Economic versus cultural differences: Forms of ethnic diversity and public goods provision*

Kate Baldwin†
John D. Huber‡

August 25, 2010

Abstract

Arguments about how ethnic diversity affects governance typically posit that groups differ from each other in substantively important ways, and that these differences make effective governance more difficult. But existing cross-national empirical tests typically use measures of ethnolinguistic fractionalization that have no information about substantive differences between groups. This paper examines two important ways that groups differ from each other – culturally and economically – and assesses how such differences affect public goods provision. Across 46 countries, the analysis compares existing measures of cultural differences with a new measure that captures economic differences between groups: between-group inequality. We show that ethnolinguistic fractionalization, cultural fractionalization, and between-group inequality measure different things, and that the choice between them has an important impact on our understanding of which countries are most ethnically diverse. Furthermore, empirical tests reveal that between-group inequality has a large, robust and negative relationship with public goods provision, whereas cultural fractionalization, ELF, and overall inequality do not.

*We are grateful for research assistance from Tom Ogorzalek and for helpful comments from Tim Frye and seminar participants at Nuffield College, Oxford and at the Russell Sage Foundation. Huber is grateful for research support from Columbia University's Center for the Study of Ethnicity and Race, and from the National Science Foundation.

†Ph.D. candidate, Department of Political Science, Columbia University and Fellow, Alexander Hamilton Center, New York University. Mailing address: 420 West 118th St., 7th Floor, New York, NY 10027. Telephone: 917-375-7614. Fax: 212-854-5670. E-mail: kab2106@columbia.edu.

‡Professor, Department of Political Science, Columbia University and Visiting Scholar, Russell Sage Foundation. Mailing address: 420 West 118th St.,#713, New York, NY 10027. Telephone: 212- 854-7208. Fax: 212-222-0598. E-mail: jdh39@columbia.edu. Corresponding author.

Introduction

Ethnic diversity is widely held to make governance more difficult. Such diversity is associated with low production of public goods, poor economic growth, and high levels of corruption, violence and civil conflict. But diversity hardly sentences a country to poor political and economic outcomes. Latvia, for example, has better governance indicators than Brazil, and Zambia has better governance indicators than Nigeria, even though these pairs of countries have similar levels of ethnolinguistic fractionalization. Why, then, do some countries cope more successfully with ethnic diversity than others?

This paper addresses this question by focusing on the nature of substantive differences between groups. The vast majority of cross-national evidence about ethnic diversity and governance utilizes the standard measure of ethnolinguistic fractionalization (ELF) (e.g., Alesina et al. 2003; Alesina and La Ferrara 2005; Collier 2000; Easterly and Levine 1997; La Porta et al. 1999). This measure contains information about the identity and size of groups but incorporates no other information about groups' substantive characteristics. Existing arguments about how ethnic diversity affects governance, however, are typically grounded in the assumption that groups differ from each other in substantively important ways, and posit that these differences underlie governance problems in multiethnic societies.

This paper examines two important types of differences between groups – cultural and economic. Our goal is to understand the empirical relationship between such differences and public goods provision across countries. Is diversity more problematic for governance in countries when this diversity is based on strong cultural or economic differences between groups? Do standard empirical results about ethnolinguistic fractionalization still hold when controlling for the cultural or economic differences between groups?

The focus on cultural differences has received substantial attention in the literature on ethnic diversity. Scholars argue that such differences make it more difficult for individuals to cooperate across groups. This may be true for a number of reasons, as Habyarimana et al. (2009) describe. One reason is that ethnic similarities make it easier for individuals to communicate with each other (e.g., Deutsch 1966; Bacharach and Gambetta 2001). The shared languages and social networks of ethnically similar individuals allow them to assess each other's intentions and trustworthiness, and to communicate goals and necessary actions. These individuals experience lower transaction costs when cooperating towards common ends. In addition, ethnically similar individuals should find it easier to sanction each other for failing to cooperate (e.g., Fearon and Laitin 1996; Grief 1994; Miguel and Gugerty 2005). Thus, as the cultural differences between groups in a country grow, public goods should be harder to produce. Although these arguments have been subject to limited cross-national empirical research, a recent paper by Desmet, Ortuño and Weber (2009) shows that redistribution by the government is lower in countries that have higher levels of linguistic diversity, which is a key indicator of cultural differences.

Economic differences between groups have received less attention in the literature, in part because there have existed no empirical measures of group-based economic differences across countries. However, there are good reasons to expect that such differences will affect governance. Group-based economic differences can lead to different group needs with respect to public goods, feelings of alienation or discrimination by some groups, different attitudes toward redistribution across groups, and different "class" identities by different groups. The effect of group economic differences on the policy preferences of group members are likely to be particularly important in affecting governance. If the different economic statuses of groups lead them to prioritize different public goods, it will be difficult for them to reach agreement on which public goods to provide (e.g., Alesina and Drazen 1991; Alesina and Spolaore 1997; Alesina, Baqir and Easterly 1999). Under these circumstances, politicians may try to win re-election by providing private goods for each group, especially when the number of groups is not too large (Fernández and Levy 2008).

This paper demonstrates that it is possible to measure differences in the economic well-being of groups using existing cross-national surveys of citizens. Specifically, we use between-group inequality (BGI), which is a weighted average of the differences in mean incomes across groups in a country, as a measure of economic differences between groups. We then argue that between-group inequality can be satisfactorily measured using surveys like the Afrobarometer, the World Values Survey, and the Comparative Study of Electoral Systems. We combine the measure of between-group inequality with existing measures of ELF, and with existing measures of cultural differences between groups that are based on language differences (Fearon 2003; Desmet, Ortuño and Weber 2009). The data show these variables measure different things, and that the choice between them has an important impact on our understanding of which countries are most ethnically diverse.

Which measure of ethnic diversity shows the strongest association with public goods provision? We find no robust empirical relationship between either the standard ELF measure or measures of cultural difference and public goods provision. The tests do reveal that between-group inequality has a large, robust negative relationship with public goods provision. Countries with higher levels of inequality between groups have lower levels of public goods, a finding that has important implications for understanding the pathways by which ethnic diversity creates governance problems.

The paper is organized as follows. The next section introduces the measures of ethnic differences – cultural fractionalization and between-group inequality – that are central to the analysis. Since cross-national measures of between-group inequality do not currently exist, the following section describes how they can be created from cross-national surveys. We then compare the measures of cultural fractionalization, between-group inequality and ethnolinguistic fractionalization with each other. Our main empirical tests follow. We first treat each country as a unit of analysis and use OLS models to test the relationships between each measure and public goods provision. The results show that only BGI has a robust relationship. Between-group economic differences, however, can be caused by policies related to public goods. We therefore also estimate models aimed at exploring whether BGI has a causal effect on public goods provision. The final section concludes the paper.

Measures of cultural and economic differences between groups

The well-known index of ethnolinguistic fractionalization, or ELF, measures the probability that two randomly chosen individuals will belong to different groups. It is written as

$$ELF = 1 - \sum_{i=1}^{n} p_i^2, \tag{1}$$

where p_i is the proportion of individuals who belong to group i, and n is the number of groups in society.[1] As noted above, ELF does not include information on the extent of cultural or economic differences across groups. But ELF can be altered to incorporate information about group-based differences.

To measure cultural differences, one approach is to consider language differences between ethnic groups. The importance of language to cultural identity has been emphasized by many scholars (Gellner 1983; Laitin 1994, 1998). Linguistic differences can lead to divergent preferences on linguistic and educational policy; in addition, they make communication between individuals more difficult, and they are often correlated with social networks (Milroy 1987). As a result, a version of ELF that incorporates information about the extent of linguistic differences between ethnic groups provides a useful measure of the level of cultural differences between groups.

In order to measure cultural differences, we use a measure first proposed by Greenberg (1956). This measure is a variation on ELF that captures the expected linguistic similarity between two randomly selected individuals in a society, and it has been used by Fearon (2003) to measure cultural fractionalization within societies. The measure is written:

$$CF = 1 - \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j r_{ij}, \tag{2}$$

where r_{ij} is a measure of the linguistic similarity of the languages of ethnic groups i and j that ranges between 0 and 1. CF will take the value 0 if all groups speak the same language, and will take its maximal value of 1 when all individuals are their own group and speak highly dissimilar languages.

We use Fearon's measure of linguistic similarity, which he constructed based on the distance between different languages in language trees. Linguists use language trees to classify languages into families, and then within each family they sub-classify languages into different branches. Fearon's measure is based on the premise that the more branches two languages have in common, the more similar the languages are to each other. In Spain, for example, Spanish and Basque are very different, as they come from two different language families (Indo-European and Basque). In contrast, Spanish and Catalan are quite similar, only branching apart from each other at the eighth junction in their language tree.[2] Formally, Fearon sets $r_{ij} = (l/15)^{\alpha}$, where l is the number of shared classifications (or branches) between i and j, and 15 is the maximal number of classifications for any language in the data set. Fearon sets $\alpha = \frac{1}{2}$. If two groups speak the same language, $r_{ij} = l = 1$ and if two groups speak languages from different linguistic trees, $r_{ij} = l = 0$; otherwise, r_{ij} takes a value between 0 and 1 that is increasing in the number of shared branches of the two group's languages. Desmet, Ortuño and Weber (2009) discuss different ways of constructing measures of linguistic resemblance and argue that it makes sense to set α smaller than the $\frac{1}{2}$ advocated by Fearon. They set $\alpha = \frac{1}{20}$, which gives less weight to having large numbers of shared branches within language trees as compared to being part of the same language tree. We are agnostic about where to set α, and thus include both Fearon's and Desmet, Ortuño and Weber's measures. [3]

Although factors other than language may create cultural barriers between groups, focusing on language differences has considerable merit because it employs an objective criteria that can be measured consistently across countries. While it is possible to imagine factors unrelated to language that lead to cultural differences in a country, it is more difficult to describe such factors in a way that is as amenable to the level of objectivity and cross-national measurement reliability as the linguistic measures. Cultural fractionalization is therefore a good proxy for the underlying level of cultural differences across groups in a society.

