. Weber wishes to acknowledge the support of the Ministry of Education and Science of the Russian Federation, Grant No. 14.U04.31.0002, administered through the NES CSDSI.

✉ Klaus Desmet
kdesmet@smu.edu

Ignacio Ortuño-Ortín
iortuno@eco.uc3m.es

Shlomo Weber
sweber@smu.edu

1 Department of Economics, Southern Methodist University, Dallas, USA

2 CEPR, London, United Kingdom

3 Department of Economics, Universidad Carlos III, Madrid, Spain

4 Department of Economics, SMU, Dallas, USA

5 New Economic School, Moscow, Russia

transfers. Public goods and transfers are therefore substitutes in their reaction to a change in peripheral diversity.

1 Introduction

Empirical evidence has shown that countries that are ethnolinguistically more diverse exhibit lower levels of transfers; those same countries also tend to display a worse provision of public goods.[1] As argued by La Porta et al. (1999) and Alesina et al. (2003), one reason may be that in more diverse societies people are less willing to pay taxes to finance transfers and public goods. It is therefore not surprising that when analyzing the effects of ethnolinguistic diversity, public goods and transfers are often put in the same bag: both suffer from lower solidarity in more diverse societies.

Although public goods and transfers have much in common, in this paper we advance the hypothesis that politically deciding on public goods is much harder than on transfers. To illustrate this in the case of public goods, there may be disagreement over the particular shape of the country's road network, and it may be difficult to decide on where to locate the nation's capital. In contrast, in the case of transfers, there is much less discussion over their shape or their type, because, after all, "money is money".

How does the difficulty to decide on public goods relate to a country's diversity? Casual observation suggests that the political conflict over public goods often arises from the antagonism between the minorities (the "periphery") and the dominant group (the "center"), rather than from tension between all groups. For instance, the attempt at making Hindi into India's sole national language in 1965 gave rise to widespread protests against "Hindi imperialism". This was a conflict between the periphery and the center, not between the peripheral groups themselves. Another example is 19th century France, where schooling was used as a "a major agent of acculturation", eliminating local culture and unifying the country (Weber 1979). This process was not exempt of tension, with many peripheral regions, especially in the Pyrenees and Brittany, resisting the center. Something similar occurred in 19th century Italy, where through compulsory education, the Northern elite was able to impose Italian as the common language, in spite of not being popular (Alesina and Reich 2015). As a last example, in present-day Spain one point of contention in the political conflict between the center and the regions is the country's star-shaped infrastructure network, with many of the roads and railroads passing through Madrid.

This discussion suggests that two types of diversity may matter in determining public goods and transfers. When it comes to people's willingness to pay taxes, it depends on a society's overall diversity, whereas when it comes to the political tension surrounding decisions on public goods, it depends on a society's tension between the center and the periphery. In the theory, we define a society's overall diversity as the expected ethnolinguistic distance between any two randomly drawn individuals — this measure is of course nothing else than Greenberg's B-index. In addition to a society's overall diversity, which derives from tension between any two individuals, we also

[1] See, e.g., La Porta et al. (1999), Alesina et al. (2003), Desmet et al. (2009), and Desmet et al. (2012).

define a society's peripheral diversity, which stems from the antagonism between the dominant group and the minorities. The main difference between both measures is that peripheral diversity ignores any potential tension between minority groups, whereas overall diversity treats all groups symmetrically.

We then propose a simple theory of the relation between peripheral diversity, overall diversity, public goods and transfers. A society is made up of individuals who belong to different ethnolinguistic groups. Their preferences are quasi-linear in public consumption and private consumption, such that any change in income is absorbed by private consumption. Public goods and transfers are financed by a proportional tax. Decisions about public goods, transfers and taxes are taken by the median voter. We then make two assumptions. First, consistent with La Porta et al. (1999) and Alesina et al. (2003), the cost of tax enforcement is increasing in society's overall diversity because of people's reduced willingness to comply. Second, consistent with the hypothesis above, the cost of public goods is increasing in society's peripheral diversity.

This simple theory yields four predictions. First, the level of public goods is decreasing in the level of peripheral diversity. This happens because an increase in peripheral diversity makes the provision of public goods more costly, leading to a drop in their provision. Second, the tax rate does not depend on society's peripheral diversity. Together with the first prediction, this implies that an increase in peripheral diversity leads to more transfers. This means that public goods and transfers are substitutes in how they react to a change in peripheral diversity. Third, the tax rate declines in the level of overall diversity. This implies that higher overall diversity lowers transfers. Fourth, because preferences are quasi-linear, a higher level of overall diversity does not affect the provision of public goods, despite its negative effect on tax revenues.

We then test these four predictions using detailed data on language use and linguistic distances from Ethnologue. These data enable us to compute measures of Greenberg's B-index and peripheral diversity for 226 countries. With these indices in hand, we analyze the relation between peripheral diversity, overall diversity, public goods and transfers in a large cross-section of countries. Consistent with the first two theoretical predictions, we find that an increase in peripheral diversity lowers the provision of public goods, but increases transfers. Consistent with the last two theoretical predictions, an increase in overall diversity has no effect on the provision of public goods, but lowers the level of transfers. Our most important conclusion is that public goods and transfers act as substitutes when the tension between the center and the periphery increases. Once again, the intuition is that the antagonism between the center and the periphery complicates political decision-making, and this disproportionately hurts public goods.

The rest of the paper is organized as follows. Section 2 proposes a peripheral diversity index and develops a theory of diversity, public goods and transfers. Section 3 tests the theory using cross-country data. Section 4 concludes.

2 Theory

We develop a simple theory of a society with both public goods and transfers, financed by a proportional tax. Collecting taxes is challenging, especially in diverse societies.

This makes the cost of tax enforcement an increasing function of a country's ethnolinguistic diversity. Additionally, in countries with a high degree of tension between the center and the periphery, drawn-out political discussions increase the cost of providing public goods. This theory yields predictions for the relation between peripheral diversity, overall diversity, public goods and transfers. These theoretical predictions will serve as a basis for our empirical investigation. Before presenting the model, we start by proposing a framework that micro-founds a peripheral diversity index which captures the alienation that arises between the dominant center and the peripheral minority groups. An early version of this index appeared in a working paper by Desmet et al. (2005).

2.1 A general model of peripheral alienation

Consider a country with total population normalized to 1. There are $K+1$ distinct groups, labeled $0, 1, \ldots, K$. The groups are defined along one dimension of identity, such as ethnicity, language, religion or place of origin.[2] In the theory we refer to them generically as ethnolinguistic groups, whereas in the empirics we focus, for reasons that will become apparent later, on linguistic groups. One group, 0, called "center" or "dominant group", has a share s_0 of the population, whereas the other K groups, called "minorities" or "peripheral groups", have population shares s_k. Each citizen of the country belongs to one and only one group. Hence, the vector $(s_0, s_1, \ldots, s_K)$ belongs to the $k+1$ -dimensional simplex Δ and $\sum_{k=0}^{K} s_k = 1$. Our model focuses on the frequently observed cases where the "dominant" group contains at least as many individuals as any of the minority groups:[3]

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

Hence, we examine the subset of vectors $\mathcal{S} \subset \Delta$ given by:

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

A crucial element of our model is the introduction of ethnolinguistic distances 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 that $\tau_{kl} = \tau_{lk}$. The set of such matrices is denoted by $\mathcal{T}$. In the empirical part of the paper τ_{kl} is the linguistic distance between the language spoken by group k and the language spoken by group l.[4]

2 We thus ignore the possible overlap between multiple dimensions of identity.

3 There are of course cases where the dominant group does not correspond to the biggest group. Examples include the Tutsis during different periods of Rwandan history and the Afrikaners of South Africa before the end of Apartheid.

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

The population shares and ethnolinguistic distances will be enough to determine the level of *peripheral diversity* which reflects the tension between the center and the peripheral groups. 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 such that the value of inter-group alienation experienced by group k towards group l is given by

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

which depends on the size of both groups and the ethnolinguistic distance between them. Because of our focus on alienation between the center and the periphery, it is natural to allow for different functional forms, one for alienation towards the center and another for alienation towards the periphery. In particular, function $f_{pc}(s_0, s_k, \tau_{0k})$ gives the centrifugal alienation experienced by each of the $k = 1, \ldots, K$ minority groups towards the center, whereas the function $f_{cp}(s_0, s_k, \tau_{0k})$ gives the centripetal alienation experienced by the center towards each of the $k = 1, \ldots, K$ minority groups. At this point, the functions f_{pc} and f_{cp} have been constructed from the notion that alienation originates directly between groups. In the next subsection we will discuss how we can derive the functions f_{pc} and f_{cp} from an alienation function at the individual level.

The country's total level of *peripheral alienation* is then the sum of the alienation from the minority groups towards the center and from the center towards the minority groups. Formally, for every vector $s = (s_0, \ldots, s_K) \in \mathcal{S}$, distance matrix $T \in \mathcal{T}$ and alienation functions f_{cp} and f_{pc}, we define the total level of *peripheral alienation* $PA(s, T)$ as

$$PA(s, T) = \sum_{k=1}^{K} \left(f_{cp}(s_0, s_k, \tau_{0k}) + f_{pc}(s_0, s_k, \tau_{0k}) \right). \tag{1}$$

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

Condition 1 (*Continuity*) The functions f_{cp} and f_{pc} are continuous on $\mathcal{S}$.

Condition 2 (*Alienation is increasing in distance*) For every k and every $s \in \mathcal{S}$, the functions $f_{cp}(s_0, s_k, \cdot)$ and $f_{pc}(s_0, s_k, \cdot)$ 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 $f_{cp}(s_0, \cdot, \tau)$ is 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 $f_{pc}(\cdot, s_k, \tau)$ is concave on the interval $[\max[s_k, \frac{1}{k+1}], 1 - s_k]$.