Next consider the measurement of group economic differences. A straightforward and easily interpretable measure of group-based income differences is between-group inequality (BGI). This measure is based on the familiar Gini index, but instead of calculating inequality based on each individual's income, it assigns each group's mean income to every member of that group. It can be interpreted as the expected difference in the mean income of the ethnic groups of any two randomly selected individuals. The differences in the mean incomes are weighted by (two times) the mean income of the society, a normalization that allows comparisons of BGI scores across countries with different income scales. The formula is :

$$BGI = \frac{1}{2\bar{y}}(\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \mid \bar{y}_j - \bar{y}_i \mid), \tag{3}$$

where $\bar{y}$ is the mean income in the country and $\bar{y}_k$ is the mean income of group k.

Although not widely familiar to political scientists, between-group inequality has been studied by economists interested in decomposing inequality into its different components.[4] Milanovic (2005) uses the BGI formula as a proxy for global inequality; his second measure of international inequality calculates a Gini in which each country's mean income is weighted by its population. Stewart (2003) and Mancini (2008) advocate the use of BGI as a measure of group differences in their studies of communal violence. The measure is also similar to the grouped version of the Generalized Ethno-linguistic Fractionalization (GELF) measure Bossert, D'Ambrosio and La Ferrara (forthcoming) develop and apply to the United States.[5]

Both the Gini and BGI have interpretations related to the Lorenz curve. The Lorenz curve is a graphical illustration of the cumulative distribution of a society's income over different ranges of the income distribution. In order to draw a Lorenz curve based on individual income differences, each person in society is ranked according to their individual income. The points on the Lorenz curve indicate that y percentage of the society's income accrues to the bottom x percentage of people in the income distribution. In a perfectly equal society, where the the "bottom" 10 percent of people control 10 percent of the society's income, the "bottom" 20 percent of people control 20 percent of the society's income, and so on, the Lorenz curve is equal to the 45 degree line. In any society where there is not perfect equality, the Lorenz curve is typically a convex curve below the 45 degree line. The Gini index is equal to two times the area between the Lorenz curve and the 45 degree line.

BGI is based on a ranking of people *not* by their individual income, but by their ethnic groups' mean income, and thus ignores information about income differences within groups. Each group member is ranked according to the mean income of their ethnic group. Since BGI is based on the proportion of income held by each group, the Lorenz curve for BGI will be a series of straight lines meeting at points where members of one group end and members of the next group begin. BGI is equal to two times the area between the group-based Lorenz curve and the 45 degree line.

Figure 1 illustrates BGI using the Lorenz curve for three hypothetical societies, each divided into three groups – group 1, group 2 and group 3 – with 35, 40 and 25 members respectively. Each society has 100 units of income to distribute among these groups. In a completely equal society, group 1 will receive 35 units of wealth, group 2 will receive 40 units of wealth and group 3 will receive 25 units of wealth. The Lorenz curve will be a straight line, since the group occupying the "bottom" 35 percent of the income distribution controls 35 percent of the income, the groups occupying the "bottom" 75 percent of the income distribution control 75 percent of the income, and so on. This case is depicted by the solid black line. BGI in this case is 0.

[Figure 1 about here.]

In an unequal society, the wealthiest group will control more income per person than the poorest group. The long dashed line draws the Lorenz curve if group 1 controls only 20 percent of the income, group 2 controls only 30 percent of the income, and group 3 controls 50 percent of the income. In this case, the group occupying the bottom 35 percent of the income distribution control only 20 percent of the income, and the two groups occupying the bottom 75 percent of the income distribution control only 50 percent of the income. The BGI in this case is .275. The short dashed line depicts the Lorenz curve in an even more unequal society in which group 1 controls just 5 percent of the income, group 2 controls 20 percent and group 3 controls 75 percent. In this case, the bottom 75 percent of the income distribution controls only 35 percent of the society's income and BGI is .55.

Like the Gini, BGI ranges from 0 to 1. It takes on its minimum value of 0 when the average incomes of all groups in society are the same, and it takes on its maximum value of 1 when one infinitely small group controls all of the income in society. Put another way, for any level of ELF, BGI will be 0 if all groups have the same mean income. Holding ELF constant, BGI is increasing in the income differences between groups. The effect of income differences across groups on BGI will be largest when ELF is largest. BGI will take its largest values when there are many equally sized groups (as with ELF), and when the income differences between the groups are large. Between-group inequality, then, can be viewed as an extension of the ELF that allows income differences between groups to vary. Conversely, ELF can be viewed as a restriction on between-group inequality that holds differences between all groups constant at 1.

Measuring between-group inequality using existing surveys

The main empirical challenge in analyzing the effects of between-group economic differences is constructing national-level measures of BGI for a large set of countries. A number of different scholars have constructed ELF measures (Alesina et al. 2003; Fearon 2003), and Fearon (2003) and Desmet, Ortuño and Weber (2009) have created measures of cultural fractionalization for almost all the countries in the world. But existing cross-national data sets do not include measures of between-group inequality, in part because BGI requires information on the economic well-being of groups that is not readily available from secondary sources. The construction of BGI for a particular country requires information on both the size of ethnic groups within the country and the economic well-being of these groups. This section describes how existing cross-national surveys – the World Values Survey (WVS), the Afrobarometer and the Comparative Study of Electoral Systems (CSES) – can be used to create BGI measures, and provides evidence regarding the validity and biases of the resulting measures.

The WVS, Afrobarometer, and CSES surveys all contain instruments that make it possible to identify the "ethnicity" of respondents. However, ethnic categories often nest inside broader categories, and as Posner (2004) demonstrates, choices about which groups to include can have a significant impact on the conclusions one draws about the relationship between group diversity and outcomes. A decision rule is therefore necessary to decide which groups are most relevant in a particular country. In this paper, we follow the identification of groups made by Fearon (2003) as closely as possible, as we believe that Fearon's work is the most careful and theoretically motivated classification of groups that has been completed to date. His seven criteria emphasize groups that are understood as "descent groups," and that are locally viewed as socially or politically consequential. Depending on the country, Fearon's identification of groups may be based on race (e.g., the US), language (e.g., Belgium), religion (e.g. France), tribe (e.g., many African countries), or even some combination of these factors. He draws on a range of secondary sources to identify the size of each group for the vast majority of countries in the world, and the resulting data set provides both a guide for the creation of the BGI measure, and a benchmark against which to judge it.

In classifying survey respondents into groups, we mirror as closely as possible the groups used by Fearon. In the US , for example, Fearon identifies four ethnic groups: whites, blacks, hispanics and Asians. Variable x051 in the WVS has seven categories: white, black, hispanic, other, central Asian, south Asian, and east Asian. We convert this seven-fold variable to the Fearon groups by recoding the three Asian groups to "Asian" and dropping "Other." It is not always possible to place all respondents into one of the Fearon groups. In France, for example, he identifies the groups as French, Muslim and Bretons. Using the CSES, we cannot identify the Bretons, but we can identify the French and Muslims.

To determine whether the Fearon groups are sufficiently well-identified by a survey to merit the inclusion of the survey in the data set, we have employed a *15 percent rule*. The rule works as follows. For each survey, we calculate the percentage of the population (per Fearon's data) that we cannot assign to any of Fearon's groups, and we retain the survey if this number is less than 15 percent. In France, for example, the only "missing group" – that is, the only group that Fearon identifies but that we cannot identify in the French surveys – is the Bretons, who (according to Fearon) represent 4.6 percent of the French population. We keep the French survey because 4.6 percent is less than our 15 percent threshold. In Israel, by contrast, Fearon distinguishes Arabs (15 percent of the population) and Palestinians (22 percent). Neither the CSES nor the WVS has an ethnic category for Palestinians, and since they represent over 15 percent of the population, we exclude these Israeli surveys. The analysis below also presents results that follow a *5 percent rule*, which obviously sets a higher standard for including countries in the data set.

We are interested in which types of ethnic diversity are associated with governance problems in countries that are at least *minimally* democratic. We focus on democracies because dictators have little incentive to provide public goods regardless of the ethnic diversity of their countries (Olson 1993; Sen 1999), and the effect of ELF on governance has been found to differ in democratic and authoritarian countries (Collier 2000). We therefore consider only countries that have a Polity 2 score of 1 or higher. Some countries have more than one survey, and when this occurs, we average BGI scores across the surveys for that country. The resulting 46 countries included in the analysis are listed in Table 1. The Afrobarometer surveys are from waves 2 and waves 3 (2002-2006), the CSES surveys are from all three waves (1996-2003), and the WVS surveys are from wave 4 (1996 and 1999). The sample includes democracies from all regions of the world, although we underrepresent Asia and especially Latin America in so far as these regions have a higher proportion of the world's democracies than their representation in this data set suggests.

[Table 1 about here.]

How well do the survey data reflect the composition of groups? It is straightforward to answer this by constructing a measure of ELF from the survey data and comparing this to the ELF scores from Fearon (2003). If one accepts the Fearon data as a good benchmark approximation of the true level of ethnic fractionalization, then one should have confidence in the survey-based measures only if the correlation between the survey data ELF and the Fearon ELF is strong. Figure 2 plots this relationship for the 46 countries. The survey-based ELF, which takes the average value within countries when there is more than one survey in a country, is obviously very closely related to the Fearon ELF. The Fearon-based measures typically take higher values, particularly at low levels of fractionalization, but the overall correlation is an impressive .96 (which is higher than the correlation between Fearon's measures of ELF and Alesina et al.'s (2003) measures of ELF).

[Figure 2 about here.]

The next challenge is measuring group economic well-being. Ideally, one would construct the BGI measure using fine-grained individual-level income or consumption data, aggregated by group. There currently do not exist a large number of multinational surveys that contain this type of data with appropriate information about group identity, but the WVS, Afrobarometer and CSES surveys contain coarse measures of each respondents' economic well-being that can be used to evaluate the relative well-being of different groups. Even coarse data on group income differences provides information that can be used to measure BGI.

Each of the three surveys measures respondents' income or consumption using a different metric. For example, the WVS survey asks the respondent to answer the following question:

> *Here is a scale of incomes and we would like to know in what group your household is, counting all wages, salaries, pensions and other incomes that come in. Just give the letter of the group your household falls into, after taxes and other deductions.*

The respondent is given a country-specific scale, typically with 8 to 10 categories, created to be meaningful within each country. Each CSES country survey asks a similar question to the WVS, but the CSES reports only the income quintiles of the respondents.

The Afrobarometer survey, like most surveys in developing parts of the world, does not have an income variable. For many individuals in such countries, such a question would be meaningless because the individuals have little or no cash income. Instead, the common strategy in such surveys is to ask respondents questions about their access to things crucial to their basic needs. For the Afrobarometer, each survey asks respondents a number of questions of the following type:

> *Over the past year, how often, if ever, have you or anyone in your family gone without food?*

In addition to asking about food, the survey asks about water, medical care, cooking fuel, and cash income. Each variable is coded on a five-point scale (from 0 to 4) according to how often the respondent has gone without the item. The sum of the responses from these five questions can be used to create an economic well-being metric that ranges from 1 (maximal unmet needs) to 21 (no unmet needs). This index is obviously most useful in distinguishing differences among the least well-off, masking differences that exist among the more well-to-do.

These "income" variables along with the group identity variables make it possible to calculate BGI for each survey. In cases where we have multiple surveys for a given country, we average across surveys to create a score for each country.[6] Validating this BGI measure is more difficult than validating the survey data on group composition because no benchmark data exist. However, because the measures of income are coarse, and thus do not fully characterize the income distribution, it is important to assess the validity of the measures and to be aware of biases they may possess.

We can gain some sense of the validity of the income measures by examining whether they reflect income disparities in the handful of cases where the nature of inequality between ethnic groups is widely acknowledged. It is reassuring to note that in the survey, Whites are richer than Mulattos who are richer than Blacks in Brazil; Bulgarians are richer than Turks in Bulgaria; the Flemish are wealthier than the French in Belgium; Whites are wealthier than Mestizos and Blacks in the Dominican Republic; Muslims are poorer than the “French” in France; and Whites are richer than Blacks and Hispanics in the US. In the African countries, where the income measures are most limited in that they only distinguish between differences among the poor, the data still capture the fact that Whites and people of mixed race background are wealthier than other ethnic groups in Namibia, the Ibo and Yoruba are richer than the Hausa in Nigeria, and Whites, Coloureds and Asians are richer than Blacks in South Africa. Thus, in these cases where the income relationships between groups are known, there is considerable face validity in the way that the data ranks the relative incomes of groups.

The problem, of course, is that data on between-group economic differences has not been previously collected on a large scale, so little is known about the economic differences between many other groups. To gain further insight into the validity of the BGI measures, we turned, where possible, to household income surveys with very fine-grained measures of income. Few surveys exist that identify income by the ethnic groups in our data set, but we identified household surveys that could be used to construct measures of the income or expenditures of the relevant ethnic groups for nine countries in our data set: Brazil, Bulgaria, Canada, Estonia, Finland, Germany, the US, South Africa and Zambia.[7] Although this is too few surveys to use in studying the relationship between BGI and public goods provision, these surveys allow an examination of the extent to which the survey measures used here differ from those constructed from finer-grained income data.

The correlation between the measures using our surveys and the measures from the household surveys is .68. But South Africa is a huge outlier, with a BGI in the fine-grained data that is three times higher than the next country, Brazil. If we exclude South Africa, the correlation between BGI from our surveys and BGI from the fine-grained expenditure surveys is an impressive .85. South Africa, of course, has a unique history of apartheid, which concentrated enormous wealth in the hands of whites. Our survey does not capture the magnitude of the inequality in South Africa since the Afrobarometer only measures the fact that virtually no whites in South Africa are poor. However, even so, our data ranks South Africa as having the third highest level of BGI in our data set, after the Dominican Republic and Brazil, suggesting that our data does a good job of ranking countries in terms of BGI.

As a final strategy for exploring how the Afrobarometer's focus on income differences among the poor affect measures of BGI, we conducted simulations. As described in more detail in the appendix, the simulated dataset contains thousands of societies made up of individuals whose exact income was known. To replicate the Afrobarometer coarsening technique, we assigned all people above a fixed poverty line in the simulated income data to the richest category and then divided the poor into different income levels based on their absolute level of deprivation. We then constructed measures of BGI for each society based on the "true" income variable and the "coarsened" income variable.

Figure 3 plots the true BGI (based on the fine-grained income distribution) against the coarsened income data (where income differences exist only at low income levels, as in the Afrobarometer). The dark xs represent societies where there are minimal income differences *within* ethnic groups (homogenous groups), the solid gray dots represent societies where there are some income differences *within* ethnic groups (heterogenous groups), and the open gray circles represent societies where there are large income differences *within* ethnic groups (very heterogenous groups). Figure 3 also depicts the 45 degree line, making it easy to identify whether "coarsening" leads to overestimates (points above the line) or underestimates (points below the line) of the true BGI. Not surprisingly, we can see that for any assumption about group income heterogeneity and for any of the three coarsening techniques, on average, the coarsened metrics underestimate the true level of between-group inequality. But we also find that the correlation between the true BGI and the BGI based on the coarsened data is very strong, ranging from .76 (when groups are most heterogenous) to .87 (when groups are most homogenous).

[Figure 3 about here.]

Thus, the simulations reinforce what we found in the data above: despite the fact that the Afrobarometer suppresses information about income differences among the non-poor, it provides measures of BGI that are highly correlated with finer measures of BGI. Furthermore, the Afrobarometer measures are biased in the "right" direction to the extent that they systematically underestimate the true BGI. This should make it more difficult to find differential results of ELF and BGI on public goods provision.

Comparing ELF, cultural fractionalization and between-group inequality

ELF, CF and BGI measure theoretically distinct concepts. But are the three measures also empirically distinct? Are the countries with the greatest linguistic differences between ethnic groups different from the countries with the greatest economic differences between groups?

[Figure 4 about here.]

Figure 4 depicts the relationships between the three measures of ethnic diversity, with each of the measures standarized to have a mean of 0 and a standard deviation of 1. First consider the relationship between CF and ELF, which is also discussed in Fearon (2003). For his full set of countries, he found that the two measures are highly correlated (r=.79), with the largest differences occurring in sub-Saharan Africa and Latin America. In our subset of 46 countries, the correlation between ELF and CF is .64, somewhat lower than in the Fearon's full data set. The top panel in Figure 4 shows that the correlation between the two variables is relatively strong throughout the range of ELF, but definitely weakens as ELF grows. A number of countries with high ELF – particularly those in Latin America – have groups that use similar languages. The countries whose diversity rankings are most affected by the switch from ELF to CF are listed in Table 2. Colombia ranks 13th in ELF and only 43rd in CF because the various ethnic groups in Colombia all speak Spanish. Madagascar is the second most diverse country in the data using ELF. But all the groups in Madagascar speak Malagasy or a closely related language, so Madagascar's CF score ranks only 30th in the data. Russia and Estonia, by contrast, move in the other direction, each increasing 13 places in the country rankings when one switches from ELF to CF. This is because the main language groups in these countries are from different language families.

[Table 2 about here.]

Next consider the relationship between ELF and BGI, which is depicted in the middle panel of Figure 4. As with the previous comparison, there is a strong correlation between the two variables (r= .62), but for many countries, their relative ranking depends substantially on which measure is used. Among the low-BGI countries, there is a wide range of ELF scores, and among the high ELF scores, there is a wide range of BGI scores. Table 3 lists the countries whose diversity rankings are most affected by the change from ELF to BGI. Indonesia is the country that declines the most: it is ranked 7th using ELF but only 23rd using BGI. The Dominican Republic moves sharply in the opposite direction: it is ranked only 29th using ELF, but has the highest BGI in the data set.

[Table 3 about here.]

The bottom panel in Figure 4 examines the relationship between CF and BGI. The correlation between these two variables (r=.17) is quite weak. As is clear in the figure, a number of countries, particularly in Latin America, have very high levels of between-group inequality but low levels of language difference. Furthermore, there are a number of countries, particularly in eastern and central Europe, where language differences are large but between-group economic differences are relatively small.

The analysis therefore demonstrates that the three diversity measures, though related, are clearly distinct. The relative rankings of the countries change, at times dramatically, across the different measures. Incorporating information about cultural fractionalization or the economic well-being of groups therefore alters how we understand the relative ethnic diversity of countries.

Group differences and public goods provision

Although the three measures differ, do the differences matter for understanding the relationship between diversity and public goods provision? Does one of the measures have a stronger relationship with public goods provision than the others, suggesting that it may be the more important factor driving the relationship between ethnic diversity and poor governance outcomes? This section presents results from a series of regression models that analyze the relationship between the measures of diversity and public goods provision.

The models in this section treat a country as the unit of analysis. As noted above, the survey data used to construct the BGI measures are from the 1996-2006 time period. When multiple surveys exist for a given country, we average across the surveys, minimizing measurement error that may be associated with particular surveys. We also average the WDI indicators used to measure public goods over the 1996-2006 period, which is important given that the WDI indicators are often missing in particular years. Because ELF and CF are constant over time, the strategy of using a country as a unit of analysis also facilitates comparison of the correlations between the various measures of diversity and public goods provision. However, it is difficult to settle questions about whether diversity affects public goods or public goods affect diversity in this framework, and the next section therefore estimates models that treat each survey as the unit of analysis in an effort to gain additional insight regarding causation.

To measure public goods provision, we rely on ten variables in the World Bank's World Development Indicators, each related to government-provided public goods, such as public health, education, public infrastructure, and the government's taxing capacity. There are a larger number of candidate variables in the WDI data set, but many have large numbers of missing values. We averaged the country values for each variable for the period 1996-2006 and then retained only those variables such that (a) none of the 46 countries had more than 3 missing, and (b) no variable was missing for more than 7 of the 45 countries. The resulting ten variables used in the analysis, with the number of missing countries in brackets, are:

- Expenditure per student, primary (% of GDP per capita) [7] (WDI variable name = SE.XPD.PRIM.PC.ZS)
- Public spending on education, total (% of GDP) [1] (SE.XPD.TOTL.GD.ZS)
- Immunization, measles (% of children ages 12-23 months) [0] (SH.IMM.MEAS)
- Immunization, DPT (% of children ages 12-23 months) [0] (SH.IMM.IDPT)
- Improved sanitation facilities (% of population with access) [7] (SH.STA.ACSN)
- Improved water source (% of population with access) [6] (SH.H2O.SAFE.ZS)
- Roads, paved (% of total roads) [0] (IS.ROD.PAVE.ZS)
- Procedures to enforce a contract (number) [0] (IC.LGL.PROC)
- Tax revenue (% of GDP) [4] (GC.TAX.TOTL.GD.ZS)
- Telephone lines (number per 100 people) [0] (IT.MLT.MAIN.P2)

Each of these variables alone is a noisy predictor of the overall level of public goods provision in a country, susceptible to measurement error that is likely idiosyncratic to particular countries. Therefore, rather than choosing one or more specific variables for analysis, we use the information from all ten variables to create a single measure of the overall level of public goods provision. To this end, we conduct a factor analysis on all ten variables. We first impute the values of the missing variables, where missing values on one of the ten variables are imputed as a function of the other nine.[8] We then use Bartlett scoring on the first dimension of the factor analysis to create the dependent variable, which we call *Public Goods*. The results for the first factor in this factor analysis are shown in Table 4. All ten of the variables load strongly on the first factor. The variables with the strongest relationship to the underlying factor, and which therefore receive the greatest weight in the construction of the public goods variable, are immunizations, sanitation, water and telephone lines, all with factor loadings greater than .8. Two variables have a loading that is less than .5 – Tax Revenue and Primary School Spending – but each of these variables still has a relatively strong relationship with the underlying factor. The first factor has an eigenvalue of 5.21 and explains 69 percent of the variance in these ten variables. Figure 5 plots Public Goods against the log of GDP per capita (measured in purchasing power parity). The two variables are obviously strongly correlated, but at any level of economic development, and in particular at lower levels of development, there exists variation in the level of public goods.