Condition 4 (*Supermodularity*) For every $s \in \mathcal{S}$ with $s_k < s_l$, and $\tau_{kl}^1 < \tau_{kl}^2$, the following holds:

$$f_{cp}(s_0, s_l, \tau_{kl}^1) - f_{cp}(s_0, s_k, \tau_{kl}^1) < f_{cp}(s_0, s_l, \tau_{kl}^2) - f_{cp}(s_0, s_k, \tau_{kl}^2);$$

and

$$f_{pc}(s_0, s_l, \tau_{kl}^1) - f_{pc}(s_0, s_k, \tau_{kl}^1) < f_{pc}(s_0, s_l, \tau_{kl}^2) - f_{pc}(s_0, s_k, \tau_{kl}^2).$$

Conditions 1 and 2 impose continuity and monotonicity. Condition 3 is the key 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.[5] In the case the alienation functions are differentiable, the supermodularity condition implies, in particular, that $\frac{\partial f_{cp}(s_0, s, \tau)}{\partial s \partial \tau} > 0$.

The following proposition states that peripheral alienation increases when the minority groups that are more distant from the center are larger:

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

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

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. For example, Shannon's information entropy index satisfies this property (Shannon 1949). Thus, our index $PA(s, T)$ can be interpreted as an index of *peripheral diversity*.[6] The proposition clarifies the relationship between diversity and the nature of the inter-group alienation function. Thus, whenever the functions f_{pc} and f_{cp} are concave in the size of the

[5] Assuming convexity, instead of concavity, would give us an index of polarization. We later return to this issue.

[6] Notice that Condition 3 requires concavity of the function $f_{cp}(s_0, \cdot, \tau)$. Thus, this 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 $f_{cp}(s_0, \cdot, \tau)$ is not a necessary condition to obtain the solution. For example, if the function $f_{pc}(\cdot, s_k, \tau)$ is "sufficiently" concave, the function $f_{cp}(s_0, \cdot, \tau)$ need not be concave.

group, the index *PA* 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 imposed the "opposite" condition by assuming the functions f_{cp} and f_{pc} to be *convex*. This would imply that if groups k and l have the same ethnolinguistic 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 inter-group alienation increases in the size of the group in a concave way or in a convex way, the aggregate index *PA* can be interpreted as satisfying a necessary property of either a measure of diversity or a measure of polarization. We summarize this insight in the following corollary of Proposition 1:

Corollary 1 *Assume Conditions 1, 2 and 4 hold. Then, if Condition 3 also holds, the index of peripheral alienation, $PA(s,T)$, can be interpreted as an index of peripheral diversity, $PD(s,T)$, so*

$$PD(s,T) = PA(s,T).$$

If, however, the "opposite" to Condition 3 holds, i.e. if the functions f_{cp} and f_{pc} are convex, the index of peripheral alienation can be interpreted as an index of peripheral polarization.

2.2 A specific index of peripheral alienation

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

To come up with these functions, we follow the identification-alienation framework of Esteban and Ray (1994), though we will allow for a more flexible approach.[7] An individual of ethnolinguistic group k feels identified with other individuals of the same group. 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 where small ethnic, linguistic, cultural or religious groups feel a keener sense of community and a stronger desire to assert their identity.

An agent of ethnolinguistic group k feels more alienated from someone of ethnolinguistic group l the greater the distance τ_{kl}. This alienation is influenced by the sense of identification. In particular, an individual attaches more weight to the distance τ_{kl} if his

[7] For a similar approach used to derive a variety of indices—Greenberg's *A* index, Greenberg's *B* index, Esteban and Ray's (1994) polarization index, Reynal-Querol's (2002) polarization index and a simple version of the peripheral index—see Desmet et al. (2009).

sense of identification with his own group is stronger. As defined in Esteban and Ray (1994), the alienation experienced by an individual of group k towards an individual of group l is $s_k^\alpha \tau_{kl}$. Since there is a proportion s_0 of individuals belonging to the dominant group, the centrifugal alienation of an agent of minority group 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 a dominant group, independently of how many people actually belong to that dominant language, so that we set $\beta = 0$. In that case, an individual's centrifugal alienation is $s_k^\alpha \tau_{0k}$. Setting $\beta = 0$ captures the idea that some policy choices may be imposed by the center because of it being the center and not because of its exact size. If so, it is reasonable to think that the centrifugal alienation associated with these policies are independent of the center's exact size. If each individual of minority group k feels an alienation $s_k^\alpha \tau_{0k}$ towards the center, and if a share s_k of the population belongs to group k, then the centrifugal alienation of all individuals of group k is $s_k^{1+\alpha} \tau_{0k}$. Thus, the inter-group alienation function f_{pc} is given by

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

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. There is no reason why the alienation experienced by the center towards the periphery should be independent of the peripheral groups' sizes. Hence, the centripetal alienation felt by members of the central group depends on the size of the minorities, so that

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

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

$$\begin{aligned} PA(s, T) &= \sum_{k=1}^{K} (f_{pc}(s_0, s_k, \tau_{0k}) + f_{cp}(s_0, s_k, \tau_{0k})) \\ &= \sum_{k=1}^{K} (s_k^{1+\alpha} \tau_{0k} + s_k s_0^{1+\alpha} \tau_{0k}) \end{aligned} \tag{4}$$

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*. We summarize this result in the following corollary:

Corollary 2 *If* $\alpha < 0$*, the index* (4) *satisfies Conditions* 1–4, *and can thus be viewed as an index of peripheral diversity. Hence,*

$$PD(s, T) = PA(s, T) = \sum_{k=1}^{K} (s_k^{1+\alpha} \tau_{0k} + s_k s_0^{1+\alpha} \tau_{0k}) \quad \text{if } \alpha < 0.$$

If, in contrast, $\alpha > 0$, the index (4) can be interpreted as an index of peripheral polarization.

To illustrate the difference between diversity and polarization, consider a country with three groups. Their respective sizes are s_0, s_1 and s_2. Group 0 is the dominant one, and groups 1 and 2 are minorities. Further assume that the distance between each minority group and the dominant group 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. For a given share of the dominant group, the maximum diversity is reached when $s_1 = s_2$. In other words, we face most diversity with two (equally sized) minority groups. If $\alpha > 0$, we obtain a measure reflecting polarization that attains the highest level if one of the two remaining groups disappears. In other words, the level of polarization is highest if we have only one, rather than two, minority groups. This insight does not change once we allow for different distances between groups.

2.3 Peripheral diversity, public goods and transfers

Denote the income of individual i by y_i. Average income is y, and median income is y_m, where $y_m < y$. The government provides everyone with the same level of public goods and lump-sum transfers. All individuals have the same preferences over public consumption (G) and private consumption (c), represented by the quasi-linear utility function

$$u(G, c) = 2G^{1/2} + c.$$

The government pays for public goods and transfers through a proportional tax. The cost of public goods is increasing in the political conflict incurred to reach an agreement. Our discussion in the introduction suggests that this political conflict often has a markedly center-periphery character. For example, the center and the periphery may have long drawn-out discussions about which language to use in schools and hospitals, or it may take many fights for both sides to agree on the shape of the country's road network. We therefore postulate that the cost of the public goods is proportional to the tension individuals from the periphery feel towards the center, $\sum_{k=1}^{K} s^{1+\alpha} \tau_{0k}$, and the tension individuals from the center feel towards the periphery, $\sum_{k=1}^{K} s_k s_0^{1+\alpha} \tau_{0k}$. That is, the cost of the public good, p, is an increasing function of the peripheral index, PD; for simplicity, we assume that $p = (PD + \eta)$, with $\eta > 0$.[8]

Collecting taxes is challenging in diverse societies. We assume that the cost of tax enforcement is an increasing function of the average ethnolinguistic distance between individuals in society, $\sum_k \sum_l s_l s_k \tau_{lk}$. This captures the idea in La Porta et al. (1999) and Alesina et al. (2003) that people are less tax compliant in more diverse societies. In addition, the cost of tax enforcement is assumed to be increasing in the tax rate, t.

8 As we will later argue, the price of the public good needs to be high enough. Since PD can be zero, this requires $\eta > 0$.

Hence, a tax rate t will generate government income $ty(1-\gamma t\sum_k\sum_l s_l s_k \tau_{lk})$, where $\gamma > 0$ and $\gamma \sum_k\sum_l s_l s_k \tau_{lk} < 1$. Notice that $\sum_k\sum_l s_l s_k \tau_{lk}$ is nothing else than Greenberg's B-index, which measures the average ethnolinguistic distance between two randomly picked individuals. Hence, $B = \sum_k\sum_l s_l s_k \tau_{lk}$. In our model Greenberg's B-index is meant to capture the tension between all groups, as opposed to peripheral diversity, which measures the center-periphery tension.[9] The society's budget constraint can then be written as

$$(PD+\eta)G + r = ty(1-\gamma t B).$$

The society has to choose the level of public good, G, the income tax rate, t, and the transfer that each individual receives, r. We adopt the view that society implements a Condorcet winner policy. We will now show that the optimal policy for the median income agent is a Condorcet winner. His optimal policy (G^*, t^*, r^*) is the one given by the solution to the following optimization problem

$$\begin{aligned} \max_{G,t,r} \quad & 2G^{1/2} + (1-t)y_m + r \\ \text{s.t.} \quad & (PD+\eta)G + r = ty(1-\gamma t B) \\ & 0 \le t \le 1 \\ & r \ge 0. \end{aligned}$$

The first order conditions for an interior solution of this problem yield

$$G^* = \left(\frac{1}{PD+\eta}\right)^2 \tag{5}$$

$$t^* = \frac{y-y_m}{2\gamma B y} \tag{6}$$

$$r^* = -(PD+\eta)G + ty(1-t\gamma B). \tag{7}$$

Note that condition $t \le 1$ implies that an interior solution satisfies[10]

$$\frac{y-y_m}{y} \le 2\gamma B$$

and that the level of public good G^* does not depend on the tax rate t. Furthermore, Eqs. (5) and (7) imply that[11]

$$r = ty(1-t\gamma B) - \frac{1}{PD+\eta}. \tag{8}$$

[9] Greenberg's B-index can be derived from the same identification-alienation framework. See Desmet et al. (2009).

[10] Our results can be also extended to the case of a corner solution in which the tax rate is $t = 1$.

[11] Since we assume that transfers are positive, the price of the public good $PD + \eta$ has to be high enough to ensure $r > 0$.

Because of the budget constraint $(PD + \eta)G + r = ty(1 - \gamma t B)$, when comparing two different policies we only need to focus on the level of the public good G and the tax rate t. Consider a policy (G', t') that is not optimal for the median income agent. In that case,

$$2G^{*^{1/2}} + (1 - t^*)y_m + t^* y(1 - t^* \gamma B) - (PD + \eta)G^* > 2G'^{1/2} + (1 - t')y_m + t' y(1 - t' \gamma B) - (PD + \eta)G'$$

which can be rewritten as

$$(t' - t^*)y_m > 2G'^{1/2} + t' y(1 - t' \gamma B) - (PD + \eta)G' - 2G^{*^{1/2}} - t^* y(1 - t^* \gamma B) + (PD + \eta)G^* \tag{9}$$

If $t' < t^*$, it follows that an inequality equivalent to (9) also holds for all $y_i \leq y_m$, so that at least 50% of the individuals prefers (G^*, t^*) to (G', t'). If $t' > t^*$, it follows that an inequality equivalent to (9) holds for all $y_i \geq y_m$, so that again at least 50% of the individuals prefers (G^*, t^*) to (G', t'). Note furthermore that if the tax rate t^* is implemented, all agents agree on the optimal level of public good, G^*. Thus, the optimal policy of the median income agent, (G^*, t^*), is a Condorcet winner policy.

We now analyze how the level of public goods and transfers depend on peripheral diversity and overall diversity. From (5) it is obvious that

$$\frac{dG^*}{dPD} < 0 \quad \text{and} \quad \frac{dG^*}{dB} = 0. \tag{10}$$

Combining (6) and (8), we can write

$$r^* = -\frac{1}{PD + \eta} + \left(\frac{y - y_m}{4\gamma B}\right)\left(1 + \frac{y_m}{y}\right). \tag{11}$$

Given that $y_m < y$, from (11) it follows that

$$\frac{dr^*}{dPD} > 0 \quad \text{and} \quad \frac{dr^*}{dB} < 0. \tag{12}$$

Before turning to the empirics, it is useful to provide some intuition for our findings in (10) and (12). When peripheral diversity increases, the cost of public goods goes up, but the cost of enforcing taxes does not change. As a result, the drop in public goods provision, due to its higher price, is compensated by an increase in transfers. When, instead, overall diversity goes up, collecting taxes becomes more expensive. The quasi-linear nature of the preference function then implies that the provision of public goods does not change, so that the lower tax revenues must entail a lower level of transfers.

3 Empirical analysis

In our empirical analysis we explore the relation between peripheral diversity, overall diversity, public goods provision and redistribution through transfers. In particular, we test whether greater peripheral diversity is associated with worse public goods provision, but higher levels of transfers. We also test whether higher overall diversity is associated with lower transfers. Before doing so, we discuss how we define groups, how we measure distances between them, and which parameter values to use to measure peripheral diversity (4).

3.1 Groups, distances and parameter values

Groups are defined by the language they use. Though the theory applies more broadly to any dimension of identity, we focus on language because of the availability of detailed data on both language use and linguistic distances. In fact, even studies that focus on ethnicity, such as Fearon (2003), often use language to define distances between ethnicities.[12]

Linguists use a tree structure to represent the distance between languages. This tree captures the genealogy of languages, classifying them by their family structure. For example, if Hindi and Russian got separated nearly 7000 years ago, whereas Russian and Byelorussian split apart less than 1000 years, then Hindi and Russian are more distant cousins—and hence will have less branches in common in the language tree—than Russian and Byelorussian (Gray and Atkinson 2003). The proximity of two languages can then be measured by the number of shared branches in the language tree.

In our theory we are interested in a distance measure τ_{lk} between languages l and k that is related to the alienation an individual from group l feels towards an individual from group k. A priori there is no reason why τ_{lk} should be a linear function of the number of shared branches in the language tree. For example, if alienation stems from the inability to understand each other, then even relatively small distances in the language tree may be enough to hinder communication. If, instead, alienation is rooted in deep cleavages going back thousands of years, then the number of shared branches must be very low for tension to arise. We therefore follow Fearon (2003) and Desmet et al. (2009) and measure τ_{lk} as a nonlinear negative function of b_{lk}, the number of branches shared by languages l and k. Hence,

$$\tau_{lk} = 1 - \left(\frac{b_{lk}}{m}\right)^{\delta}, \tag{13}$$

where m is the maximum number of branches between any two languages and δ is a parameter that determines how fast the distance τ_{lk} declines as the number of shared branches increases. Lower values of δ imply a more convex function, meaning that linguistic distances only become important when two languages are sufficiently apart

[12] In many countries, especially in Africa and Asia, language coincides with ethnicity.

in the language tree. Going back to our previous example, if alienation stems from the inability to understand each other, then τ_{lk} becomes important as soon as two languages are no longer close cousins in the language tree. In that case, a high value of δ would be appropriate. In our benchmark analysis we follow Fearon (2003) and use a value of $\delta = 0.5$, but explore other values in our robustness checks. As for the value of α, our focus is on peripheral diversity, rather than peripheral polarization. We therefore choose a value of $\alpha = -0.5$. To measure the number of shared branches between any two languages, we use the language tree from the Ethnologue project (Gordon 2005).

3.2 Peripheral diversity and Greenberg's B-index

In this section we provide estimates of peripheral diversity and Greenberg's B-index for a broad cross-section of countries. Recall from our theory that peripheral diversity captures the antagonism between the dominant group and the minority groups, whereas Greenberg's B-index captures the antagonism between all groups. In both cases, this antagonism is increasing in the linguistic distances between the groups. Using data from Ethnologue (Gordon 2005), Table 10 in Appendix B shows the values of peripheral diversity ($\alpha = -0.5$ and $\delta = 0.5$) and Greenberg's B-Index ($\delta = 0.5$) for 226 countries. The linear correlation between the two indices is 0.73, whereas the rank correlation is 0.92. Though high, there are notable differences between both indices. Some countries have a relatively high degree of peripheral diversity, but a relatively low degree of overall diversity. For example, Mexico ranks 22 in terms of peripheral diversity, but only 153 in terms of Greenberg's B-index. Likewise, Russia ranks 38 in terms of peripheral diversity, but 117 in terms of Greenberg's B-index. Some countries exhibit the opposite pattern, with relatively low degrees of peripheral diversity, in spite of having relatively high levels of Greenberg's B-index. Examples include Belize and Bolivia.

3.3 Public goods, peripheral diversity and Greenberg's B-index

Table 1 reports our benchmark regression of public goods on Greenberg's B-index and peripheral diversity. To give a broad overview of different public goods, we include two related to health (child mortality and measles immunization), two related to education (illiteracy and school attainment) and two related to infrastructure (access to improved sanitation and road density). In addition to our variables of interest, we also control for GDP per capita, regional dummies, absolute latitude and roughness of terrain. In general the data cover the period 1990–2010. Appendix B provides a detailed description of the data sources and their time spans. As expected, Table 1 shows that, whenever statistically significant, income per capita and distance from the equator (absolute latitude) improve public goods outcomes, whereas roughness of terrain worsens them. As for regional differences, sub-Saharan Africa fares worst.

Turning to our two variables of interest, we find that peripheral diversity tends to be associated with worse outcomes. In all but two of the cases, the coefficients are statistically significant at the 5% level. In one of the remaining cases—school

Table 1 Public goods, Greenberg's B-index and peripheral diversity

	(1) Log child mortality	(2) Measles immunization rate	(3) Illiteracy rate	(4) Log school attainment	(5) % Access to sanitation	(6) Roads (km per 1000 population)
Greenberg's B-index	0.174 [0.90]	−2.645 [−0.50]	2.471 [0.36]	0.104 [0.73]	6.471 [0.82]	−1.051 [−0.26]
Peripheral diversity	0.043** [2.59]	−2.663*** [−3.53]	1.518** [2.42]	−0.024* [−1.94]	−3.291*** [−4.31]	0.492 [1.47]
Log GDP per capita (1990–2010)	−0.480*** [−16.57]	2.537*** [3.08]	−6.126*** [−6.46]	0.140*** [7.04]	10.754*** [9.09]	2.358*** [5.21]
Absolute latitude	−0.010*** [−2.68]	0.072 [1.04]	−0.449*** [−3.91]	0.007*** [2.65]	0.110 [0.83]	0.198** [2.57]
Terrain roughness	−0.267* [−1.96]	0.143 [0.03]	−14.866*** [−2.65]	0.253 [1.20]	3.959 [0.58]	−3.956* [−1.90]
Latin America and Caribbean	−0.001 [−0.01]	1.922 [0.75]	−8.813** [−2.24]	0.168* [1.68]	−3.757 [−0.81]	1.879 [0.88]
Sub-Saharan Africa	0.493*** [3.61]	−10.663*** [−3.50]	3.580 [0.78]	−0.070 [−0.60]	−25.352*** [−4.88]	3.394 [1.59]
East and Southeast Asia	−0.437*** [−2.82]	3.667 [0.97]	−16.050*** [−3.59]	0.129 [1.44]	−2.985 [−0.50]	−1.809 [−0.73]
Constant	7.393*** [26.18]	67.122*** [9.11]	79.036*** [7.56]	0.592*** [2.71]	−9.663 [−0.83]	−17.073*** [−3.93]
Observations	171	171	140	137	172	172
R-squared	0.8852	0.5457	0.6176	0.6365	0.7804	0.3892

Robust t-statistics in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

attainment—the coefficient is still statistically significant at the 10% level, whereas in the other case—road density—the coefficient is statistically insignificant. As for Greenberg's B-index, its coefficients are statistically insignificant in all cases. These results are in line with our theory: as shown by (10) and (12), a higher degree of peripheral diversity worsens the provision of public goods, whereas a greater level of overall diversity has no effect.