[Table 4 about here.]

[Figure 5 about here.]

We estimate OLS models with robust standard errors using Public Goods as the dependent variable. The models include several controls. One is the level of economic development (*GDP/capita (ln)*, measured using purchasing power parity), which is known to have a very strong relationship with the level of public goods provision. Scholars have also emphasized the importance of democracy for public goods provision (e.g., Lake and Baum 2001), and though we focus on countries that are in some sense democratic (achieving Polity 2 scores of greater than or equal to 1), there is considerable variation within our data set regarding the development of democratic institutions and practices. We therefore include *Polity 2* as a control for the level of democratic institutionalization. It is also important to take into account the population of a country.

If there are economies of scale, public goods provision may be largest in the most populous countries. But if the most populous countries present the most challenges to governance, public goods provision may be negatively correlated with population. The correlation matrix for all right-hand side variables is given in the appendix. For models that include the BGI measure based on the surveys, there are four indicator variables that identify the survey(s) used to create an observation. These dummy variables indicate whether a measure was constructed from the *Afrobarometer*, the *WVS* (equals 1 if country uses only WVS), the *CSES* (equals 1 if country uses only CSES), or the *WVS/CSES* (equals 1 if country uses both WVS and CSES, which is the omitted category). The indicators are included because the surveys use different measures of "income," which likely results in differences in the degree to which they underestimate the true level of between-group differences. In the regression models, we standardize each continuous right-hand side variable to have a mean of 0 and a standard deviation of 1. This makes it straightforward to compare the size of the effects of the different variables. The parentheses in the tables provide the p-values for the estimated coefficients.

In Table 5, models 1-4 include the controls and one of the measures of diversity. Model 1 includes ELF. Consistent with previous research, the coefficient for ELF is in the expected negative direction and is very precisely estimated. Model 2 uses Fearon's cultural fractionalization measure of diversity. The coefficient is in the expected negative direction but is very small and estimated with considerable error. Model 3 uses the Desmet, Ortuño and Weber measure of cultural fractionalization instead of Fearon's measure. The coefficient has the wrong sign and is also estimated with considerable error. Finally, in model 4, the coefficient for BGI is negative and very precisely estimated. In models including only one measure of ethnic diversity, then, ELF and BGI have a clear relationship to public goods provision but cultural differences do not.

Next we examine the results when we include measures of each form of diversity in the same model. Model 5 uses the Fearon measure of cultural fractionalization and model 6 uses the Desmet, Ortuño and Weber measure. In both of these models, the coefficient for cultural fractionalization is very imprecisely estimated and has the wrong sign. The models also show that when BGI and ELF are included in the same model, the relationship between ELF and Public Goods disappears. The ELF coefficient now has the wrong sign and is estimated with large error, while the coefficient from BGI remains negatively and rather precisely estimated, especially in model 6.

We have estimated a wide range of additional models, and the results for cultural fractionalization are consistently estimated with very large error and often have the wrong sign. Since there is no empirical support for including this variable, and it is strongly correlated with ELF, model 7 presents results when only BGI and ELF are included (along with the controls). ELF is positive, although it is estimated with considerable error, and the coefficient for BGI remains negative and precisely estimated. These results from Table 5 therefore suggest the possibility that in previous research claiming a correlation between ELF and public goods provision, the relationship was actually being driven by between-group economic differences.

[Table 5 about here.]

How robust is this result? We first address this question by exploring the possibility of spurious correlation due to omitted variables. Scholars have argued that inequality may affect collective behavior and public goods provision. Many scholars have hypothesized that inequality is negatively associated with social cohesion (Boix and Posner 1998; Knack and Keefer 1997; Uslaner 2002; Uslaner 2008). Khwaja (2009) finds that inequality makes it more difficult for communities to create infrastructure projects, and Bardhan (2000) and Dayton-Johnson (2000) find that inequality in landholdings negatively affects the collective management of water resources. One might therefore worry that between-group inequality is simply a proxy for inequality more generally. We test this possibility using a measure of *Gini* from Solt (2009), who uses the Luxembourg Income Study to enhance the data from the United Nations University's World Income Inequality Database. The coefficient for Gini in model 8 is positive, but it is measured with considerable error. The result for BGI is robust to the inclusion of the Gini variable.

Next consider the geographic segregation of groups. Geographic segregation could make it more difficult to provide groups in remote areas with public goods. To the extent that geographic isolation could be correlated with between-group differences, we can have more confidence that the relationship between between-group differences and public goods is not spurious if we control for the level of group geographic segregation. We control for geographic segregation using a measure of *Geographic Isolation* derived from work by scholars of residential segregation. This measure is described in the appendix. It takes higher values when geographic segregation is greater, and interestingly, it has a negative correlation with BGI (r=-.64). Geographic Isolation should have a negative coefficient if it makes public goods provision more difficult.

Model 9 adds Geographic Isolation to model 8. The coefficient on the variable has the expected negative sign but is estimated with substantial error. The inclusion of this variable, however, does not affect the conclusions about ELF and BGI. ELF remains insignificant while BGI is significant and precisely estimated. The relationship between BGI and PG found in the previous regressions should not therefore be attributed to a spurious correlation between either of these variables and the geographic segregation of groups.

Next we test the robustness of the BGI result by re-estimating model 9 using different subsets of the data. The data set contains a number of very homogenous countries, many of which are quite rich and have high levels of public goods provision. Model 10 therefore examines whether the results are robust when we include only highly heterogenous countries, with an ELF greater than .25. The result for BGI is essentially unchanged when the data is restricted to this subsample of the 37 most heterogenous countries. The ELF coefficient remains estimated with extremely large error. Similarly, the previous analysis includes a number of countries that are very well-established democracies, where problems of ethnic diversity are often held to be relatively small. Model 11 therefore focuses on the less established democracies by eliminating the 16 countries that have a Polity 2 score of 10. The results indicate that when the strongest democracies are excluded, the effect of BGI remains very precisely estimated and in fact has the largest absolute magnitude of any model in Tables 5 and 6.

Finally, some may be concerned that the "15 percent" rule for replicating the Fearon groups is too lax, resulting in the inclusion of countries that have relatively large groups that are not represented in the surveys. To address this concern, we implemented a 5 percent rule for the retention of countries. That is, we aggregated the total population of the groups for which we could not assign the respondents. If this constituted less than 5 percent of the population, we retained the survey. Otherwise we eliminated the survey from the data set. The change in the retention rule eliminates eight countries and ensures that any group that Fearon identifies that does not exist in our surveys is very small. The results from the 5-percent rule, presented in model 12, are consistent with those in the other models. The coefficient for BGI remains negative, quite large in absolute value, and precisely measured.

[Table 6 about here.]

Across a range of models, then, standard measures of ethnic fractionalization and cultural fractionalization do not have a robust correlation with public goods provision. The same holds for the level of inequality.

Between-group inequality, by contrast, does have a negative relationship with public goods provision, one that is robust to a wide range of model specifications and decisions about which countries to include in the analysis. Furthermore, the effect is large. Bearing in mind that we standardized the right-hand side variables to facilitate comparisons, simply inspecting the regression output indicates that the estimated coefficient for BGI is very large relative to other variables. For example, in model 9, the model containing all controls and the full data set, in absolute value, the BGI coefficient is more than double the estimated coefficient for ELF or Gini, and roughly the same size as the estimated coefficient for Polity 2.

Does BGI lead to lower public goods provision?

The analysis above establishes that when we control for between-group inequality, standard measures of ELF and cultural fractionalization are not correlated with public goods provision, whereas between-group inequality does have a robust correlation with public goods. Existing theoretical arguments focus on the possibility that between-group economic differences make policy agreement more difficult and invite discrimination by powerful groups against less powerful ones, lowering public goods provision. But between-group inequality is also likely the result of policy decisions regarding public goods. If clean water or paved roads or education expenditures, for example, can be targeted at specific groups, then such targeting could lead to between-group economic differences. And if public goods tend to help the least well-off, public goods should diminish between-group inequality whenever there is a correlation between income and group.

This section presents several additional models exploring whether BGI causes lower public good provision (as opposed to simply being correlated with public goods). The data do not permit us to address this issue definitively, largely because the time series is very short, there are two few observations on BGI, and there is too much missing data on public goods provision. But by taking advantage of the temporal component of the data, we can provide some evidence that between-group economic differences lead to lower levels of public goods.

The first step is to redefine "income" in the Afrobarometer surveys so as to reduce the possibility that the measure of economic well-being in Africa is endogenous to the measure of public goods provision. Recall that the consumption-based measure from the Afrobarometer surveys includes five components, and two of these are related to access to public services (medical care and clean water). The measure of public goods similarly includes measures of health care (immunizations) and water (percent of population with access). It is therefore plausible that public goods provision is directly affecting the measure of BGI in the African countries included in the analysis. In the results presented above, we include the information about access to health care and clean water to create as fine-grained a measure as possible about the economic well-being of Afrobarometer respondents. This is particularly important given the skew of the "needs" variables toward the "rich" in the Afrobarometers. But given the endogeneity concern, it is important to re-estimate the models using a measure of BGI in Africa that does not incorporate information on individuals' access to public services; this measure of BGI has been constructed based only on individuals' reported access to food, cooking oil and cash income.