In terms of their economic magnitudes, the estimated coefficients on peripheral diversity are by no means trivial. The standardized β values on peripheral diversity range from 7% in the case of child mortality to −31% in the case of measles immunization. This means that a one standard deviation increase in peripheral diversity is associated with an increase in child mortality by 7% of its standard deviation, and it is associated with a decrease in the measles immunization rate by 31% of its standard deviation. To put these numbers into perspective, in the case of child mortality, a one standard deviation increase in peripheral diversity has about one-tenth of the effect of a one standard deviation decrease in GDP per capita, whereas in the case of measles immunization, a one standard deviation increase in peripheral diversity has a larger effect than a one standard deviation decrease in GDP per capita.

We conduct four types of robustness checks. In a first exercise we examine whether our results change when controlling for legal origin and religious composition. As can be seen in Table 2, our findings are unchanged. The standardized β values on peripheral diversity are slightly smaller: 6% in the case of child mortality and −26% in the case of measles immunization. Socialist legal origin tends to improve public goods, whereas a higher share of Muslims worsens some of the outcomes, such as child mortality and school attainment.

In a second exercise, we explore whether the results depend on the inclusion of particular countries or regions. Using our benchmark specification of Table 1, we exclude, one at a time, sub-Saharan Africa, Latin America and the Caribbean, and East and Southeast Asia. Table 3, Panels A through C, shows that when excluding East and Southeast Asia, the association between peripheral diversity and public goods provision becomes weaker, though it continues to be statistically significant at the 5% level for measles immunization and access to improved sanitation. In the case of measles immunization, the standardized β ranges from −27% when excluding sub-Saharan Africa to −36% when excluding East and Southeast Asia, whereas for access to improved sanitation, the standardized β ranges from −16% when excluding East and Southeast Asia to −29% when excluding sub-Saharan Africa. To further assess the importance of particular observations, we eliminate outliers using the dfbeta method.[13] Table 4 shows the results for the benchmark specification (Panel A) and for the more comprehensive specification (Panel B). In the benchmark specification peripheral diversity continues to be statistically significant at the 5% level with the right sign in three out of the six specifications, whereas in the more comprehensive specification this is true in four out of the six specifications. In the latter case, peripheral diversity only loses significance for school attainment. Once again, the magnitudes of

[13] The dfbeta measures for each variable how influential each observation is. Following Belsley et al. (1980), we drop observations with an absolute value of dfbeta on either Greenberg's B-index or peripheral diversity greater than $2/\sqrt{(\#obs)}$.

Table 2 Public goods and peripheral diversity: broader specification

	(1) Log child mortality	(2) Measles immunization rate	(3) Illiteracy rate	(4) Log school attainment	(5) % Access to sanitation	(6) Roads (km per 1000 population)
Greenberg's B-index	0.098 [0.51]	−6.335 [−1.34]	−1.448 [−0.22]	0.179 [1.46]	0.661 [0.09]	2.563 [0.60]
Peripheral diversity	0.041* [1.93]	−2.220*** [−3.03]	2.036*** [2.70]	−0.033** [−2.25]	−2.137** [−2.45]	−0.252 [−0.62]
Log GDP per capita (1990–2010)	−0.474*** [−12.74]	3.800*** [3.90]	−6.887*** [−7.43]	0.163*** [8.13]	13.661*** [9.92]	1.690*** [4.17]
Absolute latitude	−0.005 [−0.99]	0.001 [0.01]	0.053 [0.40]	−0.000 [−0.04]	0.015 [0.09]	0.126 [1.54]
Terrain roughness	−0.164 [−1.24]	1.683 [0.39]	−13.387*** [−2.94]	0.210 [1.12]	5.462 [0.77]	−3.379 [−1.58]
Latin America and Caribbean	0.183 [1.31]	3.923 [1.34]	−0.021 [−0.01]	0.047 [0.48]	2.616 [0.47]	−1.379 [−0.59]
Sub-Saharan Africa	0.645*** [4.69]	−7.257** [−2.21]	9.575** [2.26]	−0.193* [−1.74]	−16.469*** [−2.80]	0.311 [0.16]
East and Southeast Asia	−0.170 [−0.95]	2.514 [0.60]	−4.208 [−0.78]	−0.109 [−0.92]	−2.958 [−0.39]	−1.187 [−0.45]
Socialist legal origin	0.017 [0.11]	2.939 [0.85]	−17.609*** [−4.01]	0.233*** [2.86]	14.770*** [3.11]	0.231 [0.11]
French legal origin	0.063 [0.82]	−7.043*** [−2.90]	4.250 [1.32]	−0.143*** [−2.84]	−0.398 [−0.12]	−0.820 [−0.62]
German legal origin	−0.128 [−1.10]	−13.565*** [−3.01]		−0.075 [−0.64]	−1.189 [−0.22]	−6.975** [−2.36]
Scandinavian legal origin	−0.551*** [−2.74]	−0.182 [−0.04]		−0.275*** [−2.66]	−0.286 [−0.06]	2.235 [0.21]
Share of protestants	0.006** [2.42]	−0.063 [−1.24]	−0.205** [−2.13]	0.002 [1.22]	−0.084 [−0.99]	0.127* [1.86]
Share of Roman Catholics	0.001 [0.55]	−0.013 [−0.37]	−0.084 [−1.61]	−0.000 [−0.42]	−0.007 [−0.14]	0.021 [1.05]
Share of muslims	0.005*** [3.50]	0.001 [0.03]	0.028 [0.64]	−0.003** [−2.51]	0.092 [1.62]	−0.017 [−0.96]
Constant	6.898*** [20.60]	61.817*** [7.46]	73.975*** [7.58]	0.776*** [3.18]	−37.642*** [−2.62]	−9.986** [−2.52]
Observations	169	169	138	136	170	170
R-squared	0.9022	0.6141	0.7308	0.7827	0.8071	0.5007

Robust t-statistics in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3 Public goods and diversity: excluding different regions

	(1) Log child mortality	(2) Measles immunization rate	(3) Illiteracy rate	(4) Log school attainment	(5) % Access to sanitation	(6) Roads (km per 1000 population)
Panel A: Excluding sub-Saharan Africa[a]						
Greenberg	0.333* [1.73]	1.326 [0.27]	−10.684 [−1.55]	0.223* [1.67]	13.984* [1.71]	−2.220 [−0.47]
Peripheral diversity	0.047** [2.36]	−1.742*** [−3.91]	1.978*** [4.16]	−0.038*** [−3.30]	−3.795*** [−4.04]	0.758* [1.74]
Observations	124	124	95	105	125	125
R-squared	0.8477	0.2990	0.4530	0.5230	0.6133	0.3643
Panel B: Excluding Latin America and Caribbean[b]						
Greenberg	0.052 [0.21]	−3.393 [−0.48]	6.592 [0.78]	0.037 [0.21]	6.157 [0.63]	−1.657 [−0.31]
Peripheral diversity	0.045** [2.48]	−2.781*** [−3.25]	1.280* [1.86]	−0.018 [−1.18]	−3.372*** [−4.02]	0.647 [1.65]
Observations	139	139	113	112	138	140
R-squared	0.8956	0.5481	0.6218	0.6506	0.7907	0.4048
Panel C: Excluding East and Southeast Asia[c]						
Greenberg	0.308 [1.52]	1.773 [0.30]	1.043 [0.12]	0.048 [0.27]	5.205 [0.54]	−0.977 [−0.19]
Peripheral diversity	0.026 [0.86]	−4.126*** [−4.09]	1.815 [1.26]	−0.015 [−0.50]	−3.659** [−2.44]	0.383 [0.58]
Observations	158	158	129	124	160	159
R-squared	0.8875	0.5601	0.6181	0.6257	0.7822	0.4002

Robust t-statistics in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Regression include following additional controls: log GDP per capita (1990–2010), absolute latitude, terrain roughness, Latin America and Caribbean dummy and East and Southeast Asia dummy

[b] Regression include following additional controls: log GDP per capita (1990–2010), absolute latitude, terrain roughness, sub-Saharan Africa dummy and East and Southeast Asia dummy

[c] Regression include following additional controls: log GDP per capita (1990–2010), absolute latitude, terrain roughness, Latin America and Caribbean dummy and sub-Saharan Africa dummy

Table 4 Public goods and diversity: excluding influential observations

	(1) Log child mortality	(2) Measles immunization rate	(3) Illiteracy rate	(4) Log school attainment	(5) % Access to sanitation	(6) Roads (km per 1000 population)
Panel A: Benchmark regression as in Table 1[a]						
Greenberg	0.243 [1.58]	−2.553 [−0.60]	1.433 [0.20]	0.213 [1.54]	8.107 [1.16]	−0.219 [−0.11]
Peripheral diversity	0.030** [2.41]	−3.363*** [−4.60]	0.990 [0.92]	−0.024 [−1.00]	−4.816*** [−3.88]	0.311* [1.69]
Observations	157	160	129	125	155	162
R-squared	0.9191	0.5856	0.6258	0.7009	0.8258	0.4583
Panel B: Broader specification as in Table 2[b]						
Greenberg	−0.025 [−0.16]	−3.583 [−0.97]	−4.381 [−0.62]	−0.008 [−0.07]	1.043 [0.14]	−0.086 [−0.04]
Peripheral diversity	0.077*** [2.80]	−2.767*** [−3.79]	2.931*** [2.83]	0.001 [0.06]	−2.794** [−2.31]	−0.115 [−0.40]
Observations	150	159	124	124	153	157
R-squared	0.9383	0.6315	0.7626	0.8222	0.8463	0.5355

Robust t-statistics in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Excludes observations with an absolute dfbeta value for Greenberg's B-index or for peripheral diversity greater than 2 divided by the square root of the number of observations Regression include following additional controls: log GDP per capita (1990–2010), absolute latitude, terrain roughness, Latin America and Caribbean dummy and East and Southeast Asia dummy

[b] Excludes observations with an absolute dfbeta value for Greenberg's B-index or for peripheral diversity greater than 2 divided by the square root of the number of observations Regression include following additional controls: log GDP per capita (1990–2010), absolute latitude, terrain roughness, sub-Saharan Africa dummy, East and Southeast Asia dummy, Scandinavian legal origin, French legal origin, socialist legal origin, German legal origin, percentage protestant, percentage catholic and percentage muslim

Table 5 Public goods and diversity: Greenberg's B-index and peripheral diversity separately

	(1) Log child mortality	(2) Measles immunization rate	(3) Illiteracy rate	(4) Log school attainment	(5) % Access to sanitation	(6) Roads (km per 1000 population)
Panel A: Only Greenberg's B-index[a]						
Greenberg's B-index	0.410*** [2.68]	−17.192*** [−3.67]	11.002* [1.85]	−0.027 [−0.23]	−11.616* [−1.71]	1.665 [0.58]
Observations	171	171	140	137	172	172
R-squared	0.8831	0.5046	0.6090	0.6319	0.7656	0.3847
Panel B: Only peripheral diversity[a]						
Peripheral diversity	0.057*** [4.05]	−2.876*** [−4.55]	1.708*** [3.13]	−0.016 [−1.42]	−2.763*** [−4.40]	0.407* [1.81]
Observations	171	171	140	137	172	172
R-squared	0.8847	0.5451	0.6172	0.6353	0.7796	0.3889

Robust t-statistics in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Regressions include following additional controls: log GDP per capita (1990–2010), absolute latitude, terrain roughness, Latin America and Caribbean dummy, sub-Saharan Africa dummy and East and Southeast Asia dummy

Table 6 Public goods and diversity: robustness linguistic distances

	(1) Log child mortality	(2) Measles immunization rate	(3) Illiteracy rate	(4) Log school attainment	(5) % Access to sanitation	(6) Roads (km per 1000 population)
Panel A: $\delta = 0.1$[a]						
Greenberg's B-index ($\delta = 0.1$)	0.12 [0.53]	1.947 [0.34]	−10.657 [−1.31]	0.27 [1.64]	14.27 [1.54]	−0.633 [−0.13]
Peripheral diversity ($\delta = 0.1$)	0.059** [2.42]	−3.190*** [−4.20]	2.569*** [2.75]	−0.041** [−2.40]	−3.870*** [−3.94]	0.552 [1.21]
Observations	171	171	140	137	172	172
R-squared	0.885	0.5235	0.6142	0.6408	0.7748	0.389
Panel B: $\delta = 0.9$[a]						
Greenberg's B-index ($\delta = 0.9$)	0.196 [1.12]	−3.585 [−0.73]	5.984 [0.95]	0.045 [0.34]	2.828 [0.39]	−1.267 [−0.35]
Peripheral diversity ($\delta = 0.9$)	0.035** [2.42]	−2.340*** [−3.46]	1.065* [1.83]	−0.017 [−1.47]	−2.888*** [−3.99]	0.45 [1.55]
Observations	171	171	140	137	172	172
R-squared	0.8852	0.5492	0.6186	0.6353	0.7827	0.389

Robust t-statistics in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Regression include following additional controls: log GDP per capita (1990–2010), absolute latitude, terrain roughness, Latin America and Caribbean dummy, sub-Saharan Africa dummy and East and Southeast Asia dummy

the effects do not change much. For example, in the case of measles immunization the standardized β is between -26% (Panel B) and -32% (Panel A).

In a third exercise, we include Greenberg's B-index and peripheral diversity one at a time, rather than jointly. Table 5 reports the results. When only including Greenberg's B-index (Panel A), it is associated with worse public goods outcomes. Two out of the six outcomes yield coefficients that are statistically significant at the 1% level, two other outcomes give coefficients that are statistically significant at the 10% level, and the remaining two are not significant. When only including peripheral diversity (Panel B), it also tends to be associated with worse public goods outcomes. Four out of the six outcomes give coefficients that are statistically significant at the 1% level, one other at the 10% level, and the remaining one is not significant. It is not surprising that when including Greenberg's B-index and PD separately, both tend to be negatively associated with public goods. After all, Greenberg's B-index and PD are positively correlated. Note, furthermore, that the R^2 values tend to be slightly higher for the regressions that include PD than for those that include Greenberg's B-index. Consistent with this, when both are jointly included, as in Table 1, PD trumps Greenberg's B-index.

In a fourth exercise, we examine the importance of linguistic distances for our results. Recall that lower values of δ imply a linguistic distance function that is highly convex, meaning that the linguistic distance between two languages only becomes important when they are very far apart in the language tree. In contrast, higher values of δ imply that the linguistic distance between two languages already becomes important when they are still quite close in the language tree. Table 6 reports results for $\delta = 0.1$ (Panel A) and $\delta = 0.9$ (Panel B). The results do not seem to depend much on the value of δ: the lower value of δ yields coefficients on PD that are statistically slightly more significant than the higher value of δ, but the difference is small. In terms of their economic significance, we confirm our previous results. For $\delta = 0.1$, the standardized β for child mortality is 7% and for measles immunization is -29%; the corresponding figures for $\delta = 0.9$ are 6 and -31%. These numbers are virtually identical to those we found in our benchmark analysis of Table 1.

3.4 Transfers, peripheral diversity and Greenberg's B-index

We now turn to analyzing the relation between peripheral diversity, Greenberg's B-index and transfers. Our dependent variable is transfers and subsidies as a share of GDP. As in the case of public goods, the data cover the period 1990–2010. In addition to peripheral diversity and Greenberg's B-index, we also control for other possible determinants of the level of transfers. Table 7 reports our findings for different sets of covariates. In column (1), where we control for GDP per capita, absolute latitude, roughness of terrain and regional dummies, we find that transfers tend to go down when Greenberg's B-index increases. In contrast, transfers tend to increase when peripheral diversity is higher. Both coefficients are statistically significant at the 1% level. These findings are consistent with our theoretical results (10) and (12). The rest of the columns of Table 7 analyze the robustness of our findings by including different controls, such as population size, legal origin, religious composition, and share of the population 65 years and older. Our results are unchanged.

Table 7 Transfers, Greenberg's B-index and peripheral diversity

Dependent variable: transfers 1990–2010	(1)	(2)	(3)	(4)	(5)	(6)
Greenberg's B-index	−9.769*** [−4.36]	−8.742*** [−3.70]	−8.657*** [−3.65]	−5.810*** [−2.75]	−5.696** [−2.55]	−4.880** [−2.29]
Peripheral diversity	0.625*** [3.87]	0.510*** [2.85]	0.591*** [2.94]	0.461** [2.08]	0.499* [1.93]	0.554** [2.15]
Log GDP per capita (1990–2010)	1.645*** [5.30]	1.817*** [5.81]	2.473*** [6.85]	1.120*** [4.48]	1.503*** [5.07]	1.279*** [4.81]
Absolute latitude	0.229*** [5.89]	0.227*** [5.98]	0.152*** [3.05]	0.074* [1.78]	0.043 [0.92]	0.072 [1.53]
Terrain roughness	−1.373 [−0.52]	−0.773 [−0.31]	−0.730 [−0.28]	−4.850** [−2.26]	−4.643** [−2.13]	−6.718*** [−3.33]
Latin America and Caribbean	−2.015* [−1.70]	−1.761 [−1.50]	−1.786 [−1.43]	−3.168*** [−3.32]	−3.275*** [−3.56]	−5.069*** [−4.91]
Sub-Saharan Africa	0.384 [0.29]	0.774 [0.61]	1.520 [1.24]	−0.087 [−0.10]	0.267 [0.31]	−0.396 [−0.40]
East and Southeast Asia	−2.118 [−1.57]	−2.646* [−1.82]	−3.838** [−2.28]	−4.209*** [−3.64]	−4.653*** [−3.43]	−4.685*** [−3.78]
Log population (1990–2010)		0.476* [1.71]	0.633** [2.37]	0.257 [1.11]	0.330 [1.50]	0.179 [0.93]
Log population >65 years (1990–2010)				6.067*** [8.22]	5.720*** [8.31]	4.772*** [7.22]
Socialist legal origin			3.399 [1.09]		1.633 [0.64]	−3.661 [−1.24]
French legal origin			−0.298 [−0.09]		0.075 [0.03]	−5.635* [−1.73]
German legal origin			0.625 [0.15]		0.377 [0.11]	−4.117 [−1.32]
British legal origin			−1.493 [−0.49]		−1.275 [−0.50]	−5.668* [−1.93]
Share of protestants						−0.058** [−2.59]
Share of Roman Catholics						0.035*** [2.97]
Share of muslims						−0.019* [−1.71]
Constant	−8.261*** [−2.82]	−17.601*** [−3.10]	−23.715*** [−3.14]	−14.562*** [−3.15]	−17.520*** [−3.04]	−6.857 [−1.17]
Observations	131	131	131	131	131	130
R-squared	0.7318	0.7396	0.7654	0.8300	0.8403	0.8732

Robust t-statistics in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8 Transfers and diversity: different regions and influential observations

Transfers 1990–2010	(1) No sub-Saharan Africa	(2) No Latin America or Caribbean	(3) No East and Southeast Asia	(4) No sub-Saharan Africa	(5) No Latin America or Caribbean	(6) No East and Southeast Asia
Panel A: Different regions[a]						
Greenberg	−9.912*** [−3.73]	−11.046*** [−4.17]	−10.444*** [−3.73]	−4.427* [−1.99]	−5.265** [−2.08]	−1.146 [−0.48]
Peripheral diversity	0.649*** [3.57]	0.709*** [3.80]	0.643** [2.06]	0.966*** [5.05]	0.605** [2.44]	−0.339 [−1.17]
Log population >65	No	No	No	Yes	Yes	Yes
Legal origin	No	No	No	Yes	Yes	Yes
Religious composition	No	No	No	Yes	Yes	Yes
Observations	100	107	121	99	106	120
R-squared	0.6841	0.7414	0.7285	0.8830	0.8772	0.8843

Table 8 continued

Transfers 1990–2010	(1) Drop influential observations	(2) Drop influential observations
Panel B: Excluding influential observations[b]		
Greenberg	−10.199*** [−4.63]	−3.727* [−1.98]
Peripheral diversity	0.754*** [2.72]	0.157 [0.62]
Log population >65	No	Yes
Legal origin	No	Yes
Religious composition	No	Yes
Observations	122	117
R-squared	0.7937	0.9156

Robust t-statistics in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Regression include following controls: log GDP per capita (1990–2000), absolute latitude, terrain roughness, sub-Saharan Africa dummy, East and Southeast Asia dummy, and log population (1990–2010). In addition, columns (4), (5) and (6) also include log population above 65 years (1990–2010), socialist legal origin, French legal origin, German legal origin, British legal origin, percentage protestant, percentage catholic and percentage muslim.