Model 12 in Table 7 presents the results when model 9 is re-estimated using the re-calculated measures of BGI in Africa. The results are very consistent with those in model 9: BGI has a negative coefficient that is precisely measured. ELF continues to be insignificant (and has the wrong sign). There is no significant relationship between inequality and public goods. The results from the previous models were therefore not driven by the potential endogeneity between the BGI measure in Africa and the measure of public goods.

The next step is to use this recoded data (that excludes the endogenous items in the Afrobarometer) in models that treat each survey as an observation. The strategy is simple. If we measure BGI at time t and public goods using data subsequent to time t, it is more difficult to argue that public goods are influencing the level of between-group inequality. As noted above, this approach is not without problems, particularly because the measure of public goods suffers missing data issues. But using surveys as the unit of analysis provides an opportunity to examine whether BGI has a causal impact on public goods.

To measure the dependent variable, if the survey is taken at time t, we take the average value for each component of the public goods measure in times t through $t+3$ and then use the four-year averages to create the public goods measure as before.[9] The missing data problems from this approach are substantially more severe than they are in the aggregate approach because there are a number of four-year periods for which there are no data for particular variables. The contract enforcement variable, for example, is present only toward the end of the period covered by our data, and thus is missing for 53 percent of the survey-specific observations. Sanitation, water and roads are each missing for at least 37 percent of observations. Given the missing data issues, we present results for two different versions of the public goods variable. The first, which we will call *PG10*, uses all variables, whereas the second, *PG6*, uses only the six variables for which less than 25 percent of observations are missing. These are the two immunization variables, the two education variables, tax revenues, and telephones.

Models 14 through 17 in Table 7 use the survey-specific data to estimate the same specification as model 9, except that we now include a variable, Year (the year of the survey), to capture positive time trends that often exist in the provision of public goods, particularly in developing parts of the word.[10] Model 14 uses PG10 as the dependent variable and uses all available observations. The coefficient on BGI is negative with a p-value of .097. The coefficient on ELF, by contrast, has the wrong sign, and is measured with considerable error. Model 15 re-estimates the same model as 14, but using only countries with Polity 2 scores of less than 10. The estimated coefficient for BGI increases substantially in size and is much more precisely estimated (p=.032) in comparison with model 15. Models 16 and 17 use PG6 as the dependent variable. In model 16 (all data) BGI has the expected negative coefficient, but has a standard error that is larger than that found in other models. ELF continues to have the wrong sign and to be measured with considerable error. Model 17 includes only countries with Polity 2 <10, and the BGI coefficient is negative, substantively large and precisely estimated. If we assume, then, that public goods in the future cannot affect between-group inequality in the past, these results provide evidence that between-group inequality causes lower levels of public goods. This effect of BGI on public goods is particularly evident when we omit the most politically developed countries from the sample.

[Table 7 about here.]

Conclusion

The empirical analysis tells a clear story. First, and most importantly, we find a strong and robust relationship between the level of public goods provision and between-group inequality. By contrast, neither traditional measures of ELF, nor cultural differences between groups (measured using information about the languages groups speak) has such a relationship. Second, although there are clear limits on how hard we can push our data, we have suggestive evidence that between-group economic differences lead to lower public goods provision, particularly in the less established democracies. Third, we find that when controlling for group economic differences, the overall level of inequality itself has no impact on public goods provision. The analysis therefore strongly suggests that paying more attention to group economic differences will yield strong dividends in efforts to understand the impact of ethnic diversity, as well as inequality, on governance.

Several avenues for future research are worth pursuing. Although the BGI data employed in this paper provide useful information about group economic differences, it is important to continue the search for more fine-grained measures of group income to estimate BGI. It is equally important to explore the possibility of measuring cultural differences using factors other than language, such as religion. Additional insight could also be achieved by exploring other definitions of groups in efforts to determine how robust the results are to alternative categorizations of the ethnic groups themselves.

Second, the analysis here assumes that the effect of between-group differences on governance is the same across political systems. But do some institutional forms for governance mitigate or exacerbate the effect of BGI on outcomes? We find, for example, that the negative effect of BGI is largest when the most well-developed democracies are eliminated from the data set, suggesting that there may be an interaction between the level of democracy and BGI. And it may be the case that particular forms of democracy may mediate the effect of BGI. Does federalism, for example, soften the impact of between-group differences on governance by giving groups autonomy to provide the public goods they most value?

Finally, we have provided some evidence that BGI has a causal impact on public goods provision. But as we noted, there are good reasons to suspect that public goods policy also affects between-group inequality. It is certainly possible – perhaps even likely – that these two variables are mutually and negatively reinforcing: low public goods provision exacerbates group economic differences which impedes public goods provision which exacerbates group-based inequality, and so forth. We cannot use our data to explore this issue, but it is clear that group-based economic differences do not arise by chance – they are the result of political processes that unfold over time, and that are reinforced or ameliorated by government policy decisions. Between-group inequality might therefore be rightly construed as a measure of group-based discrimination, and the level of such discrimination clearly varies across countries with similar levels of ethnic or cultural fractionalization. Why, then, do some countries have higher levels of group-based inequality than others? And what role does public goods provision play in answering this question? In light of the findings in this study, addressing this question should play a large role in improving our understanding of how between-group economic differences affect policymaking.

Appendix 1: Construction of simulated data set

We constructed a simulated data set in order to examine the effect of the Afrobarometer's "coarse" and truncated measurement of income on estimation of BGI. The final data set contained 18,000 100-person societies that differed in terms of the number of ethnic groups, the size of different ethnic groups, the average income of ethnic groups and the heterogeneity of incomes within ethnic groups.

The goal was not to mirror the actual distribution of incomes across differently sized ethnic groups in the real world (which is impossible, given that no data exists on this), but rather to make a data set that contained an observation for all possible permutations across a realistic range of values for each of these four variables. The simulated data set contained countries with between two and six groups, with the size of each group varying from 5 to 95 percent of the population. For each population distribution, we created societies where the income of each group member was drawn from a normal distribution centered at 8, 9, 10, 11 and 12. The range of individual incomes was from 0 to 76. This resulted in a huge data set of close to 200,000 simulated societies.

We took a random sample of 6,000 of these societies, and then we created three versions of each selected society with different levels of income heterogeneity within ethnic groups. We chose three reasonable levels of group income heterogeneity based on data from the Luxembourg Income Study. In "homogenous" societies, each group had a normally distributed income distribution with a standard deviation equal to .53 of the group's mean income. In "heterogenous" societies, each group had a normally distributed income distribution with a standard deviation equal to .89 of the group's mean income, In "very heterogenous" societies, each group had a normally distributed income distribution with a standard deviation equal to 1.2 times the group's mean income. The final simulated data set therefore includes 18,000 societies, some with economically homogenous and some with economically heterogenous ethnic groups.

We mimicked the Afrobarometer's coarsening technique by assigning all individuals with an income of more than 10 to the first (richest) category. Individuals with an income between 10 and 8 were assigned to the second category, individuals with an income between 8 and 6 were assigned to the third category, individuals with an income between 6 and 4 were assigned to the fourth category and individuals with an income below 4 were assigned to the final (poorest) group. Then we constructed measures of BGI based on each individual's "true" income and based on the "coarsened" measure of income.

Appendix 2: Correlation matrix of right-hand side variables

[Table 8 about here.]

Appendix 3: Construction of geographic isolation measure

Our measure of geographic isolation is based on a measure used by scholars of residential segregation. Isolation ("I") measures "the extent to which minority members are exposed only to one another" (Massey and Denton 1988, p. 288). The measure of isolation of group g is given by

$$I_g = \sum_{i=1}^{n} \left(\frac{p_g^i}{P_g} \cdot \frac{p_g^i}{T_i} \right), \tag{4}$$

where i is a region, n is the total number of regions, p_g^i is the population of group g in region i, P_g is the total population of group g in the country, and T_i is the total population in region i.

I_g will increase as a group becomes more concentrated in the same region (holding the size and distribution of other groups constant). It has a theoretical maximum of 1 (which occurs when all members of a group live in a region (or regions) that have no members of other groups). And it has a theoretical minimum of 0 (which occurs when each member of a group is the only member from that group in his or her region).

To describe the aggregate geographic isolation of all ethnic groups in a country, GI is the weighted sum of the isolation scores for each group:

$$GI_k = \sum_{g=1}^{G} \left(I_g \cdot \frac{P_g}{T} \right), \tag{5}$$

where k is a country, G is the total number of groups in country k and T is the population of country k. To calculate this variable, we use the region variable that exists in each of the surveys. It is important to acknowledge that there could be substantial residential segregation of groups within the regions identified in the surveys, so this measure almost certainly underestimates the actual degree to which groups are segregated from each other. However, to the extent that groups live in different regions, the value of GI will increase.[11]

Notes

[1]The measure of ELF used in this paper is from Fearon (2003) and is based on data from the early 1990s. We discuss his system for identifying groups below.

[2]Both languages follow these seven branches: Indo-European — Italic — Romance — Italo-Western — Western — Gallo-Iberian — Ibero-Romance. At this point they diverge in their linguistic classification, with Spanish on the West-Iberian branch and Catalan on the East-Iberian branch.

[3]The main difference between these measures is the α they use in calculating cultural fractionalization, but they also rely on different data to identify ethnic groups. Fearon uses his classification of ethnic groups, based on data from the early 1990s, and Desmet, Ortuño and Weber (2009) use data from the 2005 *Ethnologue*.

[4]On the decomposition of the Gini into its between, within and overlap components, see Pyatt (1976) and Yitzhaki and Lerman (1991).

[5]The innovation of the GELF is that it does not *require* one to impose *ex ante* group partitions onto the data. However, in research where scholars are interested in comparing exogenously defined groups, GELF is very similar to BGI. The GELF-equivalent of BGI would be based on a "similarity matrix" where individuals from the same ethnic group have the same income, and where the similarity between members of any two different ethnic groups is a function of the distance between the mean incomes of the two ethnic groups (normalized to range from 0 to 1).

[6]There are three usable WVS surveys from African countries, each of which also has an Afrobarometer survey. Since the Afrobarometer measure of "income" is quite different than the measure of income from WVS, and since we are taking averages when we have multiple surveys, we do not include these three WVS surveys in our analysis. When these three WVS surveys are included, the results are not affected. For WVS and CSES we impute missing values for income using standard demographic variables. There are few missing values for the relevant variables in the Afrobarometer surveys.

[7]The surveys used were the Brazilian National Household Sample Survey (2006), the Bulgarian Integrated Household Survey (1995), the Canadian Census (2001), the Estonian Household Budget Survey (2000), the Finnish Income Distribution Survey (2000), the German Socioeconomic Panel (2000), the U.S. Current Population Survey (2000), the South African Living Standards Monitoring Survey (1993), and the Zambian Living Conditions Monitoring Survey (1996).

[8]For the imputation, we use all countries in the WDI data set, not just those for which we have ethnic diversity measures. Given the strict selection criteria employed regarding missing values and variable

selection, we needed to impute only 25 out of 460 variable values.

[9]We include year t in the measure of the dependent variable because doing so helps minimize missing data issues, and because it seems very unlikely that public goods in year t will have a simultaneous impact on between-group differences in that year.

[10]There are also some observations with missing Gini data. In such cases, we use data from the first year prior to the current year for which there is non-missing data.

[11]There is no usable region variable for Ireland.

References

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Quarterly Journal of Economics* 114 (4): $1243-1284$.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8 (2): $155-194$.

Alesina, Alberto, and Allen Drazen. 1991. "Why are Stabilizations Delayed?" *American Economic Review* 81 (5): $1170-1188$.

Alesina, Alberto, and Eliana La Ferrara. 2005. "Ethnic diversity and economic performance." *Journal of Economic Literature* 43 (3): $762-800$.

Alesina, Alberto, and Enrico Spolaore. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economic* 112 (4): $1027-1056$.

Bacharach, Michael, and Diego Gambetta. 2001. "Trust as Type Detection." In *Trust and Deception in Virtual Societies* ed. Christiano Castelfranchi and Yao-Hua Tan. Norwell, MA: Kluwer Academic Publishers.

Bardhan, Pranab. 2000. "Irrigation and Cooperation: an Empirical Analysis of 48 Irrigation Communities in South India." *Economic Development and Cultural Change* 48: 847-65.

Boix, Carles, and Daniel Posner. "Social Capital: Explaining Its Origin and Effect on Government Performance." *British Journal of Political Science* 28 (4): 686-695.

Bossert, Walter Conchita D'Ambrosio and Eliana La Ferrara. Forthcoming. "A Generalized Index of Fractionalization." *Economica*.

Collier, Paul. 2000. "Ethnicity, Politics and Economic Performance." *Economics and Politics* 12 (3): 222-45.

Dayton-Johnson, Jeff. 2000. "The Determinants of Collective Action on the Local Commons: a Model with Evidence from Mexico." *Journal of Development Economics* 62 (1):181-208

Desmet, Klaus, Ignacio Ortuño and Shlomo Weber. 2009. “Linguistic Diversity and Redistribution.” *Journal of the European Economic Association* 7(6): 1291-1318.

Deutsch, Karl. 1966. *Nationalism and Social Communication*. Cambridge: The MIT Press.

Easterly, William, and Ross Levine. 1997. “Africa’s Growth Tragedy: Policies and Ethnic Divisions.” *Quarterly Journal of Economics* 112 (4): $1203 - 1250$.

Fearon, James, and David Laitin. 1996. “Explaining Interethnic Cooperation.” *American Political Science Review* 90 (4): 715-735.

Fearon, James. 2003. “Ethnic and Cultural Diversity by Country.” *Journal of Economic Growth* 8 (2): $195 - 222$.

Fernàndez, Raquel, and Gilat Levy. 2008. “Diversity and Redistribution." *Journal of Public Economics* 92(June): 925-43.

Gellner, Ernest. 1983. *Nations and Nationalism.* Ithaca, New York: Cornell University Press.

Greenberg, Joseph H. 1956. “The Measurement of Linguistic Diversity.” *Language* 32:109-115.

Greif, Avner. 1994. “Cultural Beliefs and the Organization of Society.” *Journal of Political Economy* 102 (5): $912 - 50$.

Habyarimana, James, Macartan Humphreys, Daniel Posner, and Jeremy Weinstein. 2009. *Coethnicity: Diversity and Dilemmas of Collective Action.* New York: Russell Sage Press.

Khwaja, Asim. 2009. “Can Good Projects Succeed in Bad Communities?” *Journal of Public Economics* 93 (7-8): 899-916.

Knack, Stephen, and Philip Keefer. 1997. “Does Social Capital Have an Economic Payoff? A Cross-Country Investigation.” *Quarterly Journal of Economics* 112 (4): 1251-88.

Laitin, David. 1994. “The Tower of Babel as a Coordination Game: Political Linguistics in Ghana.” *American Political Science Review* 88 (3): 622-34.

Laitin, David. 1998. *Identity in Formation: the Russian-Speaking Population in the Near Abroad*. Ithaca: Cornell University Press.

Lake, David, and Matthew Baum. 2001. "The Invisible Hand of Democracy: Political Control and the Provision of Public Services." *Comparative Political Studies* 34 (6): 587-621.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert W. Vishny. 1999. "The Quality of Government." *The Journal of Law, Economics, and Organization* 15 (1): 22-79.

Mancini, Luca. 2008. "Horizontal Inequality and Communal Violence: Evidence from Indonesian Districts." In *Horizontal Inequalities and Conflict. Understanding Group Violence in Multiethnic Societies* ed. Frances Stewart. London: Palgrave.

Massey, Douglas S. and Nancy A. Denton. 1988. "The Dimensions of Residential Segregation." *Social Forces* 67 (2): 281-315.

Miguel, Edward, and Mary Kay Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya." *Journal of Public Economics* 89 (11-12): 2325-68.

Milanovic, Branko. 2005. *Worlds Apart: Measuring International and Global Inequality.* Princeton, NJ: Princeton University Press.

Milroy, Lesley. 1987. *Language and Social Networks*. Oxford: Wiley-Blackwell.

Olson, Mancur. 1993. "Dictatorship, Democracy and Development." *American Political Science Review* 87 (3): 567-76.

Posner, Daniel. 2004. "Measuring Ethnic Fractionalization in Africa." *American Journal of Political Science* 48 (4): $849 - 863$.

Pyatt, Graham. 1976. "On the Interpretation and Disaggregation of Gini Coefficients." *The Economic Journal* 86: 243-55.

Sen, Amartya. 1999. *Development as Freedom*. New York, Anchor Books.

Solt, Frederick. 2009. “Standardizing the World Income Inequality Database.” *Social Science Quarterly* 90 (2): 231-42.

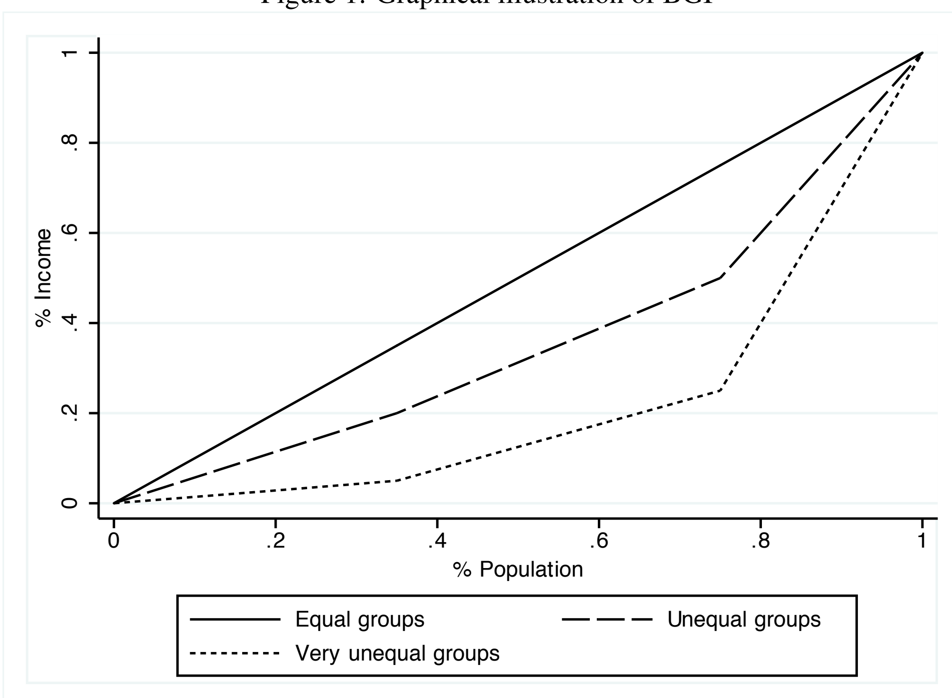
Stewart, Frances. 2003. “Horizontal Inequalities: A Neglected Dimension of Development.” CRISE Working Paper No. 1, Oxford.

Uslaner, Eric M. 2002. *The Moral Foundations of Trust*. New York: Cambridge University Press.

Uslaner, Eric M. 2008. *Corruption, Inequality, and the Rule of Law: The Bulging Pocket Makes the Easy Life*. New York: Cambridge University Press.

Yitzhaki, Shlomo and Robert Lerman. 1991. "Income Stratification and Income Inequality." *Review of Income and Wealth* 37 (3): 313 – 329.