[b] Excludes observations with an absolute dfbeta value for Greenberg's B-index or for peripheral diversity greater than 2 divided by the square root of the number of observations. Regression include following controls: log GDP per capita (1990–2000), absolute latitude, terrain roughness, sub-Saharan Africa dummy, East and Southeast Asia dummy, and log population (1990–2010). In addition, column (2) also includes log population above 65 years (1990–2010), socialist legal origin, French legal origin, German legal origin, British legal origin, percentage protestant, percentage catholic and percentage muslim

Table 9 Transfers and diversity: further robustness

Dependent variable: transfers 1990–2010	(1)	(2)	(3)	(4)
Panel A: Greenberg's B-index and peripheral diversity separately[a]				
Greenberg	−5.815*** [−3.38]	−1.652 [−1.19]		
Peripheral diversity			−0.179 [−0.67]	0.137 [0.58]
Log population >65 years	No	Yes	No	Yes
Legal origin	No	Yes	No	Yes
Religious composition	No	Yes	No	Yes
Observations	131	130	131	130
R-squared	0.7328	0.8669	0.7120	0.8660

Table 9 continued

Transfers 1990–2010	(1)	(2)	(3)	(4)
	$\delta = 0.1$	$\delta = 0.1$	$\delta = 0.9$	$\delta = 0.9$
Panel B: Different values of δ^b				
Greenberg	−9.700*** [−3.80]	−5.298** [−2.12]	−7.714*** [−3.51]	−4.250** [−2.22]
Peripheral diversity	0.567** [2.36]	0.585 [1.64]	0.450*** [2.80]	0.489** [2.15]
Log population > 65 years	No	Yes	No	Yes
Legal origin	No	Yes	No	Yes
Religious composition	No	Yes	No	Yes
Observations	131	130	131	130
R-squared	0.7412	0.8720	0.7369	0.8727

Robust t-statistics in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[a] Regression include following controls: log GDP per capita (1990–2000), absolute latitude, terrain roughness, sub-Saharan Africa dummy, East and Southeast Asia dummy, and log population (1990–2010). In addition, columns (2) and (4) also include log population above 65 years (1990–2010), socialist legal origin, French legal origin, German legal origin, British legal origin, percentage protestant, percentage catholic and percentage muslim

[b] Regression include following controls: log GDP per capita (1990–2000), absolute latitude, terrain roughness, sub-Saharan Africa dummy, East and Southeast Asia dummy, and log population (1990–2010). In addition, columns (2) and (4) also include log population above 65 years (1990–2010), socialist legal origin, French legal origin, German legal origin, British legal origin, percentage protestant, percentage catholic and percentage muslim

The magnitudes of the effects are economically meaningful. Focusing on column (1), the standardized β on peripheral diversity is 16%, and the standardized β on Greenberg's B-index is -28%. This means that a one standard deviation increase in peripheral diversity raises transfers by 16% of its standard deviation, whereas a one standard deviation increase in Greenberg's B-index lowers transfers by 28% of its standard deviation. To provide a benchmark for these figures, a one standard increase in GDP per capita raises transfers by 35% of its standard deviation. Hence, the roles of peripheral diversity and Greenberg's B-index are quantitatively relevant.

We conduct three further robustness checks. First, we analyze the importance of particular regions or observations for our results. Table 8, Panel A, shows the results when we drop different parts of the world. Focusing on the benchmark specification, reported in columns (1) through (3), the standardized β on Greenberg's B-index ranges from -27% when excluding sub-Saharan Africa to -29% when excluding East and Southeast Asia, whereas the standardized β on peripheral diversity ranges from 12% when excluding East and Southeast Asia to 16% when excluding sub-Saharan Africa. Table 8, Panel B, explores the effect of influential observations based on the same dfbeta method as before. In the benchmark regression, reported in column (1), the results hardly change. However, in the more comprehensive specification, reported in column (2), the results are weaker in that peripheral diversity is no longer statistically significant.

Second, we include Greenberg's B-index and peripheral diversity as two separate regressors. For each case, we rerun the regressions of column (1) and column (6) of Table 7. As can be observed in Table 9, Panel A, the association between our two diversity indices and transfers is weaker, and most often statistically insignificant. This is not surprising: since Greenberg's B-index and peripheral diversity are positively correlated, but have opposite effects when both included, their effects when introduced separately are ambiguous. This underscores the importance of jointly controlling for Greenberg's B-index and peripheral diversity. Not doing so would amount to omitted variable bias, since the theory predicts that both indices matter. Third, we analyze how our results change when we take different values of δ to compute linguistic distances. Table 9, Panel B, reports the results. Our findings are largely unchanged, though somewhat weaker for low values of δ.

4 Concluding remarks

In this paper we have proposed a theory that analyzes the relation between diversity, public goods and transfers. Following the existing literature, the theory assumes that people are less willing to pay taxes in countries with high degrees of overall diversity. In addition to this standard argument, we have advanced the hypothesis that in countries that suffer from greater antagonism between the center and the periphery it is harder to reach a political agreement on public goods than on transfers. To distinguish between these two arguments, we have defined two types of diversity: a country's overall diversity which captures the tension between all individuals and affects the willingness to pay taxes; and a country's peripheral diversity which captures the tension between the center and the periphery and affects the cost of providing public goods.

Our simple theory has yielded four predictions: greater peripheral diversity lowers public goods provision but increases transfers, whereas greater overall diversity has no effect on public goods provision but lowers transfers. Our empirical analysis has provided evidence in support of these theoretical predictions. An important conclusion is that public goods and transfers are substitutes in their relation to a change in peripheral diversity.

Appendix A: Proof of Proposition 1

To simplify the notation we drop 0 from the subscripts and write τ_j instead of τ_{0j}.

(1) First consider the case $\tau_k = \tau_l = \tau$. We have to show that $s_k^* \geq s_l^*$. Suppose, to the contrary, that $s_k^* < s_l^*$. Let $s' \in \mathcal{S}_{kl}(\bar{s})$ 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 $PA(s', T) < PA(s^*, T)$, it follows that

$$2 f_{pc}(s_0, x, \tau) + 2 f_{cp}(s_0, x, \tau) < f_{pc}(s_0, s_k^*, \tau) + f_{pc}(s_0, s_l^*, \tau) + f_{cp}(s_0, s_k^*, \tau) + f_{cp}(s_0, s_l^*, \tau). \quad (14)$$

By Condition 3, functions $f_{pc}(s_0, ., \tau)$ and $f_{cp}(s_0, ., \tau)$ are concave, which implies

$$f_{pc}(s_0, x, \tau) \geq \frac{1}{2} f_{pc}(s_0, s_k^*, \tau) + \frac{1}{2} f_{pc}(s_0, s_l^*, \tau) \quad (15)$$

and

$$f_{cp}(s_0, x, \tau) \geq \frac{1}{2} f_{cp}(s_0, s_k^*, \tau) + \frac{1}{2} f_{cp}(s_0, s_l^*, \tau). \quad (16)$$

It is straightforward to verify that inequalities (14)–(16) can not hold simultaneously. Thus, we have that $s_k^* \geq s_l^*$. Notice that $\tau_l = \tau_k$ implies $s_l^* \geq s_k^*$ and $s_k^* \geq s_l^*$ so that $s_k^* = s_l^*$.