Figure 1: Graphical illustration of BGI



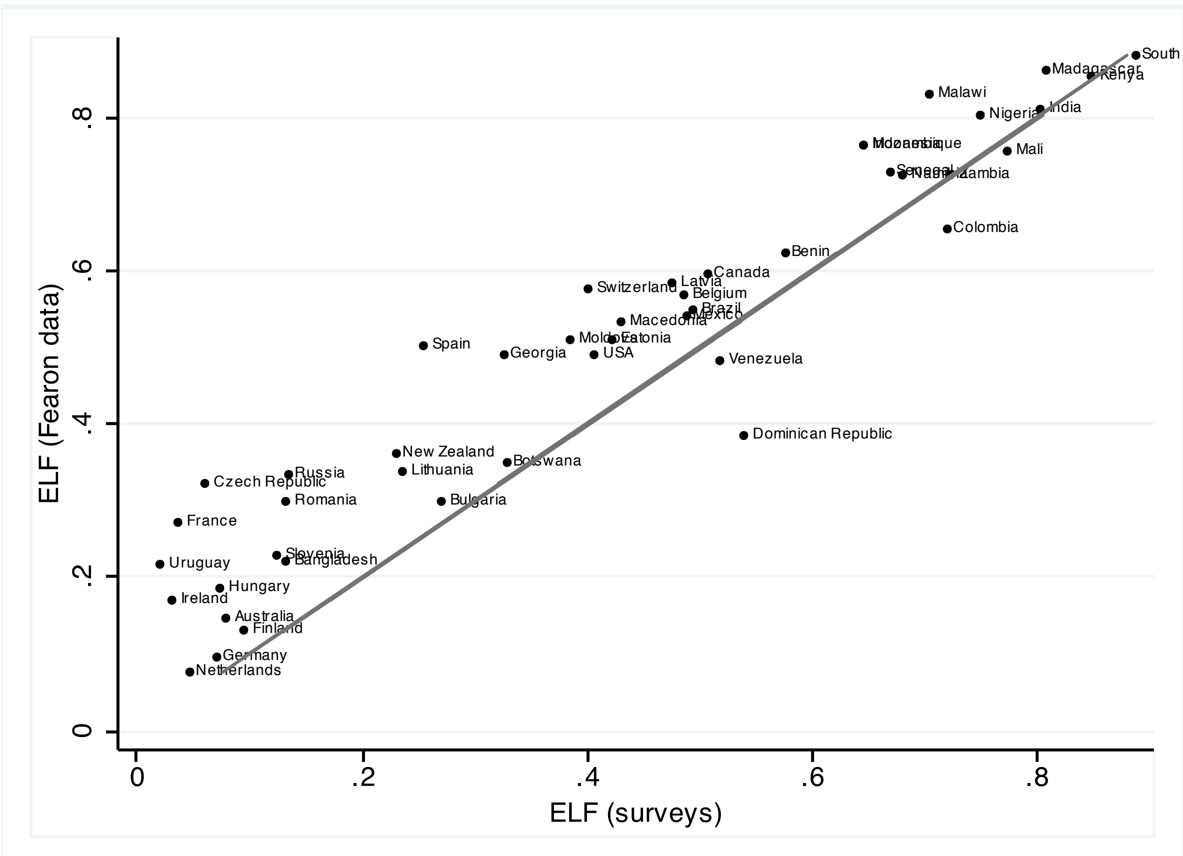


Figure 2: Comparing ELFs: Fearon data vs. survey data

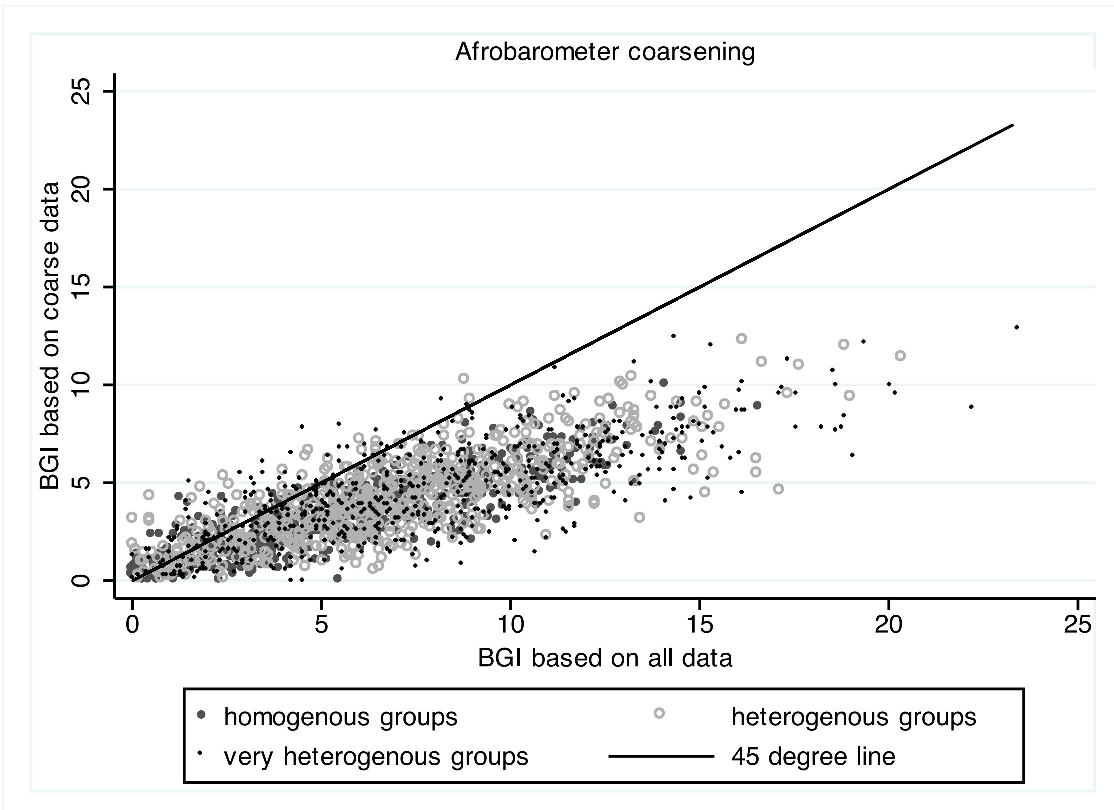


Figure 3: Simulating the effect of Afrobarometer data on estimates of BGI

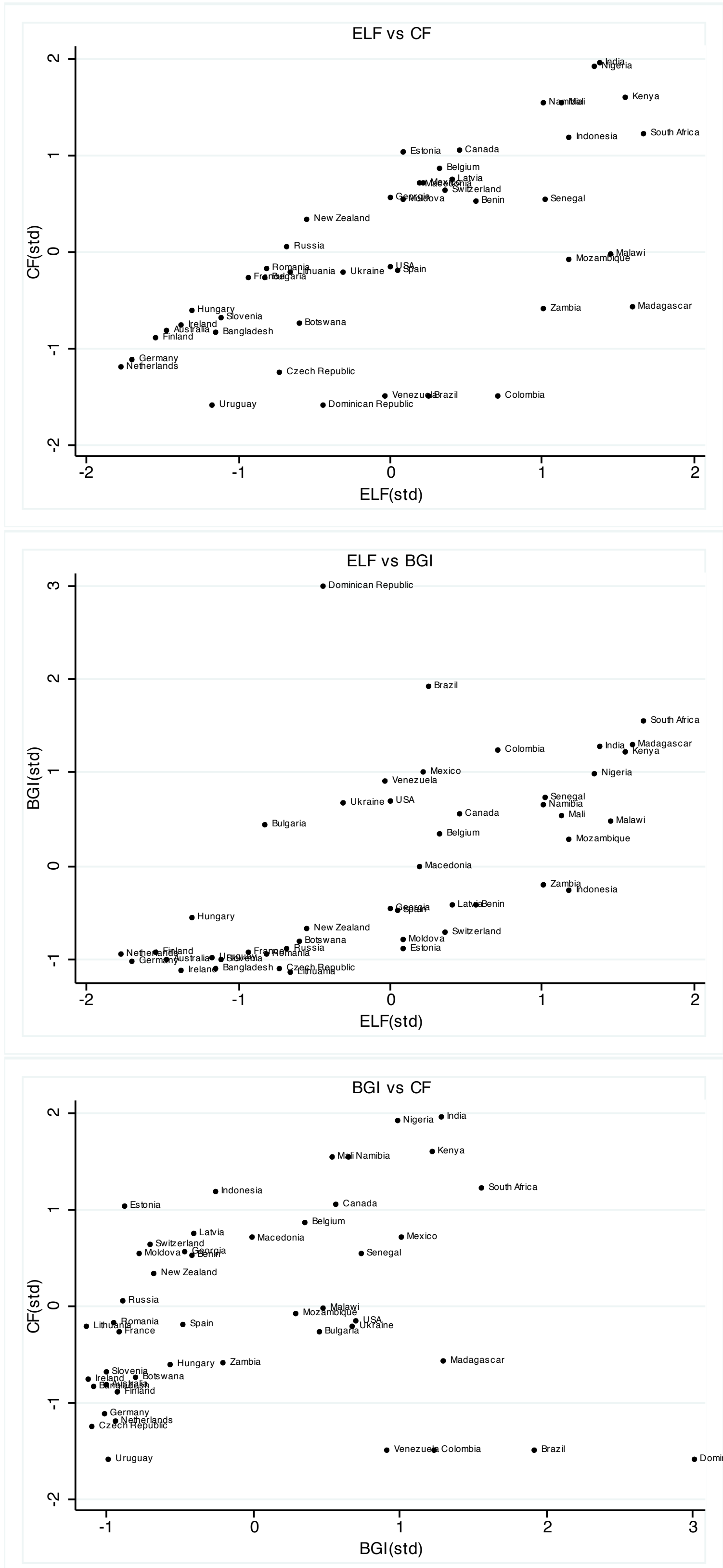


Figure 4: ELF, CF and BGI in 46 countries

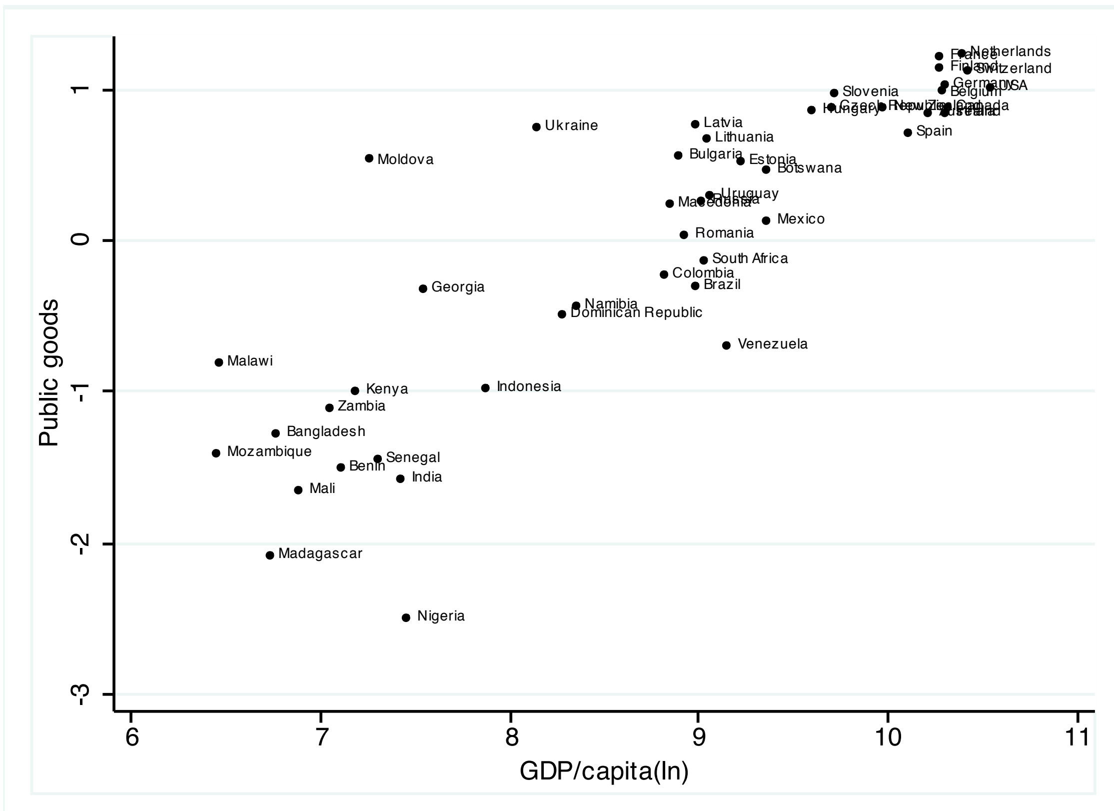


Figure 5: Public goods provision and national wealth

Table 1: Countries in study

Survey	Region					
	Western Europe	Other Europe	Asia	Latin America	Africa	Other
WVS only	Belgium Germany Ireland Netherlands Switzerland	Estonia Georgia Latvia Macedonia Moldova	Bangladesh Indonesia India	Colombia Dominican Republic Uruguay Venezuela		
CSES only	Finland France	Bulgaria Czech Republic Hungary Lithuania Romania Russia Slovenia Ukraine				
Both CSES and WVS	Spain			Brazil Mexico		Australia Canada New Zealand USA
Afrobarometer					Benin Botswana Kenya Madagascar Malawi Mali Mozambique Namibia Nigeria Senegal South Africa Zambia	
Total countries	8	13	3	6	12	4

Table 2: Countries whose diversity rankings change most when switching from ELF to CF

Countries with largest decline in rank			
Country	ELF rank	CF rank	Δ Rank
Colombia	13	43	-30
Madagascar	2	30	-28
Brazil	19	42	-23
Zambia	11	31	-20
Venezuela	27	44	-17
Malawi	4	21	-17
Dominican Republic	29	46	-17
Countries with largest increase in rank			
Country	ELF rank	CF rank	Δ Rank
Russia	33	20	13
Estonia	22	9	13
Romania	35	24	11
Georgia	26	15	11
New Zealand	30	19	11

Table 3: Countries whose diversity rankings change most when switching from ELF to BGI

Countries with largest decline in rank			
Country	ELF rank	BGI rank	Δ Rank
Indonesia	7	23	-16
Lithuania	32	46	-14
Switzerland	17	30	-13
Malawi	4	16	-13
Mozambique	8	20	-12
Countries with largest increase in rank			
Country	ELF rank	BGI rank	Δ Rank
Dominican Republic	29	1	28
Bulgaria	36	18	18
Brazil	19	2	17
Venezuela	27	10	17
Ukraine	28	13	15

Table 4: Factor analysis of public goods variables

Variable	*Factor 1*
Primary school spending	.47
Total public spending on education	.52
Measles immunizations	.82
DPT immunizations	.88
Sanitation facilities	.91
Water source	.86
Roads	.70
Contract enforcement	-.58
Tax revenue	.41
Telephone lines	.85

Eigenvalue of Factor 1: 5.21

Proportion of variance explained by Factor 1: .69

Table 5: Group differences and public goods provision I

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ELF	-0.210	–	–	–	0.009	0.072	0.096
	(0.005)				(0.950)	(0.473)	(0.327)
CF-Fearon	–	-0.007	–	–	0.089	–	–
		(0.928)			(0.383)		
CF-Desmet	–	–	-0.024	–	–	0.069	–
			(0.764)			(0.326)	
BGI	–	–	–	-0.137	-0.133	-0.166	-0.168
				(0.041)	(0.093)	(0.023)	(0.026)
GDP (ln)	0.635	0.733	0.737	0.492	0.474	0.461	0.506
	(0.000)	(0.000)	(0.000)	(0.007)	(0.007)	(0.008)	(0.004)
Population	-0.140	-0.173	-0.168	-0.200	-0.237	-0.240	-0.216
	(0.004)	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
Polity2	0.123	0.159	0.149	0.172	0.207	0.224	0.190
	(0.351)	(0.272)	(0.265)	(0.146)	(0.094)	(0.074)	(0.124)
Afrobarometer	–	–	–	-0.620	-0.684	-0.761	-0.676
				(0.038)	(0.038)	(0.025)	(0.036)
WVS	–	–	–	-0.043	-0.015	-0.061	-0.015