(2) Now consider the case $\tau_k > \tau_l$. 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 $\tau'_j = \tau_j$ for all $j \neq l$ and $\tau'_l = \tau_k$. Notice that $\tau'_j > \tau_l$. Similarly to the previous examination, let $s' \in \mathcal{S}_{kl}(\bar{s})$ 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

$$PA(s', T) < PA(s^*, T). \quad (17)$$

This implies that

$$f_{pc}(s_0, x, \tau_k) + f_{pc}(s_0, x, \tau_l) + f_{cp}(s_0, x, \tau_k) + f_{cp}(s_0, x, \tau_l) < f_{pc}(s_0, s_k^*, \tau_k) + f_{pc}(s_0, s_l^*, \tau_l) + f_{cp}(s_0, s_k^*, \tau_k) + f_{cp}(s_0, s_l^*, \tau_l), \quad (18)$$

which is equivalent to

$$\begin{aligned} & f_{pc}(s_0, x, \tau_k) - f_{pc}(s_0, s_k^*, \tau_k) + f_{cp}(s_0, x, \tau_k) - f_{cp}(s_0, s_k^*, \tau_k) \\ & \quad < f_{pc}(s_0, s_l^*, \tau_l) - f_{pc}(s_0, x, \tau_l) + f_{cp}(s_0, s_l^*, \tau_l) - f_{cp}(s_0, x, \tau_l). \end{aligned} \tag{19}$$

The argument used in case 1 above yields

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

which implies that

$$\begin{aligned} & f_{pc}(s_0, x, \tau_k) + f_{pc}(s_0, x, \tau_k) + f_{cp}(s_0, x, \tau_k) + f_{cp}(s_0, x, \tau_k) \\ & \quad > f_{pc}(s_0, s_k^*, \tau_k) + f_{pc}(s_0, s_l^*, \tau_k) + f_{cp}(s_0, s_k^*, \tau_k) + f_{cp}(s_0, s_l^*, \tau_k). \end{aligned} \tag{20}$$

By rearranging the terms we obtain

$$\begin{aligned} & f_{pc}(s_0, x, \tau_k) - f_{pc}(s_0, s_k^*, \tau_k) + f_{cp}(s_0, x, \tau_k) - f_{cp}(s_0, s_k^*, \tau_k) \\ & \quad > f_{pc}(s_0, s_l^*, \tau_k) - f_{pc}(s_0, x, \tau_k) + f_{cp}(s_0, s_l^*, \tau_k) - f_{cp}(s_0, x, \tau_k). \end{aligned} \tag{21}$$

Inequalities (19) and (21) imply

$$\begin{aligned} & f_{pc}(s_0, s_l^*, \tau_k) - f_{pc}(s_0, x, \tau_k) + f_{cp}(s_0, s_l^*, \tau_k) - f_{cp}(s_0, x, \tau_k) \\ & \quad < f_{pc}(s_0, s_l^*, \tau_l) - f_{pc}(s_0, x, \tau_l) + f_{cp}(s_0, s_l^*, \tau_l) - f_{cp}(s_0, x, \tau_l). \end{aligned} \tag{22}$$

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

$$\begin{aligned} & f_{pc}(s_0, s_l^*, \tau_k) - f_{pc}(s_0, x, \tau_k) > f_{pc}(s_0, s_l^*, \tau_l) - f_{pc}(s_0, x, \tau_l) \quad \text{and} \\ & f_{cp}(s_0, s_l^*, \tau_k) - f_{cp}(s_0, x, \tau_k) > f_{cp}(s_0, s_l^*, \tau_l) - f_{cp}(s_0, x, \tau_l), \end{aligned} \tag{23}$$

and (23) and (21) do not hold simultaneously. Hence we conclude that $s_k^* \geq s_l^*$.

Appendix B: Data sources and diversity indices

Absolute latitude The absolute value of the latitude of a country's approximate geodesic centroid, as reported by the CIA World Factbook. *Source:* Ashraf and Galor (2013).
Child mortality Log of child mortality rate per 1000 live births, 1990–2010 average. *Source: World Development Indicators, World Bank.*
GDP per capita GDP per capita, constant 2005 US$, 1990–2010 average. *Source: World Development Indicators, World Bank.*
Illiteracy Percentage of people aged 15 and above who are illiterate, 1990–2010 average. *Source: World Development Indicators, World Bank.*
Improved sanitation Percentage of population with access to improved sanitation facilities, 1990–2010 average. *Source: World Development Indicators, World Bank.*

Legal origin Socialist, French, German or British legal origin *Source:* La Porta et al. (2008).
Language data Languages spoken in each country and language trees. *Source: Ethnologue: Languages of the World, 15th Edition, SIL International, 2005.*
Major religions Share of protestants, catholics and muslims in the population. *Source:* La Porta et al. (1999).
Measles immunization Percentage of children between the age of 12 and 23 months that have been immunized against measles, 1990–2010 average. *Source: World Development Indicators, World Bank.*
Population. Total population, 1990–2010 average. *Source: World Development Indicators, World Bank.*
Population above 65 Population ages 65 and above, % of total, 1990–2010 average. *Source: World Development Indicators, World Bank.*
School attainment. Log of 1 + average years of schooling of population aged 25 or above, 1990–2010 average. *Source: Barro R. and J.W. Lee v. 1.3, 04/13.*
Road density Road network density, km per 1000 inhabitants, 2001–2010 average. *Source: World Development Indicators, World Bank.*
Terrain roughness The degree of terrain roughness of a country, calculated using geospatial surface undulation data reported by the G-ECON project (Nordhaus 2006) at a 1-degree resolution. *Source:* Ashraf and Galor (2013).
Transfers Transfers and subsidies as percent of GDP: Average for 1990, 1995, 2000, 2005 and 2010. *Source:* Gwartney et al. (2012), *Economic Freedom Dataset, Fraser Institute.*

Table 10 Indices of linguistic diversity: Greenberg's B-index and peripheral diversity

Country	Greenberg B-index	Peripheral diversity	Ranking B-index	Ranking peripheral	Difference
Afghanistan	0.5025	1.9543	46	56	−10
Albania	0.1746	0.6157	135	151	−16
Algeria	0.2266	0.8056	122	125	−3
American Samoa	0.0929	0.3736	165	165	0
Andorra	0.1951	0.5329	130	153	−23
Angola	0.2136	1.1049	126	100	26
Anguilla	0.1405	0.3488	150	170	−20
Antigua and Barbuda	0.0566	0.2570	179	184	−5
Argentina	0.1706	0.9401	137	113	24
Armenia	0.1513	0.6285	146	146	0
Aruba	0.3774	1.0115	76	106	−30
Australia	0.0972	1.1795	162	94	68
Austria	0.2430	0.8443	114	121	−7
Azerbaijan	0.3643	1.5682	82	73	9
Bahamas	0.3593	0.8462	84	120	−36

Table 10 continued

Country	Greenberg B-index	Peripheral diversity	Ranking B-index	Ranking peripheral	Difference
Bahrain	0.5457	1.7588	37	64	−27
Bangladesh	0.1525	0.7960	145	127	18
Barbados	0.0910	0.2653	167	182	−15
Belarus	0.2374	0.7477	119	133	−14
Belgium	0.4798	1.8378	54	61	−7
Belize	0.6723	1.9368	6	57	−51
Benin	0.4567	2.7947	59	27	32
Bermuda	0.0000	0.0000	219	219	0
Bhutan	0.6000	3.3685	22	18	4
Bolivia	0.6685	2.0788	7	51	−44
Bosnia and Herzegovina	0.2467	0.7221	113	137	−24
Botswana	0.1528	1.0596	144	103	41
Brazil	0.0243	0.5203	199	155	44
British Indian Ocean Terr.	0.0000	0.0000	220	220	0
British Virgin Islands	0.1671	0.3911	139	162	−23
Brunei	0.3734	1.5568	78	74	4
Bulgaria	0.2092	0.7409	127	134	−7
Burkina Faso	0.4364	2.3456	62	42	20
Burundi	0.0018	0.0353	215	214	1
Cambodia	0.1307	0.6963	155	142	13
Cameroon	0.4984	5.4310	48	7	41
Canada	0.4129	2.4837	72	36	36
Cape Verde Islands	0.0699	0.2260	171	188	−17
Cayman Islands	0.5350	1.2683	42	89	−47
Central African Republic	0.5984	6.7938	23	3	20
Chad	0.8035	7.2055	1	2	−1
Chile	0.0326	0.2097	194	189	5
China	0.3379	2.7538	90	30	60
Colombia	0.0288	0.6969	195	141	54
Comoros	0.1101	0.3554	159	168	−9
Congo	0.6050	5.1178	20	8	12
Cook Islands	0.0912	0.3058	166	175	−9
Costa Rica	0.0495	0.3572	185	166	19
Cote dIvoire	0.5350	3.4780	41	17	24
Croatia	0.0621	0.2895	176	178	−2
Cuba	0.0002	0.0060	217	218	−1
Cyprus	0.3643	0.7762	81	131	−50

Table 10 continued

Country	Greenberg B-index	Peripheral diversity	Ranking B-index	Ranking peripheral	Difference
Czech Republic	0.0477	0.2860	188	179	9
DRC	0.5497	3.9324	35	12	23
Denmark	0.0368	0.3209	192	173	19
Djibouti	0.3766	0.9986	77	107	−30
Dominica	0.3116	0.6535	98	144	−46
Dominican Republic	0.0528	0.2924	182	177	5
East Timor	0.6572	2.0903	8	49	−41
Ecuador	0.2559	1.1454	111	97	14
Egypt	0.2286	0.8285	121	124	−3
El Salvador	0.0043	0.0588	211	210	1
Equatorial Guinea	0.1842	0.7376	131	135	−4
Eritrea	0.5009	1.7238	47	65	−18
Estonia	0.4676	1.3485	56	84	−28
Ethiopia	0.5678	3.3114	30	19	11
Falkland Islands	0.0000	0.0000	222	222	0
Fiji	0.5294	1.8712	44	60	−16
Finland	0.1323	0.6435	154	145	9
France	0.1841	1.3533	132	83	49
French Guiana	0.4271	1.6219	68	70	−2
French Polynesia	0.3984	0.9369	73	115	−42
Gabon	0.3043	1.3563	100	82	18
Gambia	0.4917	1.7987	52	63	−11
Georgia	0.5386	2.2305	39	44	−5
Germany	0.1326	1.0805	152	102	50
Ghana	0.3687	2.2301	79	45	34
Gibraltar	0.4979	1.0252	49	104	−55
Greece	0.1297	0.7962	156	126	30
Greenland	0.2419	0.5056	116	156	−40
Grenada	0.0519	0.2334	183	185	−2
Guadeloupe	0.0653	0.2852	175	180	−5
Guam	0.5666	1.3898	31	81	−50
Guatemala	0.6101	3.9733	18	11	7
Guinea	0.4802	2.0787	53	52	1
Guinea-Bissau	0.5659	1.9105	32	58	−26
Guyana	0.0778	0.5553	169	152	17
Haiti	0.0002	0.0094	218	217	1
Honduras	0.0528	0.3528	181	169	12
Hungary	0.1564	0.8722	141	119	22
Iceland	0.0106	0.0597	206	209	−3

Table 10 continued

Country	Greenberg B-index	Peripheral diversity	Ranking B-index	Ranking peripheral	Difference
India	0.6788	5.5610	5	6	−1
Indonesia	0.5504	6.0363	34	5	29
Iran	0.6390	3.7850	12	15	−3
Iraq	0.4612	1.6765	57	68	−11
Ireland	0.1510	0.4314	147	159	−12
Israel	0.5316	2.7665	43	29	14
Italy	0.2775	1.2882	105	87	18
Jamaica	0.0111	0.1307	205	200	5
Japan	0.0240	0.2035	200	190	10
Jordan	0.2203	0.7958	124	128	−4
Kazakhstan	0.6297	2.5050	16	35	−19
Kenya	0.5790	2.8531	28	25	3
Kiribati	0.0225	0.1361	201	199	2
Korea, North	0.0000	0.0000	223	223	0
Korea, South	0.0030	0.0404	212	211	1
Kuwait	0.2425	0.6266	115	147	−32
Kyrgyzstan	0.5923	2.1173	25	48	−23
Laos	0.5470	3.8126	36	14	22
Latvia	0.4290	1.3445	67	85	−18
Lebanon	0.1532	0.6160	143	150	−7
Lesotho	0.0584	0.1622	178	196	−18
Liberia	0.6031	3.0010	21	23	−2
Libya	0.1809	0.6260	133	148	−15
Liechtenstein	0.0658	0.2274	174	187	−13
Lithuania	0.2491	0.7916	112	129	−17
Luxembourg	0.3494	0.9708	88	108	−20
Macedonia	0.4608	1.1860	58	93	−35
Madagascar	0.2868	1.0950	103	101	2
Malawi	0.1703	0.7008	138	140	−2
Malaysia	0.6476	3.5729	11	16	−5
Maldives	0.0047	0.0363	210	213	−3
Mali	0.6354	3.2920	14	21	−7
Malta	0.0157	0.0970	203	206	−3
Marshall Islands	0.0266	0.1295	196	201	−5
Martinique	0.0427	0.1944	190	191	−1
Mauritania	0.1713	0.6564	136	143	−7
Mauritius	0.6080	2.0883	19	50	−31
Mayotte	0.4372	0.9453	61	111	−50

Table 10 continued

Country	Greenberg B-index	Peripheral diversity	Ranking B-index	Ranking peripheral	Difference
India	0.6788	5.5610	5	6	−1
Indonesia	0.5504	6.0363	34	5	29
Iran	0.6390	3.7850	12	15	−3
Iraq	0.4612	1.6765	57	68	−11
Ireland	0.1510	0.4314	147	159	−12
Israel	0.5316	2.7665	43	29	14
Italy	0.2775	1.2882	105	87	18
Jamaica	0.0111	0.1307	205	200	5
Japan	0.0240	0.2035	200	190	10
Jordan	0.2203	0.7958	124	128	−4
Kazakhstan	0.6297	2.5050	16	35	−19
Kenya	0.5790	2.8531	28	25	3
Kiribati	0.0225	0.1361	201	199	2
Korea, North	0.0000	0.0000	223	223	0
Korea, South	0.0030	0.0404	212	211	1
Kuwait	0.2425	0.6266	115	147	−32
Kyrgyzstan	0.5923	2.1173	25	48	−23
Laos	0.5470	3.8126	36	14	22
Latvia	0.4290	1.3445	67	85	−18
Lebanon	0.1532	0.6160	143	150	−7
Lesotho	0.0584	0.1622	178	196	−18
Liberia	0.6031	3.0010	21	23	−2
Libya	0.1809	0.6260	133	148	−15
Liechtenstein	0.0658	0.2274	174	187	−13
Palau	0.0491	0.1741	186	195	−9
West Bank and Gaza	0.1012	0.2787	160	181	−21
Panama	0.3233	1.1163	94	98	−4
Papua New Guinea	0.7966	16.1864	2	1	1
Paraguay	0.3352	1.4742	91	78	13
Peru	0.3664	2.5423	80	34	46
Philippines	0.4720	3.3111	55	20	35
Pitcairn	0.0000	0.0000	226	226	0
Poland	0.0388	0.3195	191	174	17
Portugal	0.0137	0.1371	204	198	6
Puerto Rico	0.0356	0.1851	193	193	0
Qatar	0.5825	1.5415	27	75	−48
Reunion	0.0660	0.3412	173	172	1
Romania	0.1574	0.7065	140	139	1
Russia	0.2411	2.4452	117	38	79
Rwanda	0.0017	0.0373	216	212	4

Table 10 continued

Country	Greenberg B-index	Peripheral diversity	Ranking B-index	Ranking peripheral	Difference
St Helena	0.0000	0.0000	221	221	0
St Kitts and Nevis	0.0102	0.0765	207	208	−1
St Lucia	0.0198	0.1100	202	205	−3
St Pierre and Miquelon	0.1172	0.3873	157	163	−6
St Vincent and the Grenadines	0.0086	0.0960	209	207	2
Samoa	0.0020	0.0327	214	215	−1
San Marino	0.2390	0.4834	118	157	−39
Sao Tome e Principe	0.3575	0.9220	85	117	−32
Saudi Arabia	0.3554	1.3938	86	80	6
Senegal	0.4169	1.8800	70	59	11
Serbia and Montenegro	0.2850	0.9635	104	110	−6
Seychelles	0.0666	0.2934	172	176	−4
Sierra Leone	0.5875	2.1721	26	47	−21
Singapore	0.6538	2.4520	9	37	−28
Slovakia	0.2674	0.8299	107	123	−16
Slovenia	0.0929	0.4002	164	161	3
Solomon Islands	0.4970	2.8387	50	26	24
Somalia	0.0944	0.4519	163	158	5
South Africa	0.4952	1.6125	51	71	−20
Spain	0.1547	0.5279	142	154	−12
Sri Lanka	0.3109	0.7219	99	138	−39
Sudan	0.5414	4.8948	38	9	29
Suriname	0.7207	2.3556	3	39	−36
Swaziland	0.0439	0.1175	189	204	−15
Sweden	0.1345	0.9342	151	116	35
Switzerland	0.3870	1.4799	75	77	−2
Syria	0.3221	1.1149	95	99	−4
Taiwan	0.3215	1.0187	96	105	−9
Tajikistan	0.4355	1.6329	63	69	−6
Tanzania	0.3425	2.7305	89	31	58
Thailand	0.4153	2.0042	71	54	17
Togo	0.4312	2.3459	65	41	24
Tokelau	0.0538	0.1937	180	192	−12
Tonga	0.0028	0.0225	213	216	−3
Trinidad and Tobago	0.5972	1.6849	24	67	−43
Tunisia	0.0095	0.1397	208	197	11

Table 10 continued

Country	Greenberg B-index	Peripheral diversity	Ranking B-index	Ranking peripheral	Difference
Turkey	0.2729	1.4994	106	76	30
Turkmenistan	0.3186	1.4721	97	79	18
Turks and Caicos Islands	0.1455	0.3568	149	167	−18
Tuvalu	0.0512	0.1276	184	202	−18
U.S. Virgin Islands	0.3329	0.8390	93	122	−29
Uganda	0.5742	2.8565	29	24	5
Ukraine	0.2968	1.2675	101	90	11
United Arab Emirates	0.6971	2.6813	4	32	−28
United Kingdom	0.1104	1.2293	158	91	67
Uruguay	0.0480	0.2640	187	183	4
USA	0.2631	2.2270	109	46	63
Uzbekistan	0.3616	1.7139	83	66	17
Vanuatu	0.4215	2.5879	69	33	36
Vatican State	0.0000	0.0000	224	224	0
Venezuela	0.0253	0.4011	198	160	38
Viet Nam	0.2023	1.8319	129	62	67
Wallis and Futuna	0.0589	0.1815	177	194	−17
Yemen	0.2907	0.8854	102	118	−16
Zambia	0.2254	1.1548	123	96	27
Zimbabwe	0.1501	0.6180	148	149	−1

References

Alesina A, Reich B (2015) Nation-building. Harvard University, Cambridge

Alesina A, Devleeschauwer A, Easterly W, Kurlat S, Wacziarg R (2003) Fractionalization. J Econ Growth 8:155–194

Ashraf Q, Galor O (2013) The ‘Out of Africa’ Hypothesis, Human Genetic Diversity, and Comparative Economic Development. Am Econ Rev 103:1–46

Belsley DA, Kuh E, Welsch RE (1980) Regression diagnostics. Wiley, New York

Desmet K, Ortuño-Ortín I, Weber S (2005) Peripheral diversity and redistribution. CEPR Discussion Paper 5112

Desmet K, Ortuño-Ortín I, Weber S (2009) Linguistic diversity and redistribution. J Eur Econ Assoc 7:1291–1318

Desmet K, Ortuño-Ortín I, Wacziarg R (2012) The political economy of linguistic cleavages. J Dev Econ 97:322–338

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

Fearon JD (2003) Ethnic and cultural diversity by country. J Econ Growth 8:195–222

Gordon RG Jr (2005) Ethnologue: languages of the world, 15th edn. SIL International, Dallas

Gray RD, Atkinson QD (2003) Language-tree divergence times support the anatolian theory of Indo-European origin. Nature 426:435–439

Greenberg JH (1956) The measurement of linguistic diversity. Language 32:109–115

Gwartney J, Lawson R, Hall J (2012) Economic freedom of the world: 2012 annual report. Fraser Institute, Vancouver

La Porta R, Lopez de Silanes F, Shleifer A, Vishny R (1999) The quality of government. J Law Econ Organ 15:222–279

La Porta R, Lopez de Silanes F, Shleifer A (2008) The economic consequences of legal origins. J Econ Lit 46:285–332
Reynal-Querol M (2002) Ethnicity, political systems, and civil wars. J Confl Resolut 46:29–54
Shannon CE (1949) The mathematical theory of communication. University of Illinois Press, Urbana
Weber E (1979) Peasants into Frenchmen: modernization of rural France, 1870–1914. Chatto and Windus, London