				(0.747)	(0.911)	(0.679)	(0.916)
CSES	–	–	–	0.141	0.177	0.145	0.196
				(0.447)	(0.327)	(0.461)	(0.287)
Constant	0.000	0.000	0.000	0.149	0.149	0.192	0.143
	(1.000)	(1.000)	(1.000)	(0.282)	(0.277)	(0.177)	(0.293)
Adj. R-squared	0.787	0.761	0.761	0.832	0.829	0.829	0.830
N	46	46	46	46	46	46	46

P-values based on robust standard errors are in parentheses.

Table 6: Group differences and public goods provision II

	All data (8)	All data (9)	ELF>.25 (10)	Polity2<10 (11)	5% rule (12)
ELF	0.104	0.087	-0.032	0.262	0.108
	(0.296)	(0.365)	(0.795)	(0.237)	(0.398)
BGI	-0.222	-0.259	-0.279	-0.349	-0.280
	(0.019)	(0.016)	(0.007)	(0.013)	(0.025)
Gini	0.126	0.121	0.237	0.248	0.103
	(0.187)	(0.193)	(0.025)	(0.027)	(0.341)
Geo. Isolation	–	-0.074	0.052	0.097	-0.085
		(0.387)	(0.610)	(0.521)	(0.358)
GDP (ln)	0.476	0.487	0.272	0.303	0.444
	(0.006)	(0.004)	(0.114)	(0.133)	(0.020)
Population	-0.214	-0.196	-0.226	-0.325	-0.200
	(0.001)	(0.001)	(0.000)	(0.002)	(0.003)
Polity2	0.224	0.233	0.321	0.456	0.268
	(0.082)	(0.063)	(0.014)	(0.003)	(0.031)
Afrobarometer	-0.830	-0.811	-1.051	-1.304	-0.824
	(0.016)	(0.015)	(0.004)	(0.003)	(0.025)
WVS	-0.031	-0.021	-0.060	-0.164	0.008
	(0.835)	(0.896)	(0.784)	(0.535)	(0.969)
CSES	0.213	0.240	-0.025	0.237	0.234
	(0.257)	(0.231)	(0.930)	(0.535)	(0.318)
Constant	0.186	0.179	0.367	0.470	0.167
	(0.195)	(0.215)	(0.030)	(0.030)	(0.307)
Adj. R-squared	0.834	0.830	0.811	0.732	0.819
N	46	45	37	30	37

P-values based on robust standard errors in parentheses.

Table 7: Exploring the causal effect of BGI on public goods

Dependent variable:	PG	PG 10	PG 10	PG 6	PG 6
Sample:	All obs.	All obs.	Polity2<10	All obs.	Polity2<10
	(13)	(14)	(15)	(16)	(17)
ELF	0.063	0.086	0.217	0.058	0.230
	(0.499)	(0.323)	(0.224)	(0.594)	(0.266)
BGI	-0.243	-0.189	-0.301	-0.150	-0.281
	(0.012)	(0.097)	(0.032)	(0.206)	(0.049)
Gini	0.098	0.009	0.115	0.120	0.203
	(0.245)	(0.931)	(0.351)	(0.290)	(0.136)
Geo Isolation	-0.082	-0.060	0.069	-0.066	0.136
	(0.335)	(0.517)	(0.722)	(0.489)	(0.451)
GDP (ln)	0.489	0.631	0.398	0.634	0.496
	(0.005)	(0.000)	(0.053)	(0.000)	(0.024)
Population	-0.192	-0.136	-0.267	-0.206	-0.346
	(0.001)	(0.042)	(0.001)	(0.001)	(0.000)
Polity2	0.226	0.151	0.356	0.323	0.596
	(0.079)	(0.277)	(0.035)	(0.036)	(0.001)
Year	–	0.043	0.054	0.069	0.078
		(0.117)	(0.304)	(0.016)	(0.163)
Afrobarometer	-0.792	-0.685	-1.465	-0.381	-0.941
	(0.019)	(0.080)	(0.001)	(0.381)	(0.067)
WVS	-0.025	-0.200	-0.651	-0.168	-0.359
	(0.879)	(0.210)	(0.044)	(0.304)	(0.265)
CSES	0.224	–	–	–	–
	(0.263)				
Constant	0.179	-85.907	-106.888	-137.945	-156.013
	(0.221)	(0.118)	(0.308)	(0.016)	(0.165)
Adj. R-squared	0.829	0.721	0.600	0.701	0.621
N	45	69	44	69	44

P-values based on robust standard errors are in parentheses.

Table 8: Correlation table for right-hand side variables

Variables	ELF	CF-F	CF-D	BGI	GDP(ln)	Polity2	Population	Gini	GI
ELF	1.00								
CF-F	0.64	1.00							
CF-D	0.41	0.75	1.00						
BGI	0.63	0.18	0.23	1.00					
GDP(ln)	-0.63	-0.30	-0.24	-0.35	1.00				
Polity2	-0.54	-0.33	-0.36	-0.26	0.76	1.00			
Population	0.23	0.29	0.27	0.25	-0.13	0.01	1.00		
Gini	0.56	0.18	0.34	0.56	-0.43	-0.43	-0.01	1.00	
GI	-0.57	-0.13	-0.16	-0.62	0.41	0.40	0.07	-0.53	1.00